SPADVROUTE

Deploying Cisco Service Provider Advanced Network Routing

Student Guide

Volume 1

Version 1.2

Part Number: 97-3433-01

https://t.me/learningnets



Americas Headquarters	Asia Pacific Headquarters	Europe Headquarters
Cisco Systems, Inc. San Jose, CA	Cisco Systems (USA) Pte. Ltd. Singapore	Cisco Systems International BV Amsterdam, 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.

https://t.me/learningnets

Welcome Students

Students, this letter describes important course evaluation access information!

Welcome to Cisco Systems Learning. Through the Cisco Learning Partner Program, Cisco is committed to bringing you the highest-quality training in the industry. Cisco learning products are designed to advance your professional goals and give you the expertise that you need to build and maintain strategic networks.

Cisco relies on customer feedback to guide business decisions. Therefore, your valuable input will help shape future Cisco course curricula, products, and training offerings. Please complete a brief Cisco online course evaluation of your instructor and the course materials in this student kit. On the final day of class, your instructor will provide you with a URL, directing you to a short post-course evaluation. If there is no Internet access in the classroom, please complete the evaluation within the next 48 hours or as soon as you can access the web.

On behalf of Cisco, thank you for choosing Cisco Learning Partners for your Internet technology training.

Sincerely,
Cisco Systems Learning

© 2014 Cisco Systems, Inc.

https://t.me/learningnets

https://t.me/learningnets

Table of Contents

Course Introduction **1**
- Overview 1
- Course Goal and Objectives 3
- Course Flow 4
- Your Training Curriculum 5
- Additional References 8

Module 1: Service Provider Connectivity with BGP **9**

Lesson 1: Defining Customer-to-Provider Connectivity Requirements 11
- Cisco IP NGN Architecture 13
- Customer-to-Service Provider Connectivity Types 15
- Single-Homed Customer Connectivity 16
- Dual-Attached Customer Connectivity 17
- Multihomed Customer Connectivity 18
- Routing Schemes 19
- Single-Homed Customer Routing Schemes 20
- Dual-Attached Customer Routing Schemes 21
- Multihomed Customer Routing Schemes 22
- Addressing and AS Number Allocation 23
- Single-Homed Customer IP Addressing Schemes 24
- Dual-Attached Customer IP Addressing and AS Number Schemes 26
- Multihomed Customer IP Addressing and AS Number Schemes 27
- Summary 28

Lesson 2: Connecting a Customer to a Service Provider 29
- Implementing Customer Connectivity Using Static Routing 31
- Single-Homed Customer Using Static Routing and a Single IP Address 32
- Single-Homed Customer Using Static Routing and Multiple IP Addresses 35
- Dual-Attached Customers Using Static Routing in a Primary and Backup Scenario 38
- Dual-Attached Customers Using Static Routing in a Load-Balancing Scenario 43
- Connecting a Dual-Attached Customer to a Single Service Provider 46
- Conditional BGP Advertising 48
- BGP Configurations on the Service Provider PE Router 50
- Removing a Private AS Number 51
- Dual-Attached Customers Using BGP 52
- Service Provider Migrations Using Local AS 54
- Dual-Attached Customers Using BGP in a Primary and Backup Scenario 55
- Dual-Attached Customers Using BGP in a Load-Balancing Scenario 57
- Connecting a Multihomed Customer to Multiple Service Providers 61
- Customer-Implemented BGP Routing Policies in a Primary and Backup Scenario 64
- Customer-Implemented BGP Routing Policies in a Load-Balancing Scenario 67
- Service Provider-Aided BGP Routing Policy in a Primary and Backup Scenario 71
- Summary 74

Lesson 3: Module Summary 77

Lesson 4: Module Self-Check 79

Module Self-Check Answers 81
- Answer Key 81

Module 2: Scale Service Provider Network **83**

Lesson 1: Scaling BGP in Service Provider Networks 85
- Cisco IP NGN Infrastructure Layer 86

https://t.me/learningnets

Service Provider Network Routing Protocols 87
Route Propagation in Service Provider Networks 90
Route Information Exchange Between Service Providers 91
Route Information Exchange with Customers 92
BGP Next-Hop Resolution with IGP 93
Scaling BGP Routing 94
Scaling Addressing in Service Provider Core Networks 95
BGP Policy Accounting 96
Summary 99
Lesson 2: Introducing BGP Route Reflectors and Confederations 101
BGP Route Reflectors and BGP Confederations 103
BGP Split-Horizon Rule 105
Steps of Route Propagation in a Route Reflector-Enabled Network 107
Redundant Route Reflectors 109
Route Reflector Clusters 111
Additional Loop-Prevention Mechanism 113
Network Design with BGP Route Reflectors 114
Hierarchical Route Reflectors 116
Implementing BGP Route Reflectors 118
BGP Confederations Overview 122
AS Path Propagation Within BGP Confederation 123
Intraconfederation EBGP Session Properties 126
Summary 127
Lesson 3: Module Summary 129
Lesson 4: Module Self-Check 131
Module Self-Check Answers 133
Answer Key 133
Module 3: Secure and Optimize BGP 135
Lesson 1: Implementing Advanced BGP Operations 137
Cisco IP NGN Infrastructure Layer 139
Threats in Service Provider Environments 140
BGP Countermeasures Overview 142
BGP Route Limiting 143
BGP Neighbor Authentication 147
BGP TTL Security Check 149
Control Plane Policing 150
BGP Neighbor Authentication, TTL Security, and CoPP Configuration 151
Remote-Triggered Black-Hole Filtering 154
Destination-Based RTBH 155
Source-Based RTBH 157
Cisco Nonstop Forwarding 158
Cisco Nonstop Routing 161
BGP NSF and NSR Configuration 162
BGP Process Restart 163
Summary 165
Lesson 2: Improving BGP Convergence 167
BGP Route-Dampening Overview 168
Configuring BGP Route Dampening 171
Verifying BGP Route Dampening 173
BGP Convergence 175
BGP Processes 176
Improving BGP Convergence 179
Distributed BGP 181

https://t.me/learningnets

PMTU Discovery 183
PMTU Increasing Input Queue Depth 184
PMTU Discovery, Hold Queue, and Distributed BGP Configurations 185
BGP Prefix Independent Convergence 188
Bidirectional Forwarding Detection for BGP 191
BGP Timers and Intervals 194
Summary 198
Lesson 3: Improving BGP Configuration Scalability 199
BGP Peer Groups Overview 201
BGP Peer Groups as Performance Tool 204
BGP Peer Groups Configuration 205
BGP Peer Groups Limitation 206
BGP Dynamic Update Groups 207
BGP Peer Templates Overview 208
BGP Configuration Templates 209
BGP Configuration Templates Inheritance 210
Configuring BGP Configuration Templates 212
BGP Configuration Templates Verification 214
BGP Peer Templates 217
BGP Peer Templates Inheritance 218
BGP Peer Templates Configuration 220
BGP Peer Templates Verification 221
Summary 223
Lesson 4: Module Summary 225
Lesson 5: Module Self-Check 227
Module Self-Check Answers 231
Answer Key 231
Module 4: Multicast Overview 233
Lesson 1: Introducing IP Multicast 235
IP Multicast Benefits and Caveats 237
Multicast Operations High-Level Overview 238
Multicast Advantages and Disadvantages 239
Multicast Application Types 241
IP Multicast Group Address 243
IP Multicast Basic Addressing 244
RFC 2770 and SSM Addressing 246
Multicast Session Directory 248
IP Multicast Service Model 250
Functions of a Multicast Network 252
Multicast Source and Receivers 253
Multicast Protocols 254
Multicast Forwarding and RPF Check 255
Multicast Scoping 256
Summary 260
Lesson 2: Defining Multicast Distribution Trees and Forwarding 261
RPF Check 263
Types of Multicast Distribution Trees 265
Multicast Distribution Tree Identification 268
Multicast Protocols Overview 269
PIM Dense Mode and Sparse Mode High-Level Overview 271
Intradomain Multicast Routing Protocols 273
Interdomain Multicast Routing Protocols 275
Multicast High-Availability Options 278

https://t.me/learningnets

Multicast NSF with Stateful Failover 280
PIM Triggered Joins 281
IGMP Overview 282
IGMPv1 Overview 283
IGMPv2 Overview 285
IGMPv3 Overview 287
Summary 288
Lesson 3: Defining Multicast on the LAN 289
Mapping Multicast IP Addresses to MAC Addresses 291
Layer 2 Multicast Frame Switching 294
Implementing IGMP 298
IGMP join-group and static-group 300
IGMPv3 Host Stack Feature 301
Configuring IGMP Snooping 303
IGMP Fast-Leave in a Switch 305
PIM Snooping 306
Summary 307
Lesson 4: Populating the Mroute Table 309
The Mroute Table 310
Multiprotocol BGP 312
MP-BGP Capability Negotiation 314
MP-BGP Multicast Configuration 315
MP-BGP Multicast Verification 316
Summary 317
Lesson 5: Module Summary 319
Lesson 6: Module Self-Check 321
Module Self-Check Answers 327
Answer Key 327
Glossary 329

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Course Introduction

Overview

Deploying Cisco Service Provider Advanced Network Routing (SPADVROUTE) v1.2 is a course that provides network engineers and technicians with the knowledge and skills that are necessary to implement and support a service provider network.

The course focuses on using Cisco routers that are typically found in the service provider network and on various technologies that are used to offer different services to customers. Upon completing this course, learners will be able to configure, verify, and troubleshoot advanced BGP configuration, IP multicasting, and IPv6 transition mechanisms.

The course also includes classroom activities with remote labs that are useful for gaining practical skills on deploying Cisco IOS, IOS XE, and IOS XR Software features that operate and support the service .

Learner Skills and Knowledge

This subtopic lists the skills and knowledge that learners must possess to benefit fully from the course. The subtopic also includes recommended Cisco learning offerings that learners should first complete to benefit fully from this course.

https://t.me/learningnets

Learner Skills and Knowledge

- Students in this course should have attended the following classes or obtained an equivalent level of training:
  - *Building Cisco Service Provider Next-Generation Networks, Part 1* (SPNGN1) v1.2
  - *Building Cisco Service Provider Next-Generation Networks, Part 2* (SPNGN2) v1.2
  - *Deploying Cisco Service Provider Network Routing* (SPROUTE) v1.2

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Course Goal and Objectives

This topic describes the course goal and objectives.

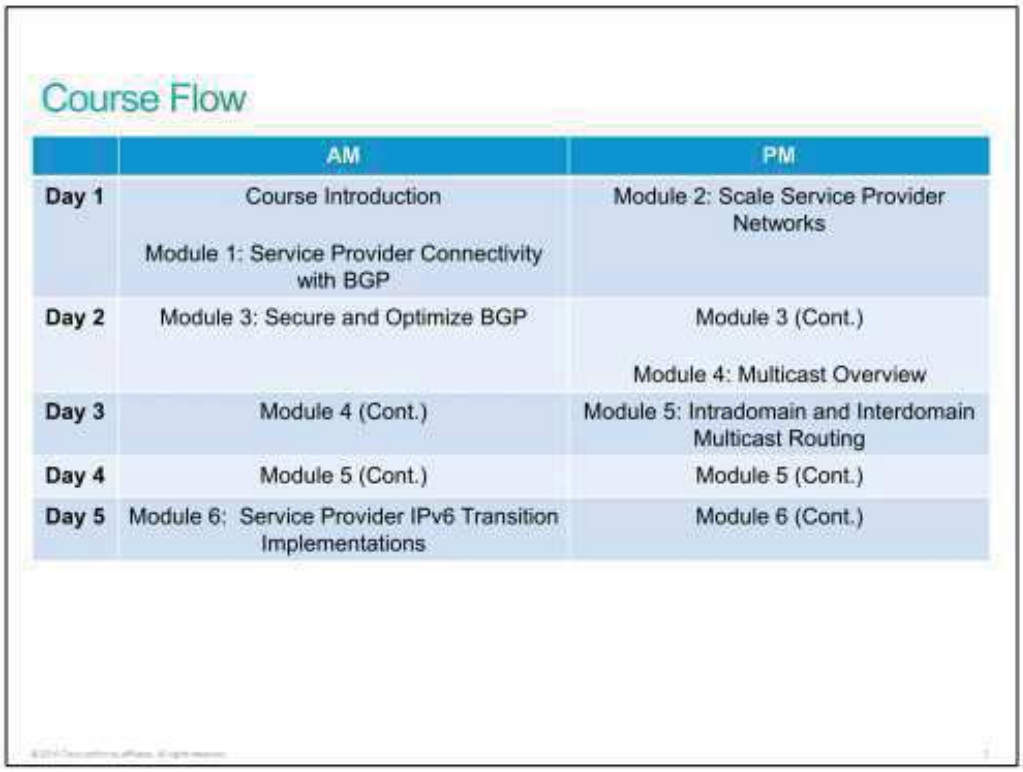
Upon completing this course, you will be able to meet these objectives:

- Configure the service provider network to support multiple BGP connections with customers and other autonomous systems
- Describe common routing and addressing scalability issues in the service provider network
- Describe available BGP tools and features to secure and optimize the BGP routing protocol in a service provider environment
- Introduce IP multicast services and the technologies that are present in IP multicasting
- Introduce PIM-SM as the most current scalable IP multicast routing protocol
- Describe service provider IPv6 transition implementations

https://t.me/learningnets

Course Flow

This topic presents the suggested flow of the course materials.



The schedule reflects the recommended structure for this course. This structure allows enough time for the instructor to present the course information and for you to work through the lab activities. The exact timing of the subject materials and labs depends on the pace of your specific class.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Your Training Curriculum

This topic presents the training curriculum for this course.



You are encouraged to join the Cisco Certification Community, a discussion forum open to anyone holding a valid Cisco Career Certification (such as Cisco CCIE®, CCNA®, CCDA®, CCNP®, CCDP®, CCIP®, CCNP® Security and CCNP® Voice, or CCSP™). It provides a gathering place for Cisco certified professionals to share questions, suggestions, and information about Cisco Career Certification programs and other certification-related topics. For more information, visit http://www.cisco.com/go/certifications.

https://t.me/learningnets

Training Curriculum

This subtopic presents the training curriculum for this course.

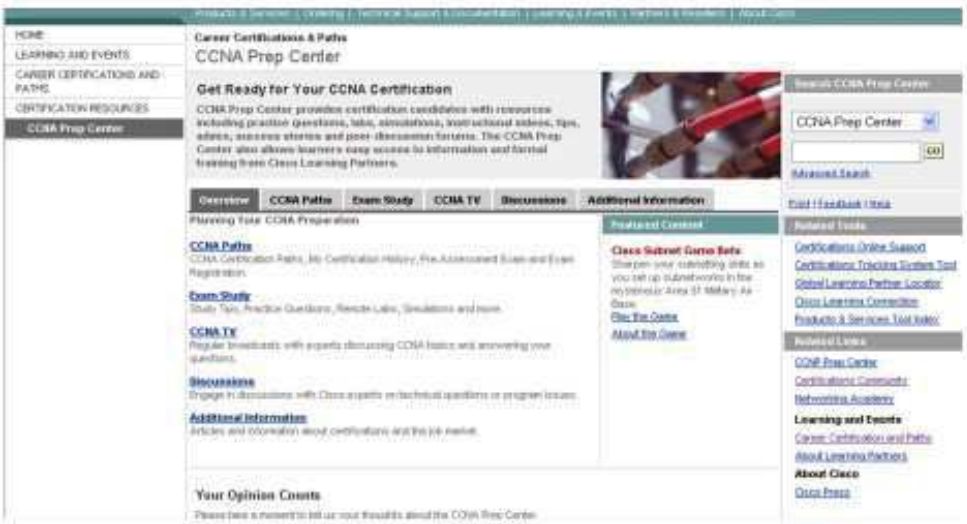


For more information on certifications, visit http://www.cisco.com/go/certifications.

Note	Cisco provides three levels of general certifications for IT professionals with several different tracks to meet individual needs. Cisco also provides focused certifications for designated areas such as cable communications and security. There are many paths to Cisco certification, but only one requirement—passing one or more exams demonstrating knowledge and skill. For details, go to www.cisco.com/go/certifications.

https://t.me/learningnets

CCNA Prep Center



Additional information is available at http://www.cisco.com/go/prepcenter.

Learner Introductions

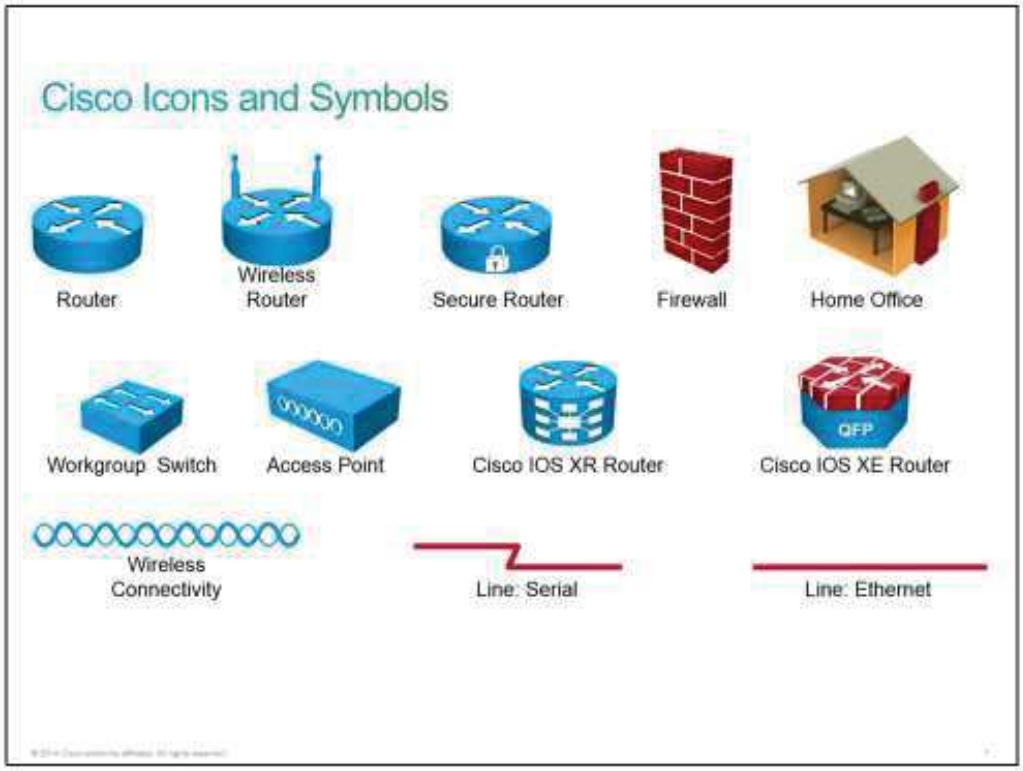
- Your name
- Your company
- Job responsibilities
- Skills and knowledge
- Brief history
- Objective



https://t.me/learningnets

Additional References

This topic presents the Cisco icons and symbols that are used in this course, as well as information on where to find additional technical references.



Cisco Glossary of Terms

For additional information on Cisco terminology, refer to the Cisco Internetworking Terms and Acronyms glossary of terms at http://docwiki.cisco.com/wiki/Category:Internetworking_Terms_and_Acronyms_(ITA).

https://t.me/learningnets

Module 1: Service Provider Connectivity with BGP

Introduction

Customers who are connecting to a service provider must consider various requirements and choose the proper connectivity type. In general, three types of connectivity are available: connecting to a single service provider, connecting to a single service provider using two links, and connecting to two independent service providers. Based on the connectivity type, the proper routing methods, IP addressing, and AS number have to be chosen. After all of these input parameters are known, the customer and service provider can implement the desired connectivity type.

This module first describes connectivity types that are available when connecting a customer to a single or multiple service providers. The module also describes which routing, IP addressing, and AS number options are available in each connectivity type. The module also describes how to configure the service provider and customer device for each connectivity type.

Upon completing this module, you will be able to:

- Define the requirements to connect customer networks to the Internet in a service provider environment
- Describe connecting a single-homed and dual-attached customer to a single service provider and describe connecting a multihomed customer to multiple service providers

https://t.me/learningnets

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Lesson 1: Defining Customer-to-Provider Connectivity Requirements

Overview

The Cisco IP Next-Generation Network is a service provider-targeted framework that provides all-IP transport for customer services and applications, regardless of access type. When planning customer connectivity to the service provider NGN, you have to consider three different connection scenarios: the customer that is connected to a service provider using single link, the customer that is connected to a single service provider using two links, and the customer that is connected to two service providers. You have to carefully examine customer requirements, including redundancy, stability, security, and flexibility, and choose the connection scenario that is based on these requirements. Based on the selected connection scenario, you will also have to choose among different routing options, select proper IP addressing, and select an AS number. This is all required to meet both the business and technical requirements of the applications and services that are planned in the network.

This lesson first briefly describes Cisco IP NGN architecture and positions customer-to-provider connectivity into the NGN concept. The lesson then describes different connectivity scenarios and routing, IP addressing, and AS selection requirements that should be met to satisfy customer requirements.

Upon completing this lesson, you will be able to:

- Show the Cisco IP NGN architecture
- Identify various connection options that are used by customers to connect to service providers
- Describe a single-homed connection
- Describe a dual-attached connection
- Describe a multihomed connection
- Identify various routing schemes that are used by customers to connect to service providers
- Describe the routing schemes used by single-homed customers

https://t.me/learningnets

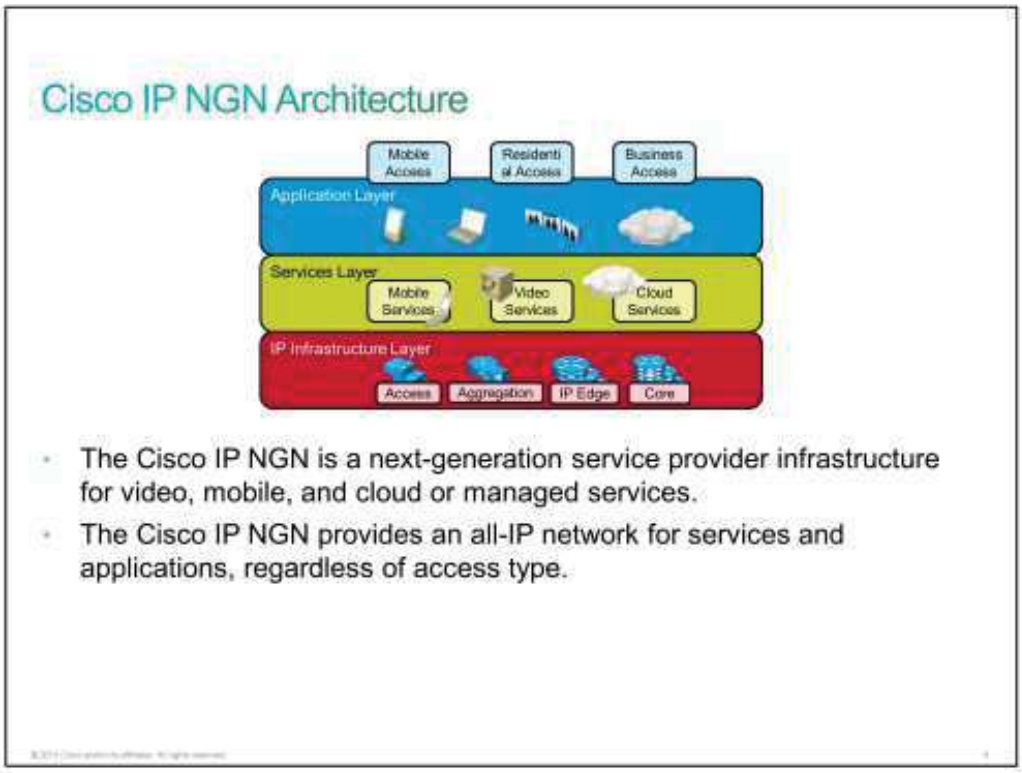
- Describe the routing schemes used by dual-attached customers
- Describe the routing schemes used by multihomed customers
- Describe IP address and AS number schemes that are used by customers to connect to service providers
- Describe IP address schemes that are used by single-homed customers
- Describe IP address and AS number schemes that are used by dual-attached customers
- Describe IP address and AS number schemes that are used by multihomed customers

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Cisco IP NGN Architecture

The architecture enables service providers to start developing fixed and mobile convergence starting with the transport in the access, aggregation, edge, and core networks. The Cisco IP Next-Generation Network targets service providers with an existing centralized wireline services edge network that are willing to maintain and evolve this network layer as part of the future services, network, and organizational evolution.



The Cisco IP NGN architecture constructs a flexible, comprehensive, and generic framework that is structured around the most common layers in the P-networks: customer premises, access network, aggregation network, edge network, core network, network management, and network admission layers. The access, aggregation, edge, and core layers are used for transport of mobile, video, and cloud or managed services.

The general idea of Cisco IP NGN is to provide all-IP transport for all services and applications, regardless of access type. IP infrastructure, service, and application layers are separated in Cisco IP NGN, thus enabling the addition of new services and applications without any changes in the transport network.

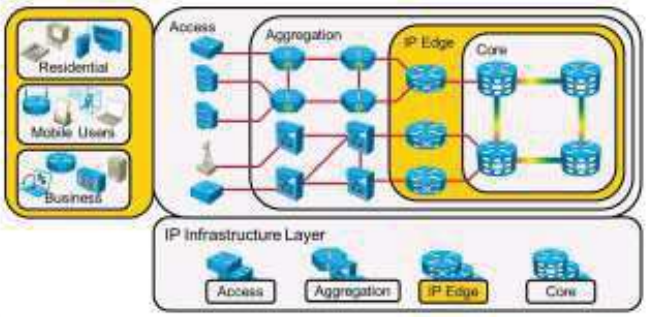
https://t.me/learningnets

Cisco IP NGN Architecture (Cont.)

Cisco IP NGN Infrastructure Layer

Customer-to-provider connectivity focuses on the:

- IP infrastructure layer of the Cisco IP NGN.
- Service provider edge and customer devices.



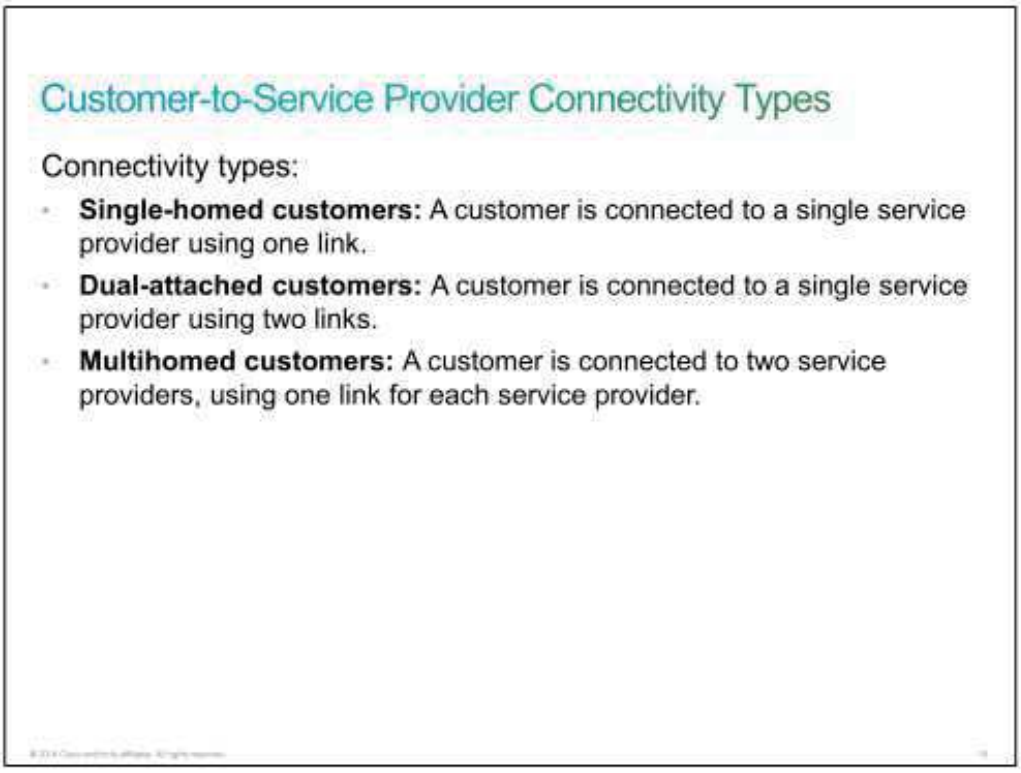
Customer-to-provider connectivity is about connecting a customer to a service provider in a way that meets various customer requirements. Customer-to-provider connectivity is a part of the IP infrastructure layer of the Cisco IP NGN. It focuses on customer-premises equipment and service PE devices.

 © 2014 Cisco Systems, Inc.

https://t.me/learningnets

Customer-to-Service Provider Connectivity Types

When customers connect to a service provider, they can choose from the following connectivity types: single-homed, dual-attached, and multihomed.



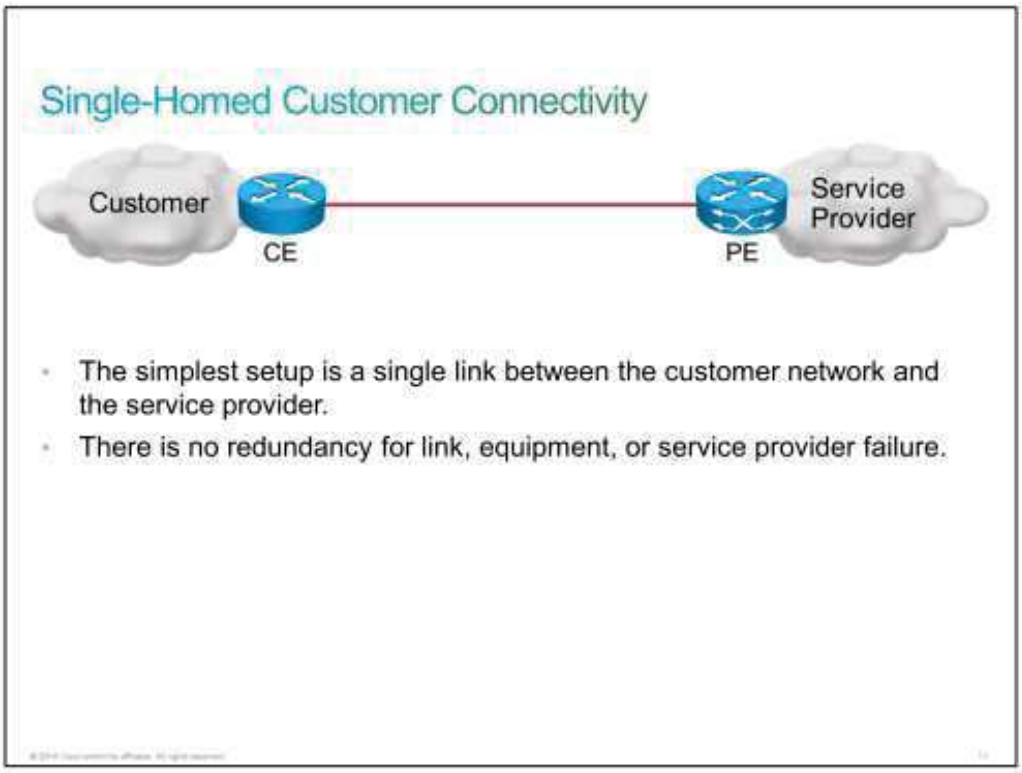
Customers should choose from the connectivity types based on the following:

- Redundancy or high availability
- Stability
- Security
- Flexibility

https://t.me/learningnets

Single-Homed Customer Connectivity

The single-homed connection is the simplest to implement.



In this type of connection, the customer connects to a single service provider using a single link. The customer has a CE router that connects to the PE router using DSL, FTTx, cable, or something equivalent.

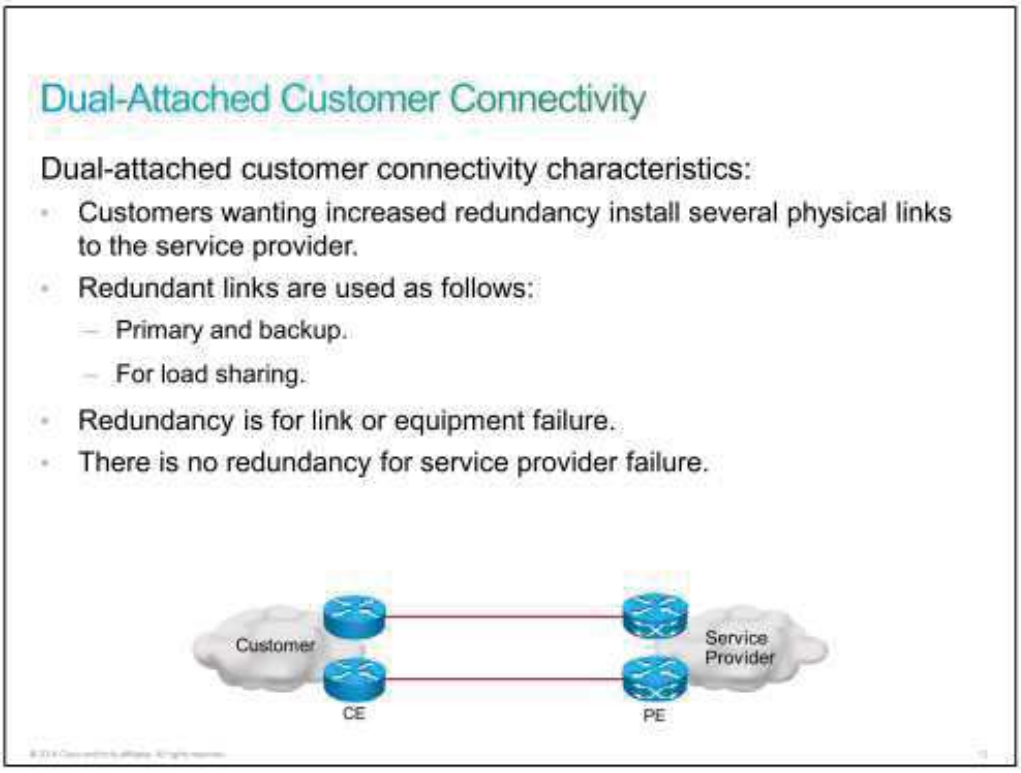
There is no redundancy in this connectivity type; failure on the CE or PE device, on the access link, or in the P-network results in a total outage of customer connectivity.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Dual-Attached Customer Connectivity

In the dual-attached connection, the customer connects to a single service provider using two redundant connections.



One CE router connects to one PE router at the service provider side, and another CE router connects to another PE router.

The two links may be implemented in two scenarios:

- **Primary and backup:** One link is always active, while another one is the backup and is used only when the primary link fails.
- **Load sharing:** Both links are used at the same time to provide load balancing.

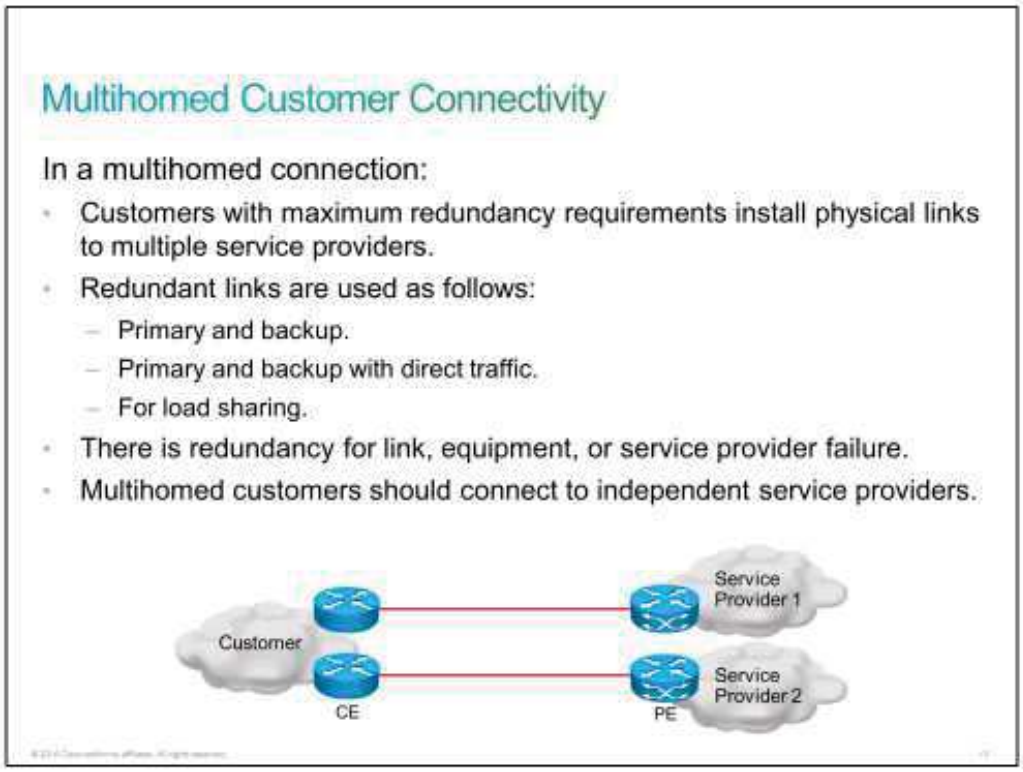
In both scenarios, the customer has to cooperate with a service provider to achieve the desired setup.

Redundancy in this connectivity is provided for link and CE or PE device failure, but the customer will still experience a connectivity outage if a network inside the service provider fails.

https://t.me/learningnets

Multihomed Customer Connectivity

In the multihomed connection, the customer connects to two independent service providers, using one link to each service provider.



Link and device failures are covered in the same way as in the previous example. However, if one service provider fails, traffic will still flow through another service provider, and the highest level of redundancy is achieved.

In this setup, three scenarios are possible:

- **Primary and backup:** One link is always active, while another one is the backup and is used only when the primary link fails.
- **Primary and backup with direct traffic to the backup service provider:** One link is active, and another one is the backup, except for the traffic that should go directly to service provider 2.
- **Load sharing:** Both links are used at the same time to provide load balancing.

In all scenarios, the customer has to cooperate with a service provider to achieve the desired setup. Load sharing between the links can never be optimal. Load distribution over the links is achieved based on destination addresses. Slowly adjusting the router parameters and observing the link traffic load changes can help you achieve an acceptable distribution of traffic between the two links.

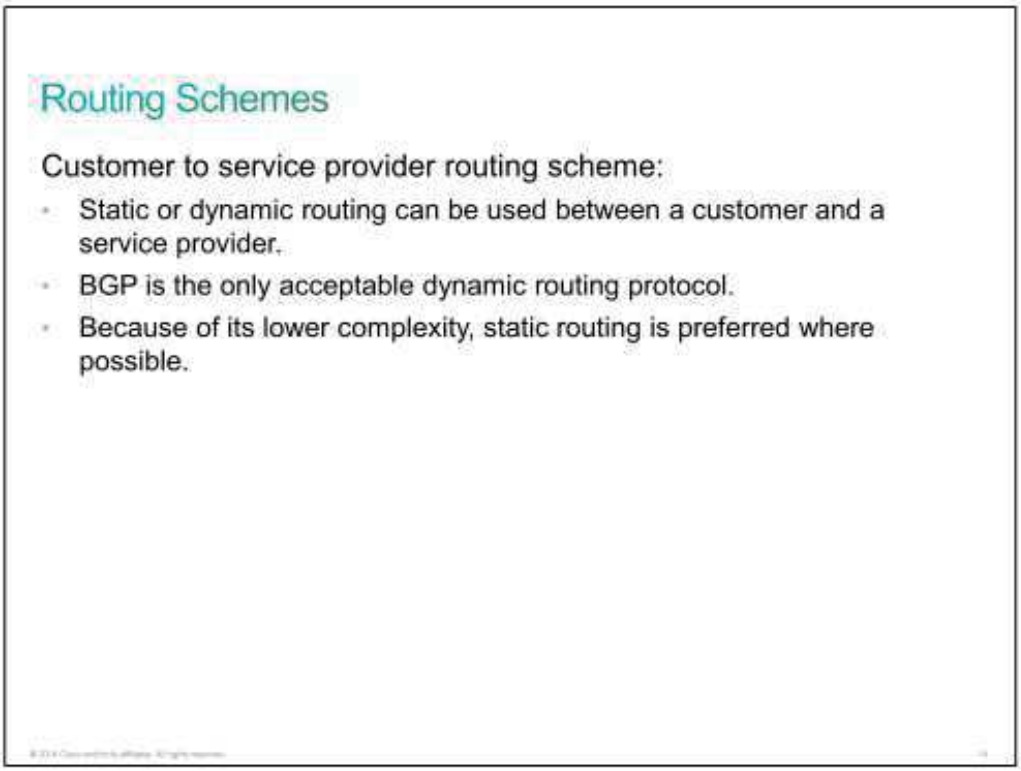
In the multihomed customer connectivity scenario, when there is outage to a local carrier, the customer will lose both connections.

 © 2014 Cisco Systems, Inc.

https://t.me/learningnets

Routing Schemes

Different solutions for connecting a customer network to the network of a service provider require different methods of routing information exchange.



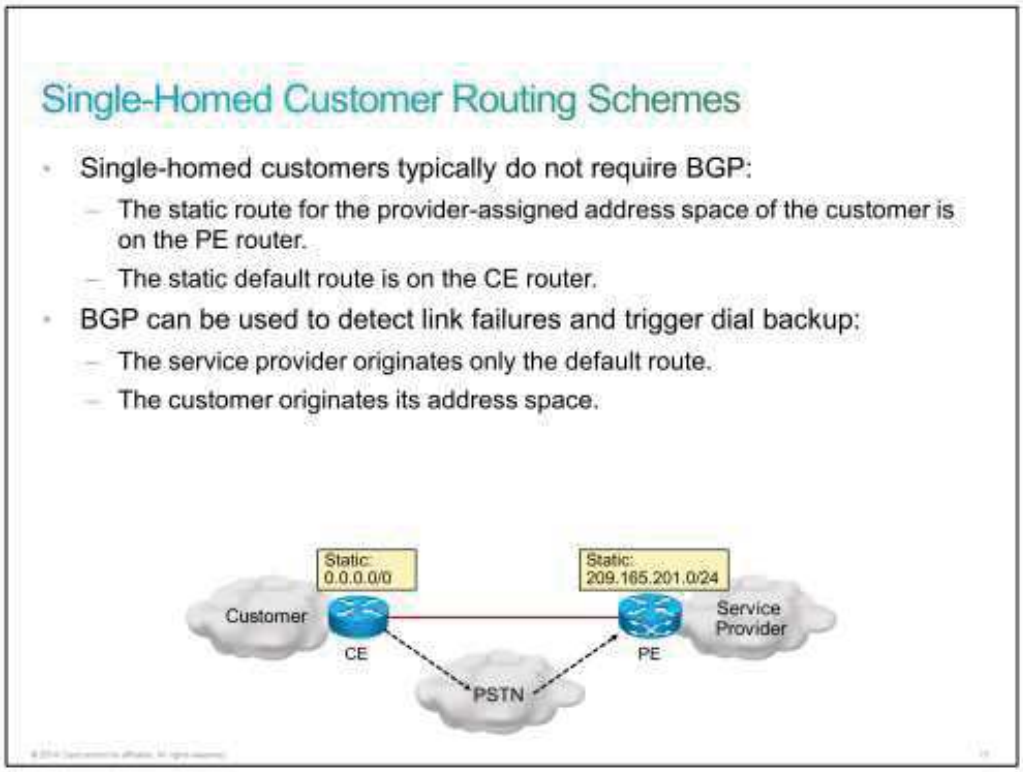
Routing options:

- **Static routing:** Static routing is preferred because of its lower complexity. In a normal case, the customer network must have a default route to the P-network, and the P-network must have a route to the IP prefixes that the customer has in its network. As always, static routing provides very low, if any, redundancy.
- **Dynamic routing:** Dynamic routing provides redundancy. The customer and the P-networks must be configured to exchange a common routing protocol. BGP is the only choice because of the large volumes of routing information, the inherent security mechanisms of BGP, and the ability of BGP to manage routing policies. In the MPLS environment, other routing options are also possible: IGP or static routing.

https://t.me/learningnets

Single-Homed Customer Routing Schemes

When the customer has a single connection to the service provider, static routing is usually adequate. The physical topology does not provide any redundancy, and it is therefore unnecessary to add the complexity of dynamic routing. If there is static routing, a default route is needed on the CE router that points to the PE router, and a static route for the customer network is needed on the PE router, which points to the CE router.

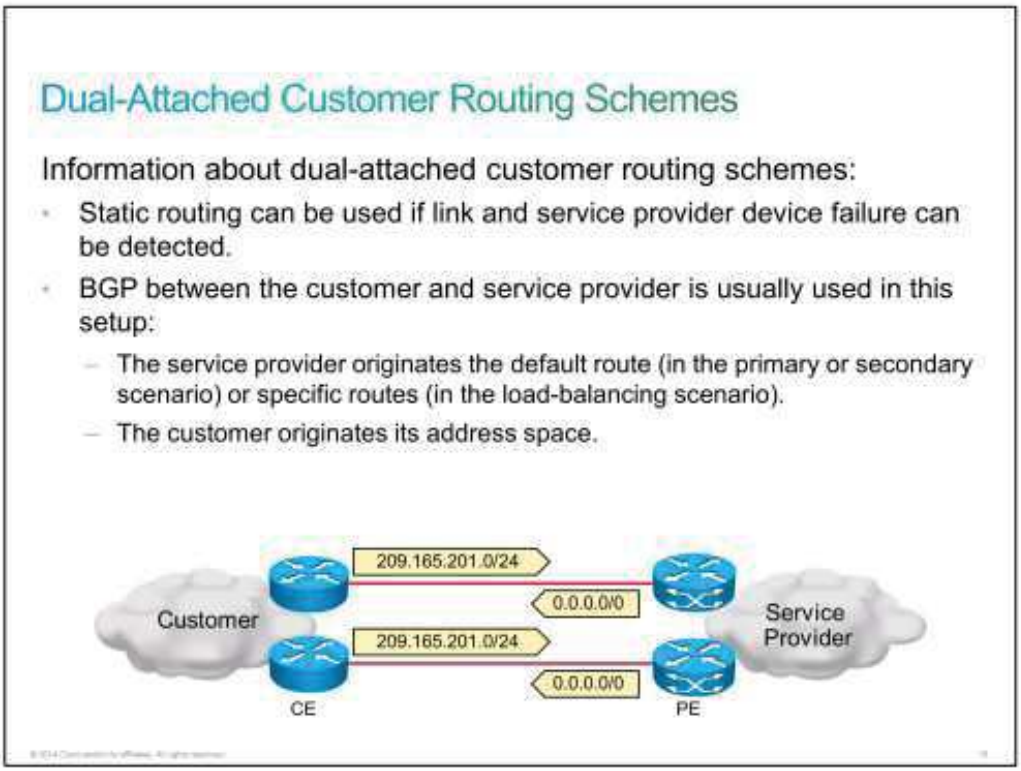


Some single-homed customers may deploy a dial-backup solution in which it is beneficial for them to receive notification if their primary path has failed. If a failure occurs, the BGP session will go down, and the customer can initiate a dial backup connection. If BGP is used, the service provider sends a default route to the customer, and the customer advertises its own address space to the service provider.

https://t.me/learningnets

Dual-Attached Customer Routing Schemes

You can also use static routing for dual-attached connections between two different routers on the customer network and two different routers on the P-network if the failures can be detected by the link-level procedures.



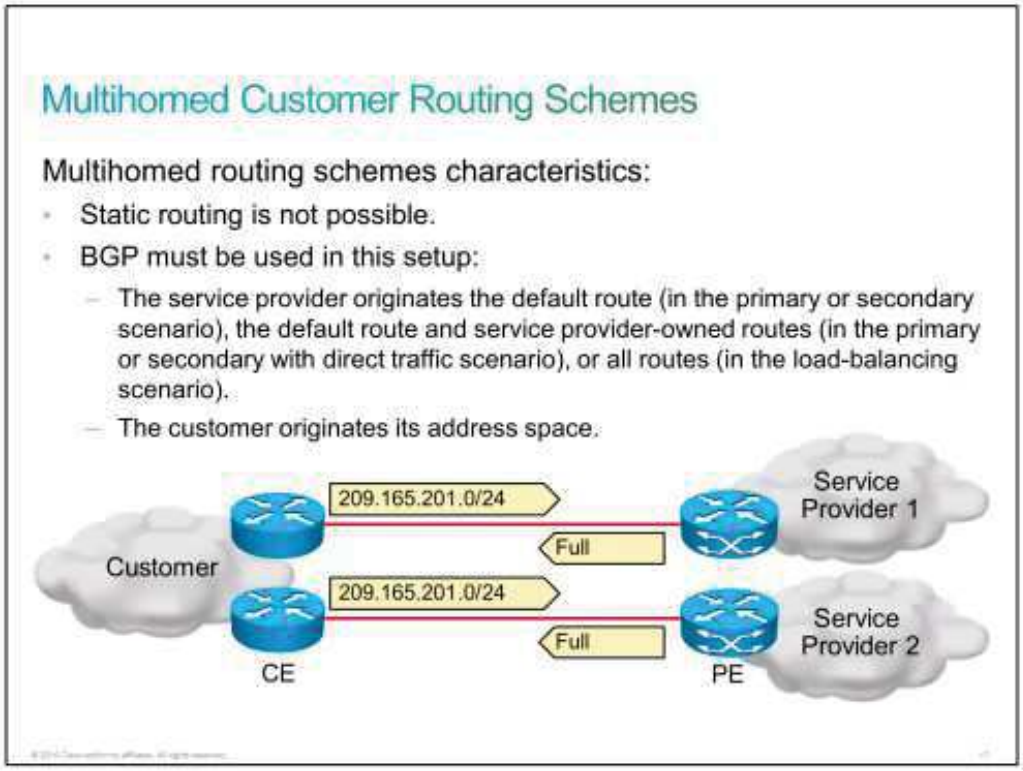
Both CE routers are configured with the default route, which points to the corresponding PE router, and both PE routers are configured with the static route for the customer network, which points to appropriate CE routes. When one of the connections is lost, the link-level procedure detects this loss and places the interface in a down state. Because the interface is in the down state, the static default route that points out of the down interface becomes invalid. As a result, the CE router stops the redistribution of the default route into the customer IGP, and the PE router stops the redistribution of the static route that points the customer network into the service provider BGP.

A better option to detect link failures is using BGP between the CE and PE routers. Because BGP uses handshaking and reliable transfer, it always detects a failed link or failed remote router. BGP can also be used to achieve a primary or secondary scenario or load balancing of traffic over both links.

https://t.me/learningnets

Multihomed Customer Routing Schemes

In the multihomed customer connection, static routing cannot be used.



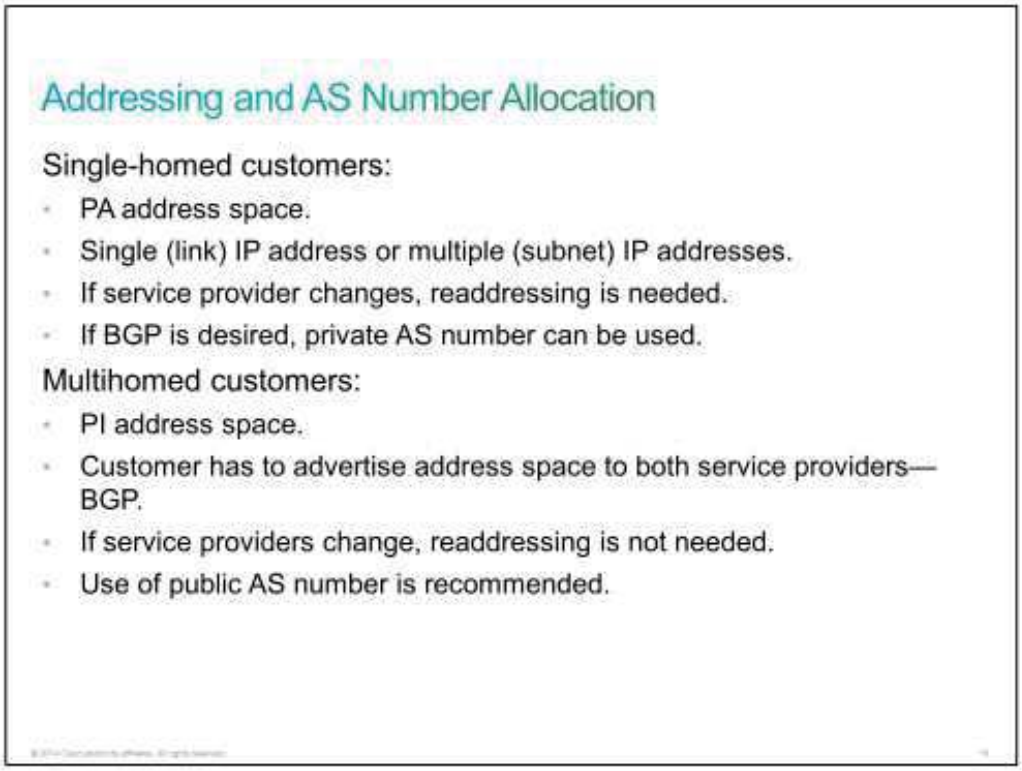
In this setup, BGP has to be used to coordinate which link will be used. CE routers advertise the customer network to service providers, and PE routers advertise only the default route or the default route plus the service provider-owned networks or all routes to the CE routers. The level of routes that are advertised by the service provider is determined based on the required customer scenario.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Addressing and AS Number Allocation

Customers that are connected to a single service provider usually have their address space assigned by the service provider, known as PA addresses. Customers can either request a single IP address, which is used on the access link as well, or they can request an entire IP subnet. In the latter case, the service provider provides another small subnet for the access link only. A service provider is usually assigned a large address space to delegate to its customers. Because all customers of one service provider get their addresses from one address space or a few address spaces, it is very likely that the service provider is able to aggregate the customer addresses before sending the routes to the rest of the Internet. If BGP is desired by a customer to detect link failures, the service provider can assign a private AS number to the customer. In this case, the service provider has to remove this private AS number from the AS path when advertising the network to upstream service providers.



If the customer should decide to change its service provider, the customer must return its PA addresses to the old service provider and receive a new assignment of PA addresses from the new service provider. Otherwise, the service providers are no longer able to perform efficient address aggregation. The consequence for the customer is that the customer has to renumber its network when it changes its service provider.

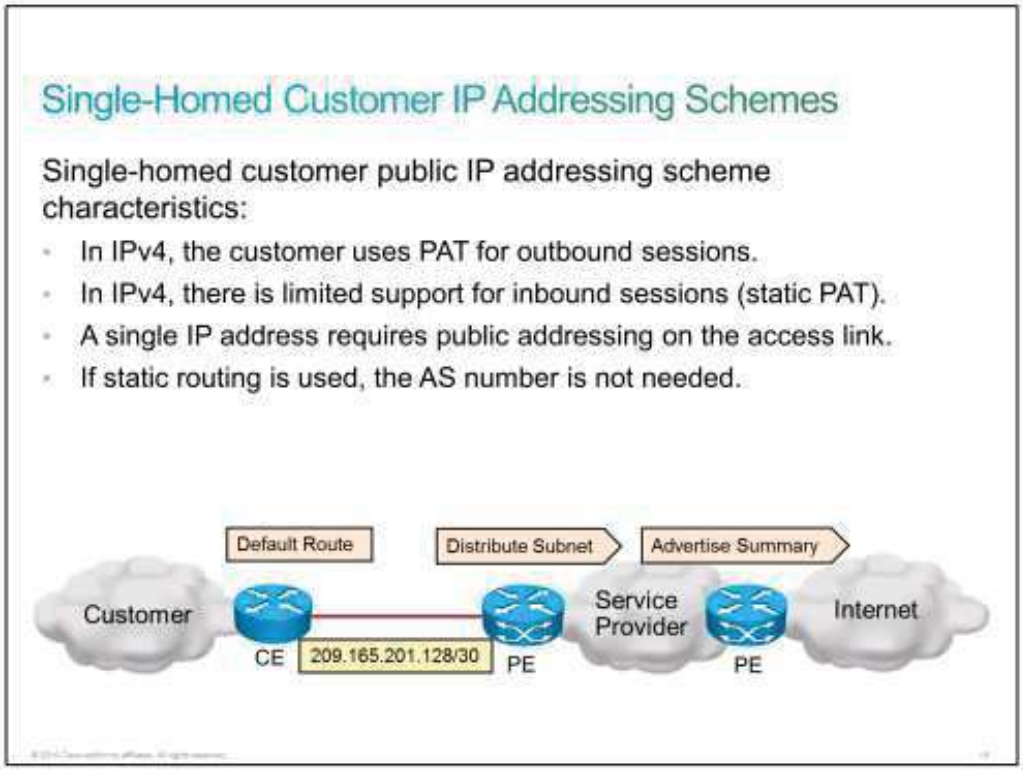
Customers that are multihomed should use a provider-independent IP subnet. This IP subnet can be obtained from the LIR, which has been allocated a large block of IP addresses from the RIR. PI address space is not advertised by any service provider, and the customer has to advertise it to both service providers using BGP. Because BGP is used, the customer should be assigned a public AS number.

If the customer should decide to change its service provider, renumbering of the network is not needed.

https://t.me/learningnets

Single-Homed Customer IP Addressing Schemes

Single-homed customers may use single or multiple IP addresses.



The figure shows an example of a single IP address from a small subnet (/30) that is assigned to the customer. This example applies only to IPv4. This IP address is also used on the CE router for the access link. An IP address from the same subnet is assigned to the PE router for the access link. The customer has to perform NAT with PAT to provide connectivity for inside hosts to the service provider and to the rest of the Internet. Because a single, public IP address is available, the only way to provide support for inbound sessions is to use static PAT. Because BGP is not needed in this example, the AS number is not required by the customer.

The customer configures the default route on the CE router, which points to the PE router. The service provider redistributes the access link subnet into the service provider BGP on the PE router. This is usually done by using redistribution of static routes into BGP. The service provider should summarize all customer networks to a larger network when advertising the network to upstream service providers.

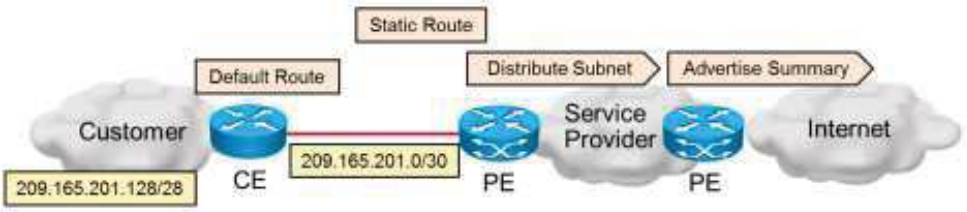
 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Single-Homed Customer IP Addressing Schemes (Cont.)

Multiple IP addresses schemes:

- The public subnet is assigned to the customer from the service provider pool.
- The access link uses public addresses.
- Multiple IP addresses support customer services that require inbound sessions.
- If static routing is used, the AS number is not needed.



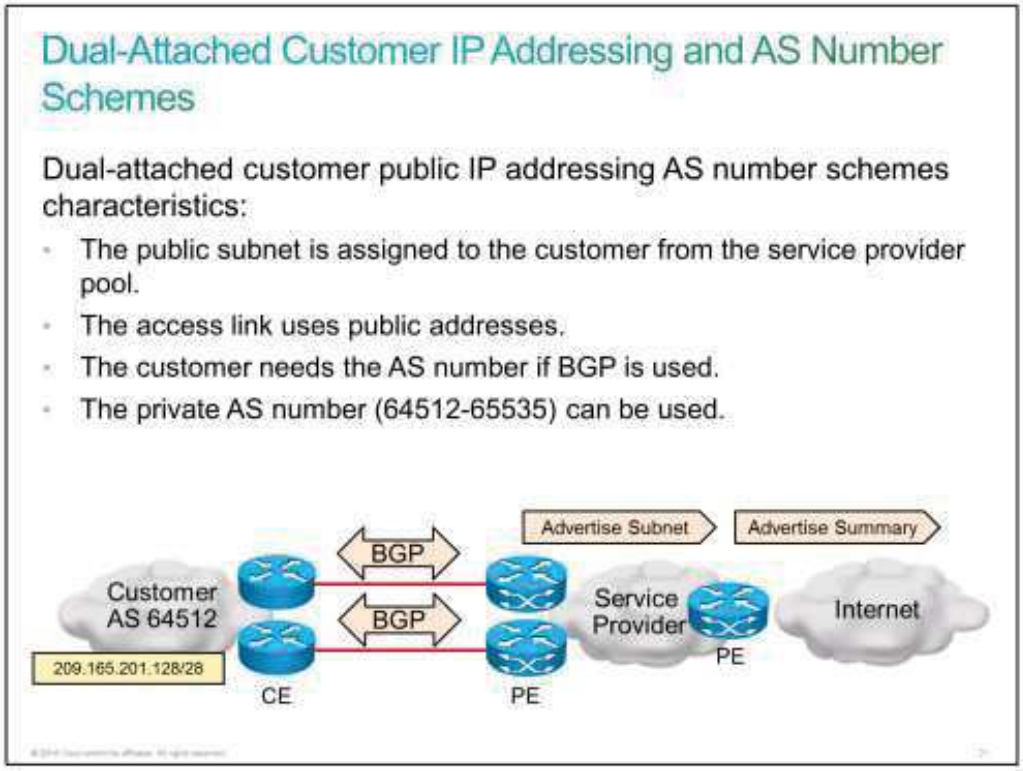
In the example, the entire subnet (/28) is assigned to the customer. This IP subnet is assigned from the service provider address pool. Another small public subnet (usually /30) is assigned to the access link. This example better supports customer services that require inbound sessions because there are many public IP addresses that are available. In this setup, static routing is usually used; therefore, the AS number is not needed by the customer.

The customer configures the default route on the CE router, which points to the PE router. The service provider configures the static route for the customer network, which points to the CE router. The service provider then redistributes the static route into the service provider BGP on the PE router. The service provider should summarize all customer networks to a larger network when advertising the network to upstream service providers.

https://t.me/learningnets

Dual-Attached Customer IP Addressing and AS Number Schemes

In the dual-attached customer example, addressing remains the same as in the previous example. The customer is assigned a public subnet from the service provider pool, and two additional IP subnets are used for access links. In this scenario, the use of BGP is preferred to exchange routing information, enable primary and backup routing or load balancing, and have the ability to detect failed links. A customer that uses BGP to exchange routing information with only one service provider does not require a public AS number. This customer can use a private AS number. When the customer can use a private AS number, the service provider must allocate one from the range of private AS numbers (such as 64512 to 65535). The service provider must make sure not to assign any of the private AS numbers to more than one customer.



When the service provider receives BGP routes from the customer, the service provider routers see the private AS number in the AS path and treat the private number as any other AS number. However, before the service provider propagates any of these routes to the rest of the Internet, it must remove the private AS numbers from the AS path because the same AS number may be in use by someone else. After the private AS number is removed, the route appears as belonging to the public AS of the service provider.

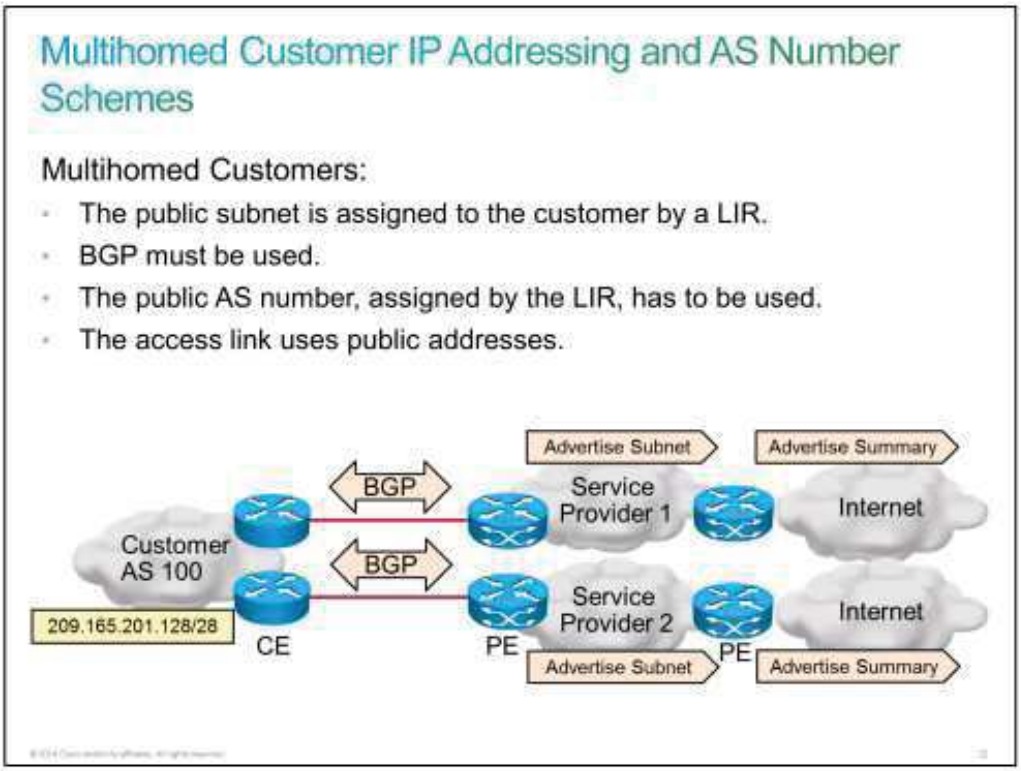
The customer runs BGP on the CE routers, peers with the service provider, and advertises the customer subnet. The service provider advertises the subnet further on the PE routers. The service provider should summarize all customer subnets before advertising them to upstream service providers and to the Internet. Both the customer and service provider should manipulate BGP routing updates to achieve either primary or secondary setup or load-balancing setup. The BGP route updates manipulation will be further described in the next lesson.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Multihomed Customer IP Addressing and AS Number Schemes

In the multihomed customer example, the customer is assigned a provider-independent public address space. This address space can be requested from a LIR. Each service provider also assigns a small subnet from its address space to the access link. Because the customer has to advertise the address space to both service providers, the customer has to use BGP. Because the PI address space should appear in the Internet as originating from the customer, the public AS number should be used by the customer. (Recall that a service provider has to remove all private AS numbers from an AS path when propagating routes to the Internet.) The public AS number can be requested from a LIR as well.

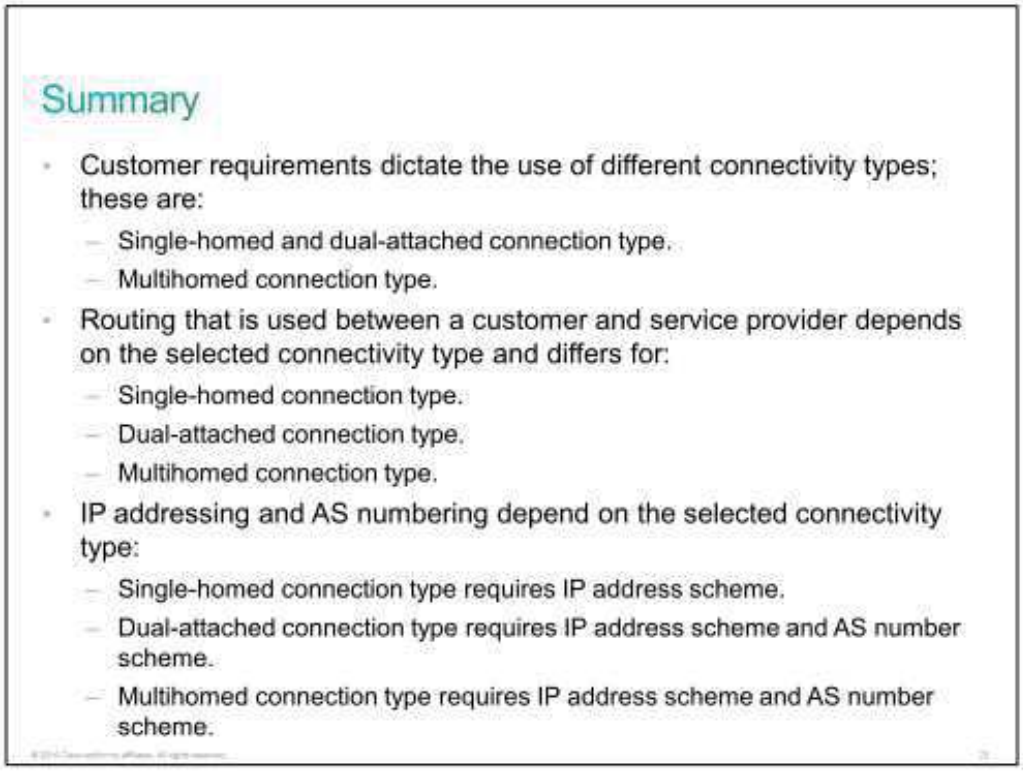


The customer runs BGP on the CE routers, peers with service providers, and advertises the customer subnet. Service providers advertise the subnet further on the PE routers. The service provider should summarize all customer subnets before advertising them to the Internet, if possible. The customer and service providers should manipulate BGP routing updates to achieve either primary or secondary setup or load-balancing setup. BGP route updates manipulation will be further described in the next lesson.

https://t.me/learningnets

Summary

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



 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Lesson 2: Connecting a Customer to a Service Provider

Overview

When a customer connects to one or more service providers, different routing options are available. The simplest scenario, where a customer connects to one service provider using a single link, requires only static routing. However, this connectivity offers no redundancy if there is an equipment, link, or service provider failure. Static routing could also be used when a customer connects to a single service provider using two redundant links. This scenario offers redundancy if there is equipment and link failure but not if there is service provider failure. When a customer connects to one service provider using two links, BGP can be used as an alternative to static routing. The most resilient setup to failures is multihoming with two different service providers. In multihoming, the only acceptable routing solution is BGP.

This lesson first describes static routing and describes which scenarios that static routing could be used to connect a customer to a service provider. The lesson then continues with a customer that is connected to a single service provider using two links and using BGP as a routing solution. Both the primary and backup and load-balancing scenarios are discussed. Finally, the lesson describes how to connect a customer to two service providers to achieve the highest level of redundancy. Again, both the primary and backup and load-balancing scenarios are discussed.

Upon completing this lesson, you will be able to:

- Describe customer connectivity using static routes in single- and dual-attached scenarios
- Describe customer connectivity using static routes in single-homed scenarios using a single IP address
- Describe customer connectivity using static routes in single-homed scenarios using multiple IP addresses
- Describe customer connectivity using static routes in dual-attached scenarios using a primary and a backup path
- Describe customer connectivity using static routes in dual-attached scenarios using load balancing

https://t.me/learningnets

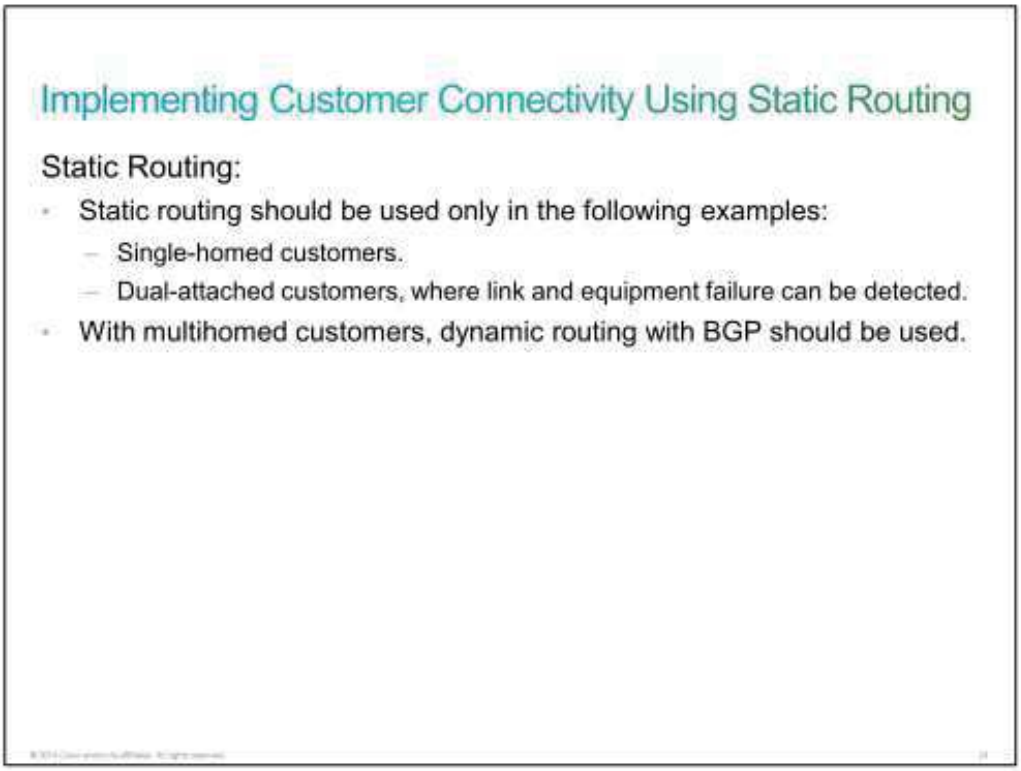
- Describe customer connectivity using BGP in a dual-attached scenario
- Explain using conditional BGP advertising at the CE router in a dual-attached scenario
- Describe BGP configuration on the PE routers in a dual-attached scenario
- Describe PE router removal of a private AS number from the AS path in a dual-attached scenario
- Explain CE and PE router BGP configurations in a dual-attached scenario
- Show an example of how the local AS BGP feature is used
- Describe customer connectivity using BGP in dual-attached scenarios using a primary and a backup path
- Describe customer connectivity using BGP in dual-attached scenarios using load balancing
- Describe customer connectivity using BGP in a multihomed scenario
- Describe customer connectivity using BGP in multihomed scenarios using a primary and a backup path
- Describe customer connectivity using BGP in multihomed scenarios using load balancing
- Describe a service provider-aided routing policy using BGP in multihomed scenarios using a primary and a backup path

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Implementing Customer Connectivity Using Static Routing

Static routing is the best solution to implement when there is no redundancy in the network topology. A single connection between the customer network and the P-network does not provide any redundancy. If the link goes down, the connection is lost regardless of which routing protocol is configured in the customer or P-network. When there are redundant connections between the customer network and the network of a single service provider, static routing can be used under specific circumstances.



A static default route must be conditionally announced by the CE routers that are using an interior gateway protocol IGP. If the link to one of the CE routers goes down, then the router must be able to detect the failure and invalidate the static default route. Announcement of this router as a default gateway that is using an IGP must now cease. Likewise, on the provider edge PE routers, the static routes that are pointing to the customer networks must be invalidated if the link between them goes down, and redistribution to BGP is therefore stopped.

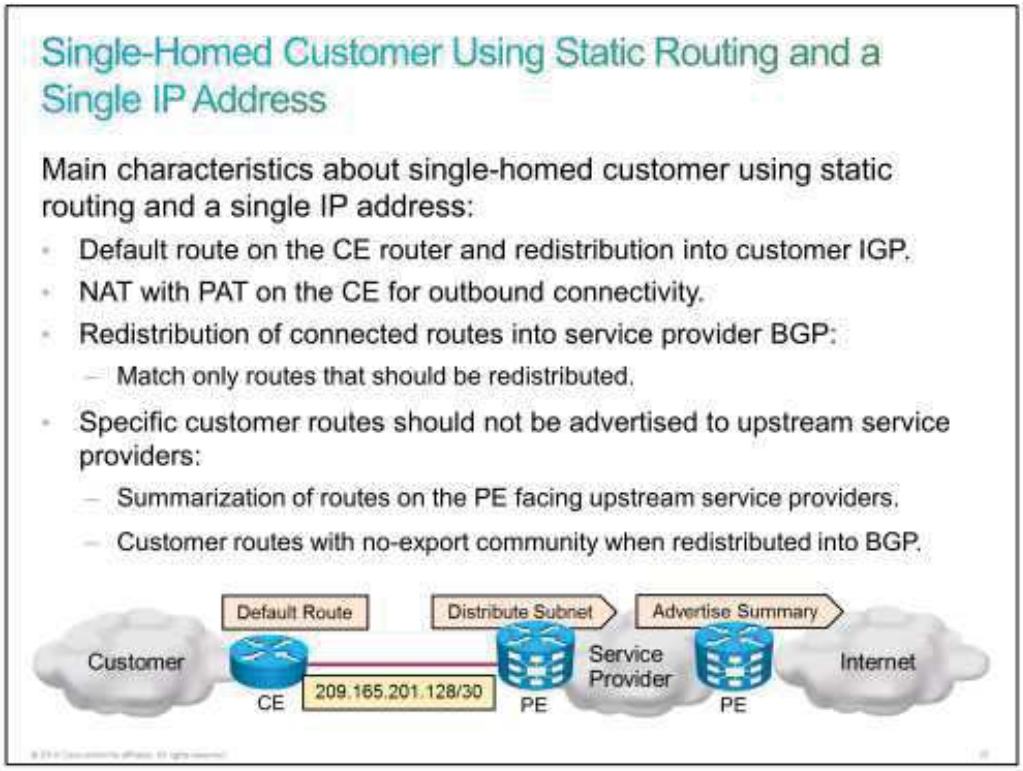
If link-level procedures cannot detect a link failure, the interface remains in the up state. The static routes are not invalidated, and packets are forwarded into a "black hole." In these cases, because the router cannot detect a failure at the link level, BGP must be used between the customer and the service provider.

BGP must also be used between the customer and the P-networks when the customer is multihomed. This is true regardless of which link failure detection mechanisms are in use.

https://t.me/learningnets

Single-Homed Customer Using Static Routing and a Single IP Address

The most common option for small customers and residential users is to receive a single IP address from the service provider. In this scenario, public IP addresses are assigned to the access link only, usually with mask /30. The customer uses its own, private addressing space inside the customer network and uses NAT with PAT, where private customer IP addresses are translated into the public IP address of the CE interface pointing to the service provider. When a service provider provides a single IP address only, and support for inbound sessions is required, static PAT (or port forwarding) should be used.



Routing is simplified; on the customer side, only static default route is needed. This default route should be redistributed into the customer IGP, if used.

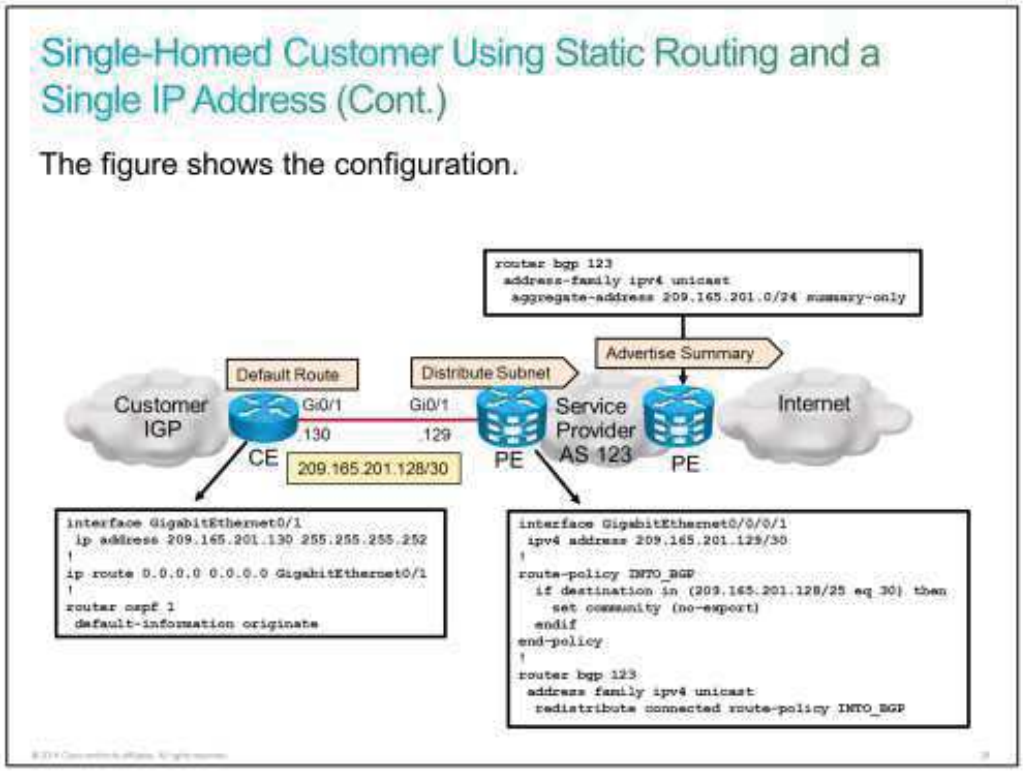
On the service provider side, no routes are needed because the access link is directly connected to the PE router. Directly connected access links should be redistributed into BGP to send the routing information further through the P-network and to upstream service providers. The service provider should carefully redistribute only those connected routes that are used by single IP address customers.

Note	Access links IP addresses can also be advertised inside the service provider IGP (that is, OSPF or IS-IS). In this case, access links IP addresses are redistributed into BGP at the border with upstream service providers.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Because the service provider probably uses contiguous address space (apart from the service provider-owned address space) for access link addressing, all customer-assigned IP addresses could be summarized. It is important that the service provider advertises only a summary route to upstream service providers. Individual routes should not be advertised to upstream service providers. You can make individual routes ineligible for advertising outside of the local AS by setting the predefined BGP no-export community. If a router in a P-network receives an update carrying this community, the router will not propagate the update to any external BGP neighbors, meaning the update is not propagated to other service providers. The no-export community is set on the route when the route is redistributed to BGP on the customer-facing PE router.



The example shows a configuration on CE and PE routers, where the access link is used and directly connected routes are redistributed into BGP on the PE router. The CE router is configured with static default route, which points to the PE router. The default route is also redistributed into OSPF, which is used in the customer network.

The PE router, which runs Cisco IOS XR Software, is configured with BGP. An IBGP neighbor relationship is established between the PE router and other routers in the P-network. (The configuration is not shown.) The connected route that is redistributed into BGP is marked with a predefined community called no-export. The edge routers of the P-network then see out of the no-export community and filter those routes before sending the update to EBGP neighbors. Communities are set using RPL. The route policy in the example matches routes with /30 prefix length, which fall into the 209.165.201.128/25 major subnet. Recall that the set action inside the route policy allows a matched route to be redistributed.

The Internet-facing PE router is configured with a summarization of all customers that are assigned IP addresses into 209.165.201.0/24. Only the summary route is then advertised into the Internet.

Configuration of the customer-facing PE router, running Cisco IOS and IOS XE Software, is as follows:

https://t.me/learningnets

```
interface GigabitEthernet0/1
 ip address 209.165.201.129/30
!
ip prefix-list CUSTOMER_ROUTES permit 209.165.201.128/25 ge 30 le 30
!
route-map INTO_BGP permit 10
  match ip address prefix-list CUSTOMER_ROUTES
  set community no-export
!
router bgp 123
 redistribute connected route-map INTO_BGP
 neighbor 209.165.201.10 send-community
```

To match routes that should be redistributed into BGP, the prefix list is used. The prefix list matches the first 25 bits of the prefix 209.165.201.128. The subnet mask must be greater than or equal to 30 and less than or equal to 30, which means exactly 30.

Configuration for IPv6 is very similar. All IPv4 addresses should be changed with IPv6 addresses. On the CE router, OSPFv3 should be configured to route IPv6 routes. On the PE router, redistribution of static routes should be done under the IPv6 address family.

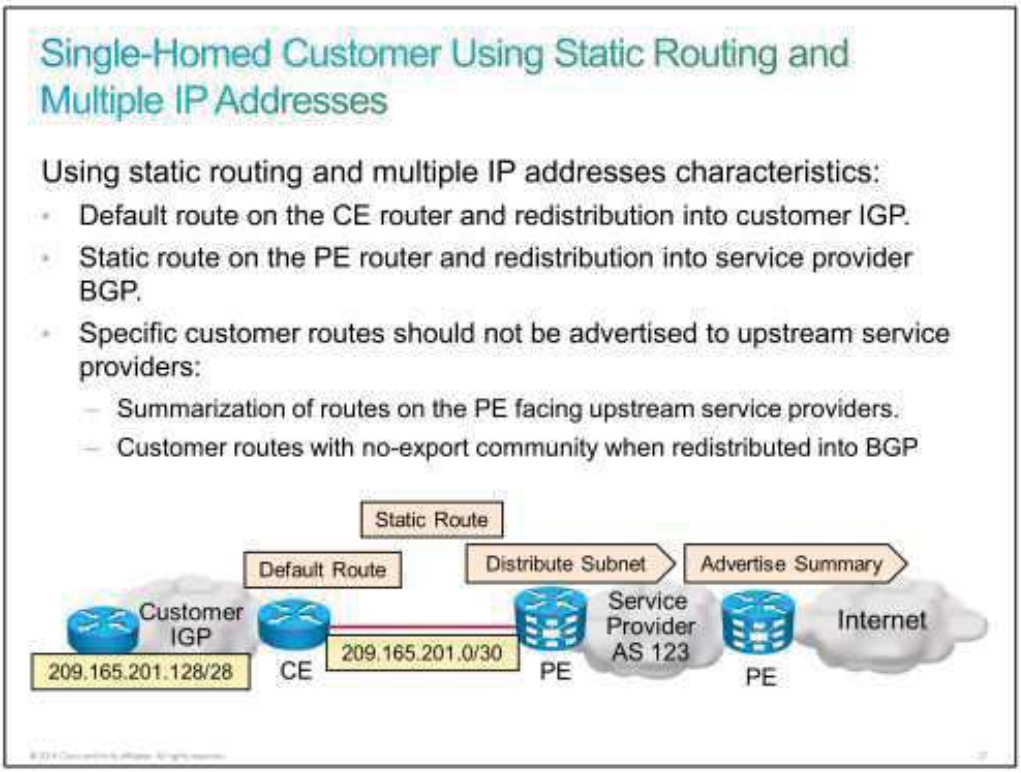
Note	For a complete command reference, refer to the Cisco IOS, IOS XE, and IOS XR Software Command Reference guides on www.cisco.com.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Single-Homed Customer Using Static Routing and Multiple IP Addresses

For customers that need multiple IP addresses, the service provider can assign a separate address space for a customer. One small public subnet (/30) is still used for access link addressing. The assigned IP subnet in a single-homed scenario can be provider assigned.



The figure shows an example of a PA address. The service provider assigns IP addresses that are public to the access link. The customer receives a public subnet from the provider pool. NAT with PAT could still be used if there are not enough public IP addresses for the customer network.

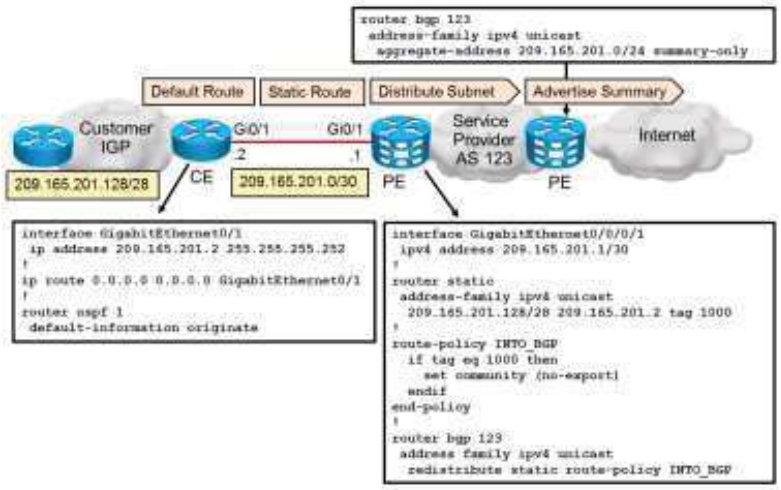
On the CE router, a static default route, which points to the PE router, should be configured. This default route should be redistributed into the customer IGP, if used.

On the customer-facing PE router, a static route for the customer network should be configured. The static route should be tagged with a tag. The tag is used to match static routes that are eligible for redistribution into BGP. Redistribution into BGP ensures that all other service provider routers and the rest of the Internet will receive information about the provider-assigned subnet of the customer. The service provider should summarize provider-assigned subnets before sending the routes to the rest of the Internet. To make sure that individual customer networks are not propagated outside of the service provider, the routes are marked with no-export community when redistributed into BGP.

https://t.me/learningnets

Single-Homed Customer Using Static Routing and Multiple IP Addresses (Cont.)

Configuration:



The example shows a configuration on CE and PE routers, where PA address space is assigned to the customer and advertised inside the P-network and beyond using BGP.

The CE router, which runs Cisco IOS Software, is configured with static default route, which points to the PE router. The default route is also redistributed into OSPF, which is used in the customer network.

The PE router, which runs Cisco IOS XR Software, is configured with BGP. An IBGP neighbor relationship is established between the PE router and other routers in the P-network. (The configuration is not shown.) On the PE routes, a static route for the customer network is configured, which points to the CE router. The static route is tagged with tag 1000. Routes that are tagged with tag 1000 are then matched inside the route policy and set with the BGP no-export community.

The Internet-facing PE router is configured with a summarization of all customers that are assigned IP addresses into 209.165.201.0/24.

Configuration of the customer-facing PE router, running Cisco IOS and IOS XE Software, is as follows:

```
interface GigabitEthernet0/1
 ip address 209.165.201.1/30
!
ip route 209.165.201.128 255.255.255.240 209.165.201.2 tag 1000
!
route-map INTO_BGP permit 10
  match tag 1000
  set community no-export
!
router bgp 123
 redistribute static route-map INTO_BGP
```

Configuration for IPv6 is very similar. All IPv4 addresses should be changed with IPv6 addresses. On the CE router, OSPFv3 should be configured to route IPv6 routes. On the PE router, static routes, BGP configuration, and redistribution of static routes should be done under the IPv6 address family. This applies to all examples in the lesson.

 © 2014 Cisco Systems, Inc.

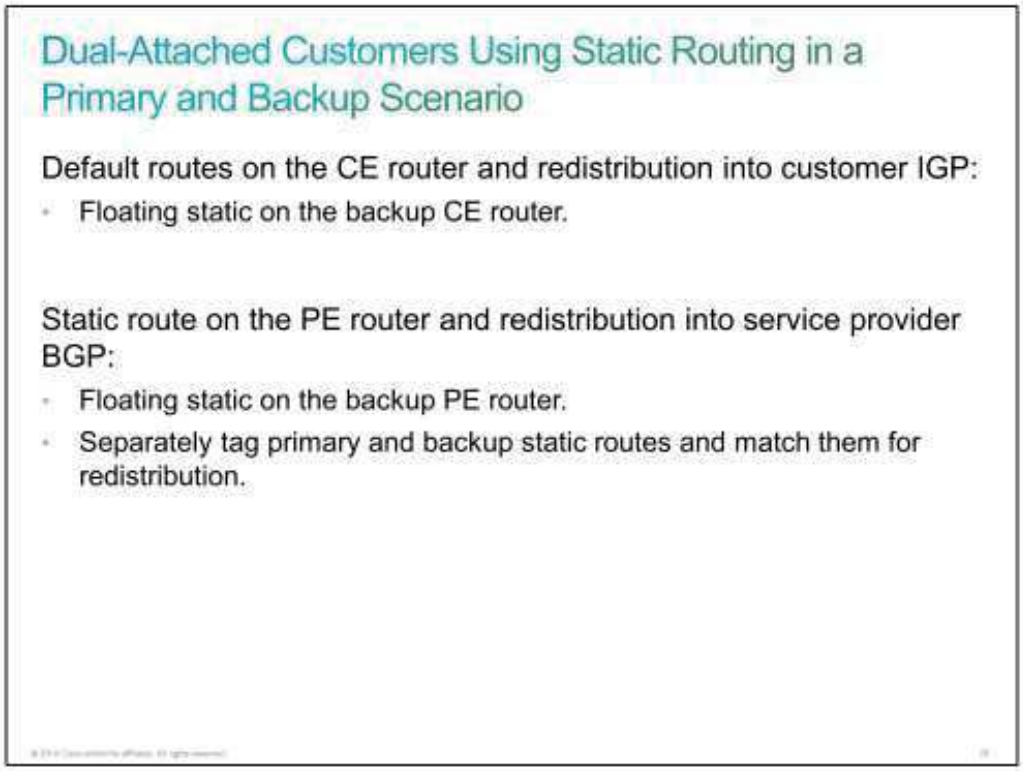
https://t.me/learningnets

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

https://t.me/learningnets

Dual-Attached Customers Using Static Routing in a Primary and Backup Scenario

If a customer requires a higher level of redundancy but wants to be connected to a single service provider, the service provider can offer two connections. A customer that is connected to a single service provider is called a dual-attached customer. The example illustrates a case where the customer network has two connections to a single service provider, using two routers on each side. One connection between the customer network and the service provider is the primary connection, and the other connection is used for backup purposes only. If link-level procedures can detect link failures and a failure in the remote router, then static routing can be used instead of a dynamic routing protocol between the customer and P-networks.



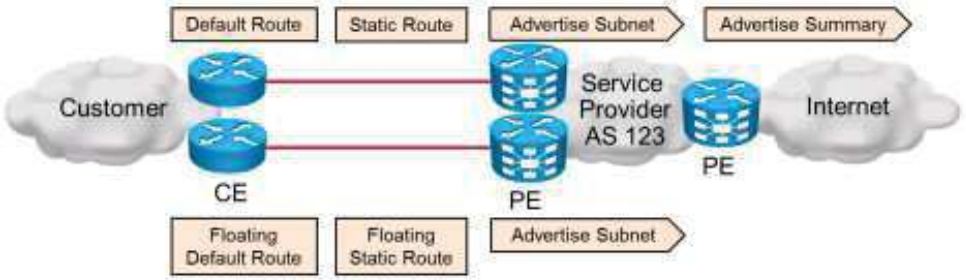
 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Dual-Attached Customers Using Static Routing in a Primary and Backup Scenario (Cont.)

Specific customer routes should not be advertised to upstream service providers:

- Summarization of routes on the PE facing upstream service providers.
- Customer routes with no-export community when redistributed into BGP.



As in the previous example, where no backup link is available, the CE router has a static default route toward the service provider, and the primary PE router has a static route toward the customer. The customer router redistributes the static default route into its IGP. The service provider router redistributes the static route into BGP.

If the primary link goes down, the link-level procedures set the interface to the down state, causing the static routes pointing out through the interface to be invalid and removing the routes from the routing table. When the interface changes back to the up state, the static route will reappear in the routing table.

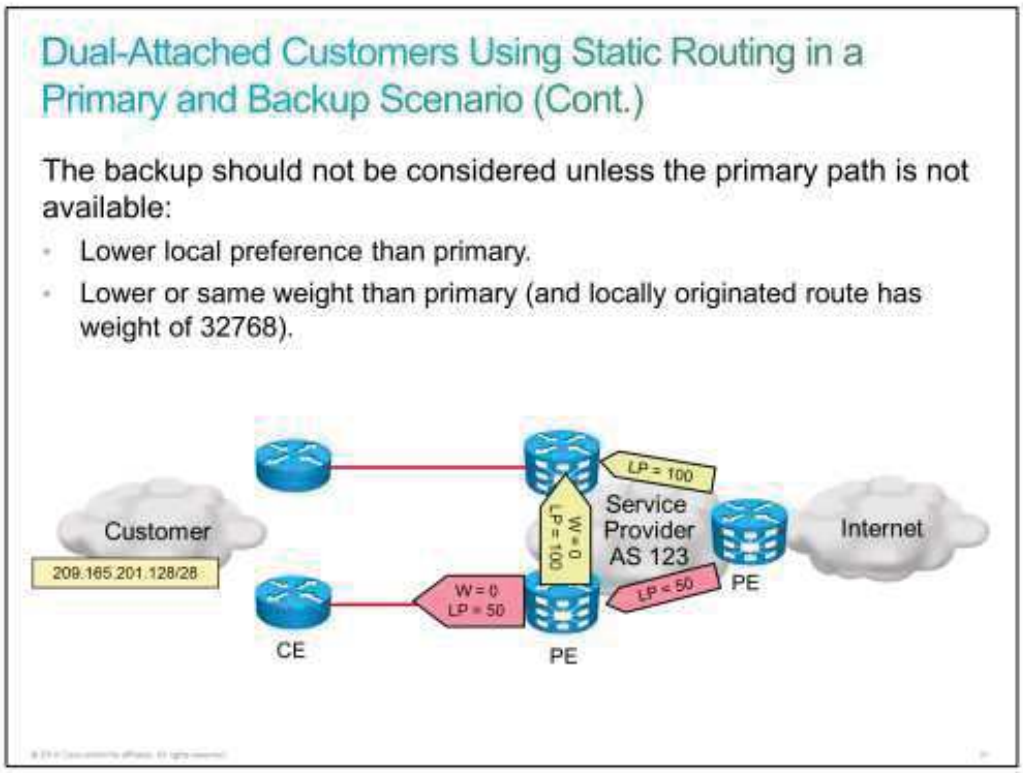
Redistribution of routes into any routing protocol is conditioned by the appearance of the route in the routing table. Thus, if the interface goes down, the router removes the static route from its routing table, and the route is withdrawn from the routing protocol. When the static route reappears, the redistribution process inserts the static route into the routing protocol again.

The backup CE router also uses a default static route toward the service provider via the backup link. The backup CE router is also redistributing the default route into the IGP. However, the static route that is used is a floating static route, which is assigned a high administrative distance—higher than the administrative distance of the customer IGP. As long as the primary link works, the IGP provides the backup CE router with the primary default route. Because of the higher administrative distance, the backup static default route is not installed into the backup router routing table. Because the static route is not in the routing table, it is not redistributed. If the primary link fails, the IGP no longer feeds the backup CE router with a default route. The backup static default route is the only remaining default route. Therefore, the router will install the floating default route into its routing table and then redistribute it into the IGP.

On the PE routers, static routes for the customer network should be configured and redistributed into BGP. On the backup PE router, the floating static route should be configured and have an administrative distance that is set to a higher number than any other routing protocol (for example, 250). Both static routes should be tagged. However, they should be tagged with different tags because a separate BGP routing policy is needed for the primary and backup routes.

https://t.me/learningnets

Similarly, then, in the previous scenarios, the service provider should summarize provider-assigned subnets before sending the routes to the rest of the Internet. To make sure that individual customer networks are not propagated outside of the service provider, the routes are marked with no-export community when redistributed into BGP.



The example shows why the backup route needs a separate BGP routing policy.

Static routes that are configured on the PE routers are redistributed into BGP. The backup PE router learns two routes for the customer network through BGP: one from the primary PE router over the IBGP session, and another that is locally originated from redistribution. The locally originated route will always take precedence because the weight attribute is set to 32768. (The weight for other routes is set to 0 by default.)

Based on BGP route selection rules, the redistributed floating static route will always remain the preferred path if additional BGP configuration is not performed on the PE router. This preference means that regardless of whether the primary link is available, the backup router selects the locally sourced route as the best route. Therefore, the backup router continues to announce a path toward the customer network.

To solve this problem, you will need to take additional configuration steps to ensure that the BGP route selection algorithm selects the primary route as the best BGP route when both routes are available:

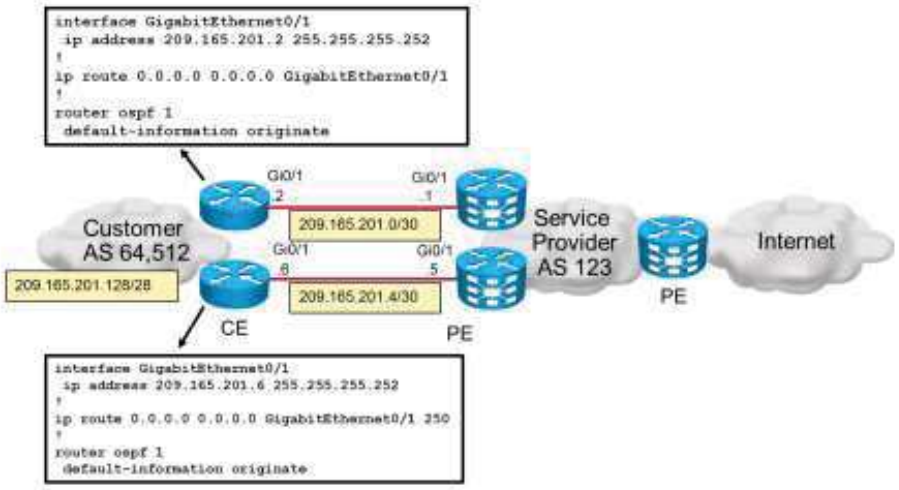
- When a router redistributes a floating static route into BGP, the weight value that is assigned to the floating static route must be set to 0. Otherwise, the floating static route will always be selected as the best BGP.
- Local preference values must also be assigned by the router to the floating static route so that the floating static route has a lower local preference than the primary route. This assignment ensures that the primary route is selected as the best BGP on other routers within the P-network.

Because the static routes require different BGP attribute manipulation, the static routes should be tagged with different tags. One tag should be used for primary routes that do not need any BGP manipulation. Another tag should be used for backup routes, where weight and local preference should be set by using RPL.

https://t.me/learningnets

Dual-Attached Customers Using Static Routing in a Primary and Backup Scenario (Cont.)

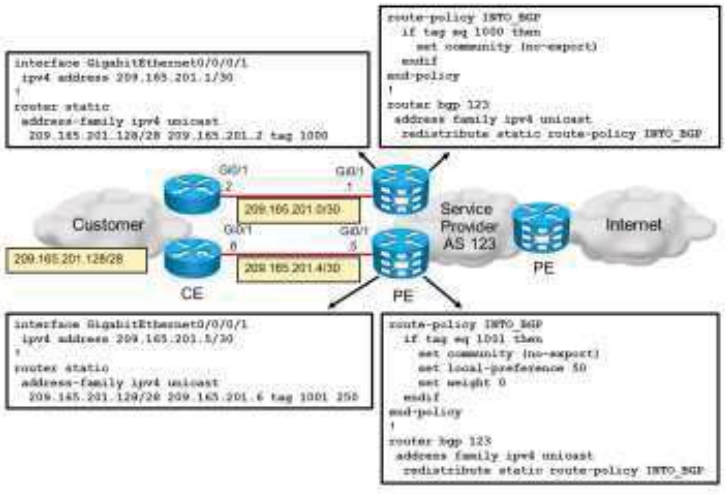
Configuration of CE:



The CE routers, which run Cisco IOS Software, are configured with static default route, which points to the PE router. The default routes are also redistributed into OSPF, which is used in the customer network. The static route on the backup CE router is configured as floating static route, which has the administrative distance set to 250.

Dual-Attached Customers Using Static Routing in a Primary and Backup Scenario (Cont.)

Configuration of PE:



https://t.me/learningnets

The PE routers, which run Cisco IOS XR Software, are configured with BGP. An IBGP neighbor relationship is established between the PE routers and other routers in the P-network. (The configuration is not shown.) On the primary PE router, a static route for the customer network is configured, which points to the CE router. On the secondary PE router, a floating static route for the customer network is configured, which points to the CE router. The floating static route has the administrative distance set to 250. The static route on the primary PE router is tagged with tag 1000. The BGP no-export community is then added to the route, and the route is redistributed into BGP. The static route on the backup PE router is tagged with tag 1001. The route is then matched, and no-export community is added to the route. The local preference is changed to 50, and the weight is set to 0 when the route is redistributed into BGP.

Configuration of the backup PE router, running Cisco IOS and IOS XE Software, is as follows:

```
interface GigabitEthernet0/1
 ip address 209.165.201.129/30
!
ip route 209.165.201.128 255.255.255.240 209.165.201.6 tag 1001 250
!
route-map INTO_BGP permit 10
  match tag 1001
  set community no-export
  set weight 0
  set local-preference 50
!
router bgp 123
 redistribute static route-map INTO_BGP
 neighbor 209.165.201.10 send-community
```

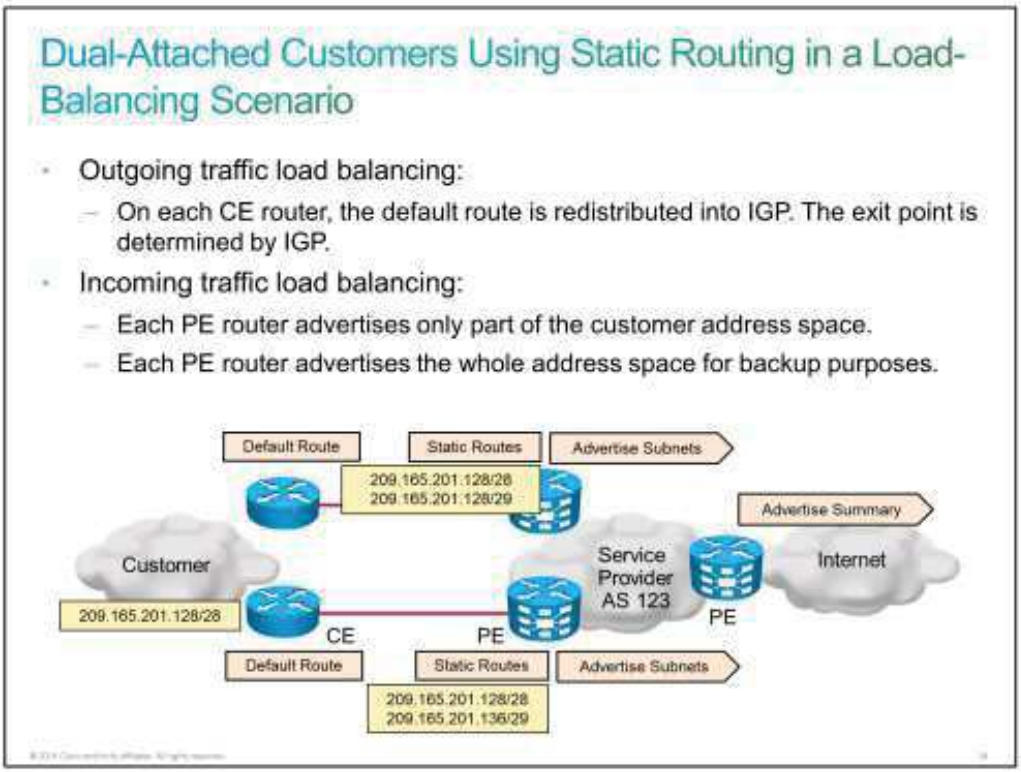
Note	For a complete command reference, refer to the Cisco IOS, IOS XE, and IOS XR Software Command Reference guides on www.cisco.com.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Dual-Attached Customers Using Static Routing in a Load-Balancing Scenario

A customer that is dual-attached to a service provider can also decide to balance traffic over both links. If both links are active, traffic should be balanced across both links. If there is link failure, one link should care for all traffic.



Load sharing of outgoing customer traffic is accomplished by configuring a standard default static route on both CE routers. Each static route is valid as long as the link from the CE router to the PE router is up. When both static routes are valid, both CE routers announce the default route into the customer network.

The remaining routers in the customer network see two candidate gateways of last resort. These remaining routers choose the closest one, with respect to the IGP metric. The part of the network that is closer to the uppermost exit point uses that exit point for all outgoing traffic. The other part of the network uses the other (lower) exit point.

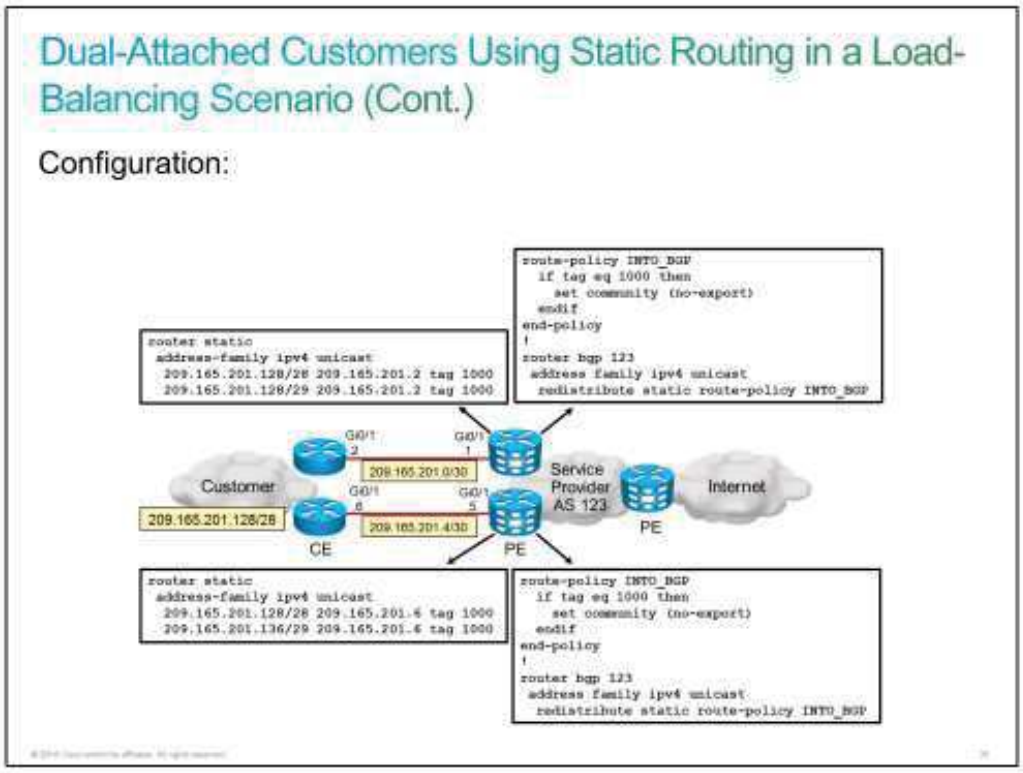
If both exit points are colocated, they are equally distant from each of the other routers in the customer network. Each router within the customer network therefore uses load sharing of traffic that is sent out of both exit points.

Service provider routers know about two routes to the customer network. Those routes are learned through BGP. However, BGP was designed to select only one, which is the best route from a collection of routes to the destination. Because of this, BGP cannot select more than one route to the customer network, although two are available. But you can achieve load balancing of incoming traffic by advertising different parts of the customer network with different PE routers. The upper PE router could advertise half of the address space, and the lower PE router could advertise the other half. For backup reasons, the PE routers also should advertise the entire address space as a larger route summary.

https://t.me/learningnets

As long as both paths are available, the traffic from the service provider to the customer uses the most explicit route. In this case, two explicit routes are used to send traffic representing one half of the address space over one link and traffic representing the other half of the address space over the other link.

Load sharing in this way does not result in an equal load on the links, but rather a statistically based distribution of the traffic load over the links.



Configuration of the CE routers is not shown in the example because it does not significantly differ from the configuration in previous scenarios.

The PE routers, which run Cisco IOS XR Software, are configured with BGP. An IBGP neighbor relationship is established between the PE routers and other routers in the P-network. (The configuration is not shown.) On each PE router, a static route for half of the customer network (on upper PE router: 209.165.201.128/29, on lower PE router: 209.165.201.136/29) is configured. For backup purposes, each PE router is configured with a static route for the complete customer network (209.165.201.128/28). The static routes on the PE routers are tagged with tag 1000. The route is then matched, the no-export community is added to the route, and the route is redistributed into BGP.

Configuration of the lower PE router, running Cisco IOS and IOS XE Software, is as follows:

```
interface GigabitEthernet0/1
 ip address 209.165.201.129/30
!
ip route 209.165.201.136 255.255.255.248 209.165.201.6 tag 1000
ip route 209.165.201.128 255.255.255.240 209.165.201.6 tag 1000
!
route-map INTO_BGP permit 10
  match tag 1000
!
router bgp 123
 redistribute static route-map INTO_BGP
 neighbor 209.165.201.10 send-community
```

 ©2014 Cisco Systems, Inc.

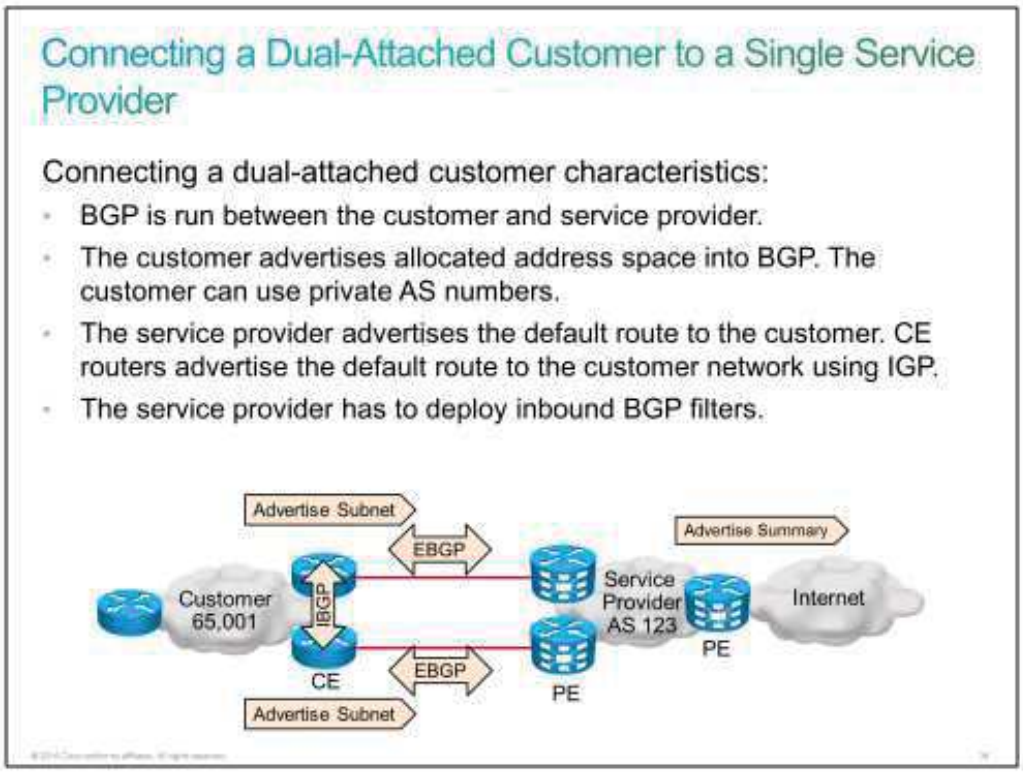
https://t.me/learningnets

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

https://t.me/learningnets

Connecting a Dual-Attached Customer to a Single Service Provider

Another routing option when a customer is connected to a service provider using two links is using BGP. In the example, the customer is connected to the P-network using multiple permanent links. BGP is used to exchange routing information between the customer and service provider.



Selecting BGP as the routing protocol between the customer and P-network ensures that a link failure or the failure of a remote router is detected. In this scenario, the customer does not require the use of a public AS number or full Internet routing. Instead, a private AS number is assigned to the customer network, and the service provider sends a default route to the customer through BGP.

The significant difference in this case, compared to a network scenario where static routes and redistribution are used, is that routers within the private AS of the customer now advertise the customer routes via BGP. Thus, the customer is responsible for announcing its own address space. The service provider receives routes from the customer and conditionally propagates them (similar to static routing). If the customer uses PA address space and the service provider can summarize the address space, the service provider will not propagate the explicit routes from the customer to the Internet. The private AS number in the AS-path attribute must be removed before the service provider can propagate any of the customer routes.

Because the customer is now creating BGP routes that are received by the service provider, any error that is made by the customer can influence routing operation within the P-network and, if propagated, within the Internet as a whole. Announcing a route to a network to which the customer has not been assigned may cause routing problems. There is always a risk that these routing problems can occur in a P-network. However, the risk is much greater when the customer, whose network administrators usually have less experience with BGP, enters the configuration.

 ©2014 Cisco Systems, Inc.

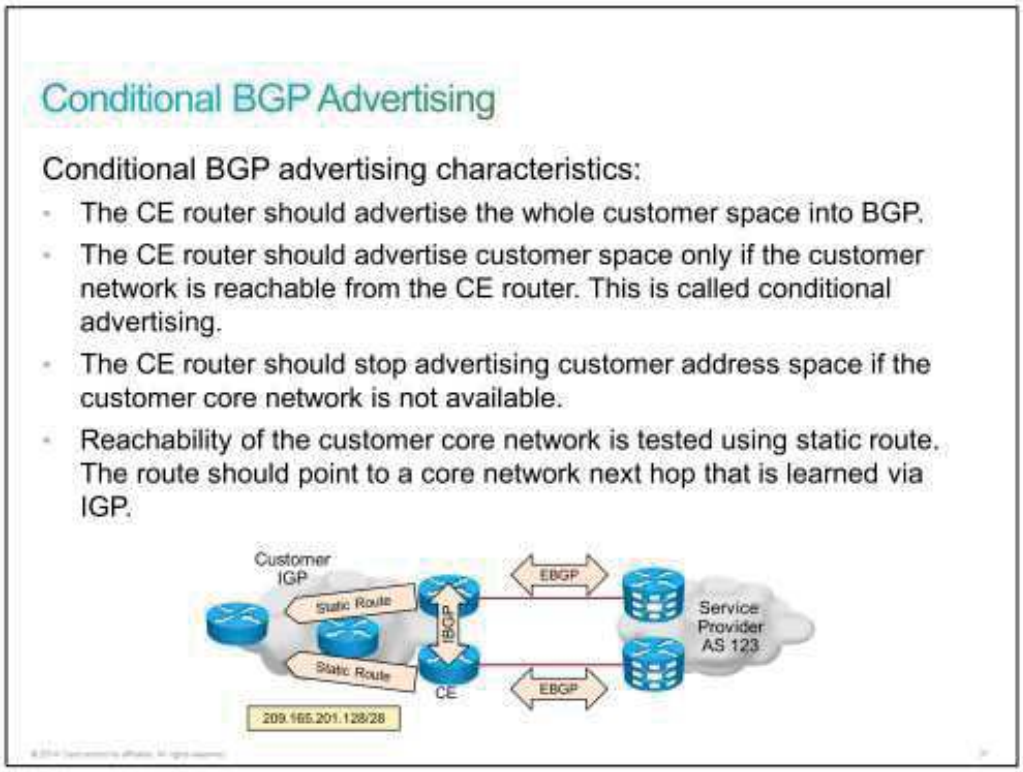
https://t.me/learningnets

To reduce the risk of erroneous route advertising, the service provider should always filter any BGP information that it has received from the customer network. The service provider should reject routes to networks that are not expected to be in the customer AS. Routes that contain an AS path with unexpected AS numbers should also be rejected.

https://t.me/learningnets

Conditional BGP Advertising

The customer should announce addresses as large as possible to the service provider (that is, the larger the address space that can be aggregated, the better). The BGP advertisement is configured on the CE routers using the **network** command. The route advertisement is conditioned by the appearance of a corresponding network or subnet in the routing table of the edge router. Because there is a chance that the customer subnetted the address space into smaller subnets, the complete customer address space will not be presented in the CE routing table. The solution is to manually enter the route into the routing table by configuring a static route to null 0. However, the static route is always present in the routing table, and the BGP advertisement is always performed from the CE router, even if the CE router cannot access the rest of the customer network.



If the CE router loses connectivity to the rest of the customer network but is still connected to the P-network, the BGP advertisement must cease. In this case, the BGP advertisement can be stopped if BGP advertisements are bound to the reachability status of a specific subnet in the core of the customer network, according to the customer IGP.

The problem with using a static route to null 0 is that it conditions the network statement in the BGP configuration so that BGP always advertises the route. If the CE router loses connectivity with the rest of the customer network, the router continues to advertise the entire customer address space. The P-network receives a valid route from the CE router. Traffic is sent to this router, but because the router has lost connectivity with the rest of the network, the traffic is dropped (or routed to the null 0 interface using the static route).

 ©2014 Cisco Systems, Inc.

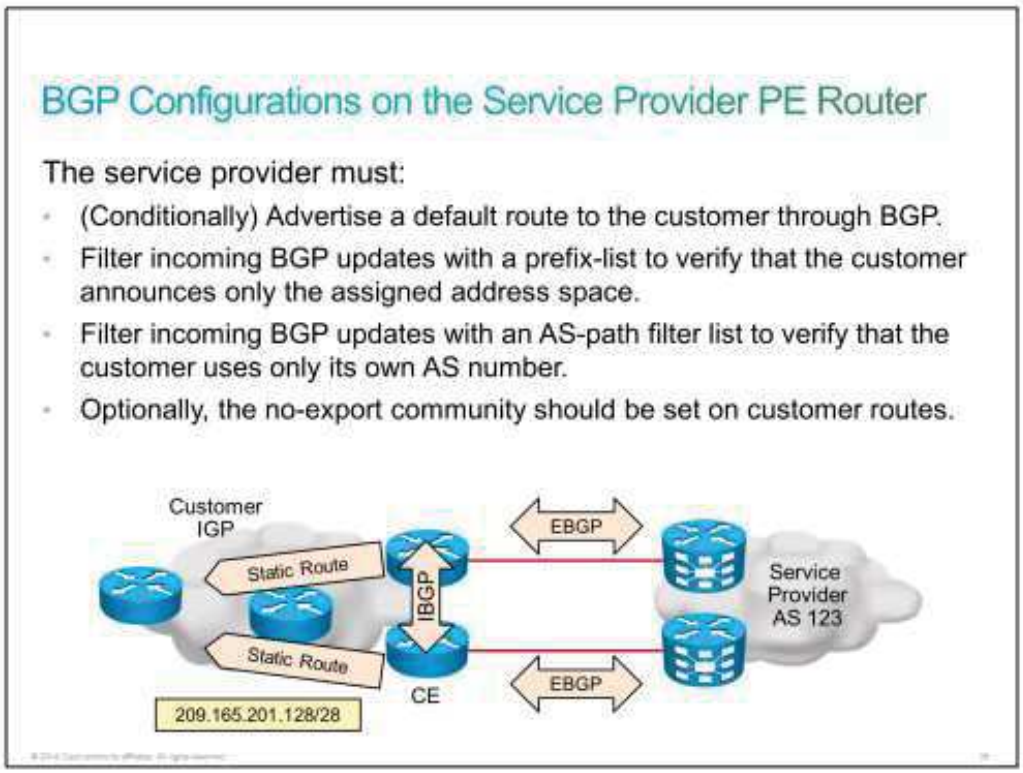
https://t.me/learningnets

The solution is to create a static route for the customer-assigned addresses, which points to a next hop that was learned through an IGP. The static route will be in the routing table only if the next hop is reachable. The static route would be valid and the route would be advertised through BGP. If the CE router loses connectivity from the rest of the customer network, the next hop will become unavailable and the static route will not be valid anymore. Because the customer route will not be present in the routing table anymore, BGP will cease to advertise the customer route to the service provider. The customer network will then be accessible over the backup connection, if available.

https://t.me/learningnets

BGP Configurations on the Service Provider PE Router

In the P-network, the two PE routers must have BGP sessions with the customer. There is no point in feeding the full Internet routing table to the customer because the table contains the same set of routes for both links and the customer always uses the service provider for all traffic toward the Internet. Injection of a default route in the customer network would accomplish the same task.



The customer is responsible for its own advertisements. Because customers are much less likely to be experienced in BGP configuration than the service provider, they are more likely to make errors. Therefore, the ISP must protect itself and the rest of the Internet from those errors.

The service provider should use a filter that allows only customer-assigned routes and denies any other route to ensure that private address space or any other illegal networks that are erroneously announced by the customer never reach the service provider BGP table. Filtering based on the AS path also provides some protection from customer configuration errors. Only routes that originated within the customer AS should be allowed.

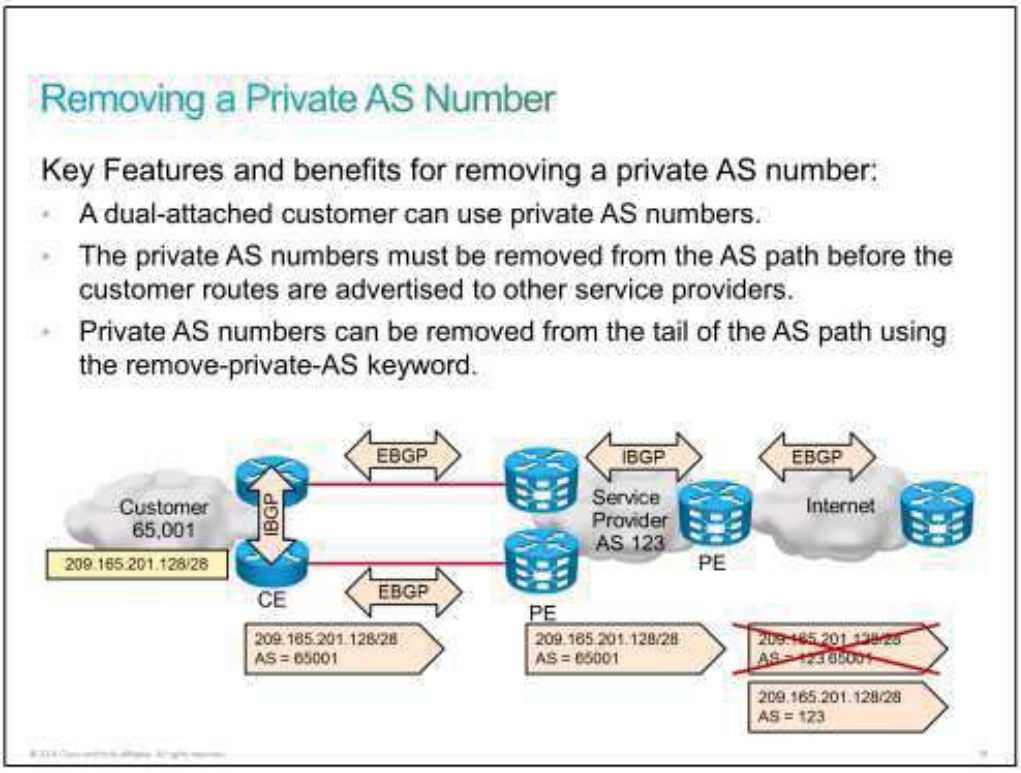
If the customer address space is PA address space and represents only a small part of a larger block that is announced by the ISP, the explicit BGP routes that are received from the customer need not be advertised to the rest of the Internet. The service provider can announce the large block, attracting any traffic toward any subnet within the block. After the traffic enters the P-network, the more explicit routes to the customer network are available and used. In this case, the PE router can tag the BGP routes that are received from the customer with the no-export, well-known community, restricting them from being sent by the service provider to any other AS.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Removing a Private AS Number

Routes that are received by the service provider from the customer are propagated to the rest of the Internet only if they are part of the PI address space.



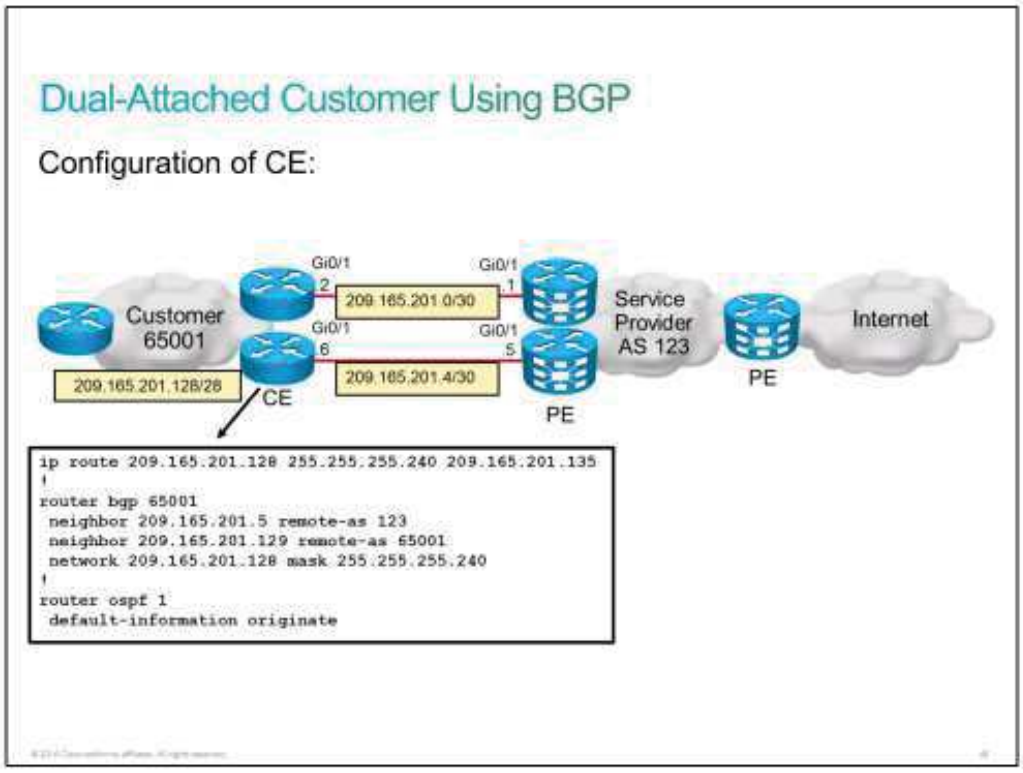
When the service provider receives BGP routes from the customer, the AS-path attribute of the received routes contains only the AS number of the customer. If the customer uses AS-path prepending, there may be several repetitions of the customer AS number in the AS path. If the customer routes are propagated by the service provider to the Internet, the AS number of the customer will be present in the AS path, unless it is explicitly removed. If the customer has been assigned a private AS number, this AS number must never be advertised by any router to the rest of the Internet, and the service provider has to make sure that the private AS number is removed from the BGP route and that the service provider AS number is present in the tail of the AS path.

Removal of a private AS number from the AS path is accomplished by using the **remove-private-as** command on the service provider EBGP sessions with the rest of the Internet. In the figure, removal of the private AS number occurs on the EBGP session between AS number 123 and the Internet.

https://t.me/learningnets

Dual-Attached Customers Using BGP

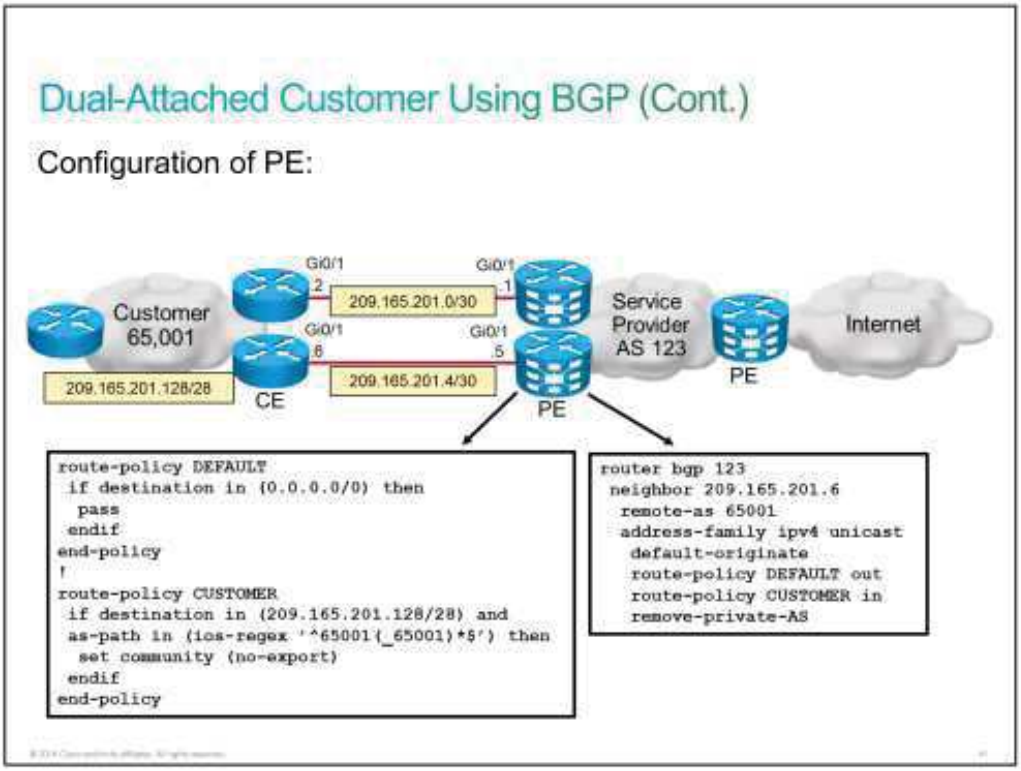
The CE routers, which run Cisco IOS Software, are configured with static route for the customer network. The PE routers, which run Cisco IOS XR Software, are configured with BGP.



The static route points to a next hop, which is learned through IGP. The CE router is configured with an IBGP session with another CE router and configured with an EBGP session with a PE router. The CE router advertises the customer network using the **network** command. The CE router will advertise the customer network only if the static route for the network is valid. (The next hop, learned through the IGP, should be available.) Because the CE router will receive the default route from the PE router through BGP, the default route is also propagated into the customer network through the IGP (or OSPF in the example).

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets



An IBGP neighbor relationship is established between the PE routers and other routers in the P-network. (The configuration is not shown.) The PE router is configured to announce the default route to the customer using the **default-originate** command. The PE router is also configured with two route policies. The DEFAULT policy allows only the default route to be sent to the CE router. The CUSTOMER policy allows only the customer route with 65001 in the AS path to be received from the customer. The service provider allows more than one occurrence of the customer AS in the AS path if the customer performs AS-path prepending. The PE router connecting to the Internet is configured to remove any private AS number that occurs in the tail of the AS path of a BGP route.

Configuration of the lower PE router, running Cisco IOS and IOS XE Software, is as follows:

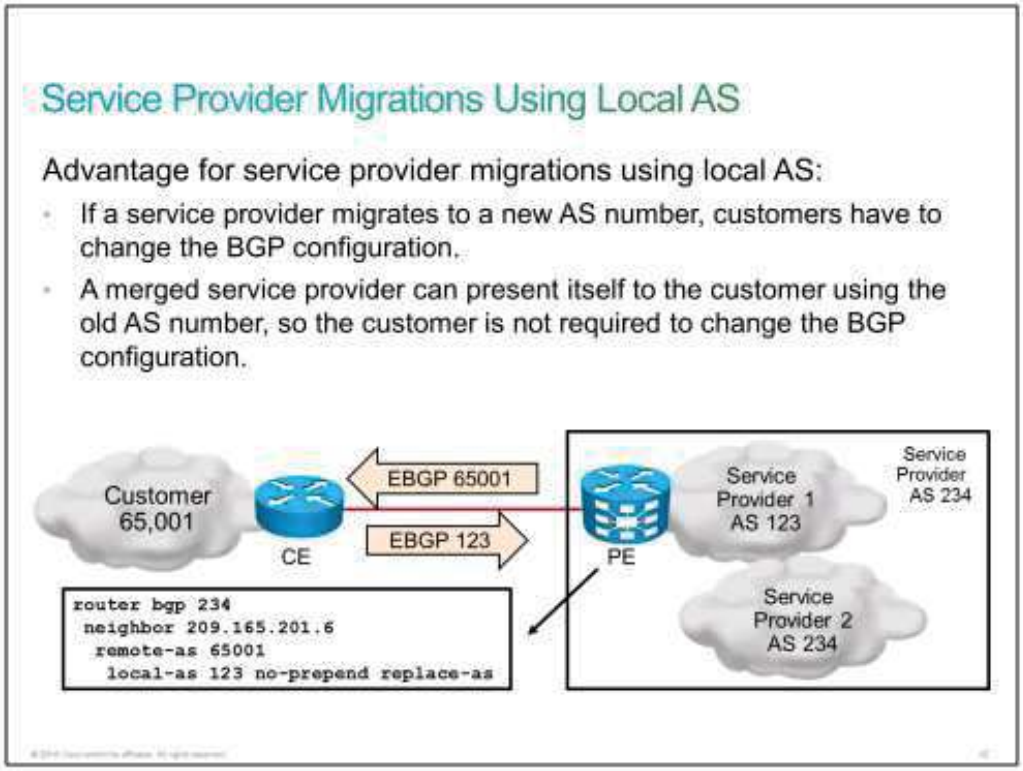
```
ip as-path access-list 15 permit ^65001(_65001)$
ip prefix-list DEFAULT permit 0.0.0.0/0
ip prefix-list CUSTOMER permit 209.165.201.128/28
!
route-map CUSTOMER_MAP permit 10
  match ip address prefix-list CUSTOMER
  set community no-export
!
router bgp 123
 neighbor 209.165.201.6 remote-as 65001
 neighbor 209.165.201.6 send-community
 neighbor 209.165.201.6 filter-list 15 in
 neighbor 209.165.201.6 prefix-list DEFAULT out
 neighbor 209.165.201.6 route-map CUSTOMER_MAP in
```

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

Service Provider Migrations Using Local AS

In the scenario when the customer connects to the service provider with AS number 123. Another service provider with AS number 234 buys the first service provider and wants to migrate complete BGP configuration to AS number 234. All customers would then be required to change BGP configuration on their routers as well. Another solution is to use the local AS feature that is available on BGP routers. Using the local AS feature, the PE router in AS number 234 can present itself as a BGP router from AS number 123. In this case, the customer is not required to change BGP configuration on its routers.



The figure also shows configuration on the PE router. The BGP routers using the local AS feature prepend the local AS number in inbound EBGP updates. Because it is not desired to have local AS numbers in the AS path inside the P-network and further in the Internet, the **no-prepend** option is used. This option instructs the router not to prepend the local AS number to outgoing updates. BGP routers using the local AS feature also prepend both the actual AS number and local AS number in outbound EBGP updates. Because it is desired that the customer sees BGP routes as coming from the service provider with AS number 123, the **replace-as** option is used. This option instructs the PE router not to prepend the local AS number to the updates received from the customer.

Configuration of the PE router, running Cisco IOS and IOS XE Software, is as follows:

```
router bgp 234
 neighbor 209.165.201.6 remote-as 65001
 neighbor 209.165.201.6 local-as 123 no-prepend replace-as
```

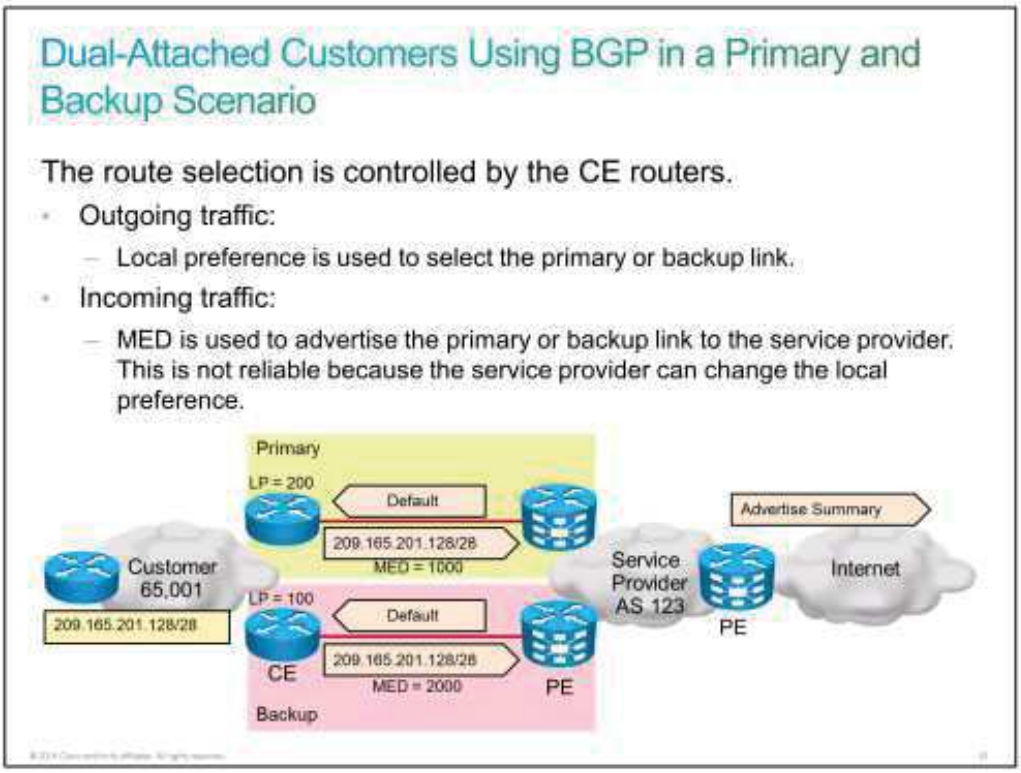
Note	For a complete command reference, refer to the Cisco IOS, IOS XE, and IOS XR Software Command Reference guides on www.cisco.com.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Dual-Attached Customers Using BGP in a Primary and Backup Scenario

When a customer uses BGP on multiple links between its network and the P-network, the customer is solely responsible for controlling how it uses the links. The customer can choose to use its links in a primary or backup scenario or in a load-sharing scenario.



If one link is primary, then the other link should be used for backup only. The customer can use the local preference configuration to direct all outgoing traffic over the primary link.

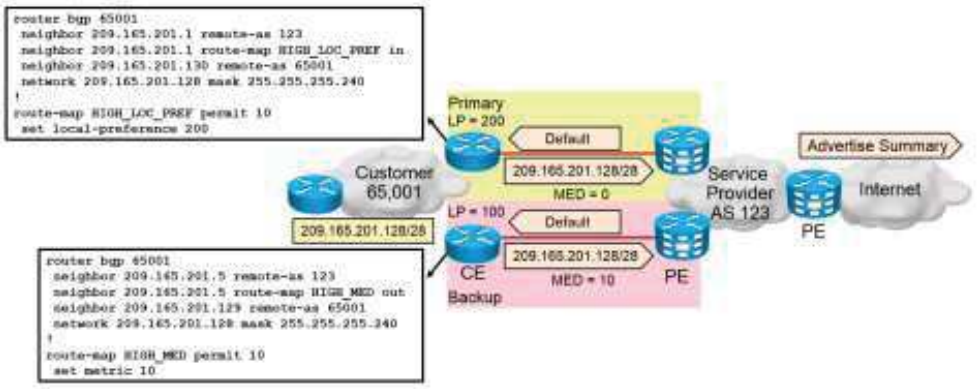
Incoming traffic to the customer is controlled by using either AS-path prepending or the MED. Because the customer has multiple connections to the same AS, the MED is the ideal attribute to use. When the customer announces its routes to the service provider, a worse (or high) MED value on the backup link and a good (or low) value on the primary link are set.

The MED and AS-path length are checked by the receiving EBGP peer only if the weight and local preference attributes have not been configured. In this case, the service provider should not use any of these configuration options. The service provider should rely solely on the attributes that it has received from the customer. However, if the service provider sets, for example, the local preference for customer routes, the service provider will control path selection for traffic to the customer. This path selection is why you cannot rely on the fact that your traffic distribution scenario will work as intended.

https://t.me/learningnets

Dual-Attached Customers Using BGP in a Primary and Backup Scenario (Cont.)

Configuration:



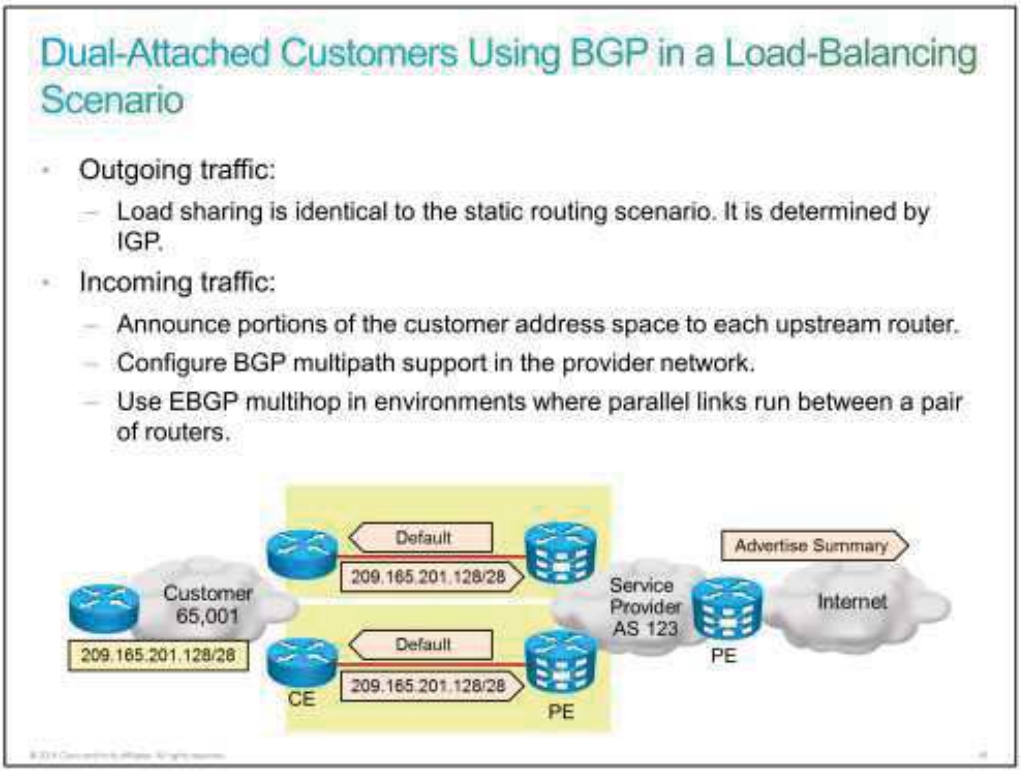
The CE routers, which run Cisco IOS Software, are configured with an IBGP session with another CE router and configured with an EBGP session with a PE router. The CE router advertises the customer network using the **network** command. The primary CE router sets the local preference of received routes from the service provider to 200 using a route map. This makes the primary router the preferred one. Recall that the default local preference on the backup CE router is 100. The secondary router advertises the customer network to the service provider with a MED of 10. This makes the primary router the preferred one from the service provider point of view. Recall that the default MED on the primary CE router is set to 0, which is better than 10.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Dual-Attached Customers Using BGP in a Load-Balancing Scenario

Load sharing of outgoing traffic from the customer network is identical to the static routing scenario. The customer IGP is configured to send information about a gateway of last resort. There is no difference whether the edge router gets its default by static routing or by incoming EBGP updates.



Load sharing of the return traffic coming back to the customer network from the service provider can be implemented in a number of ways:

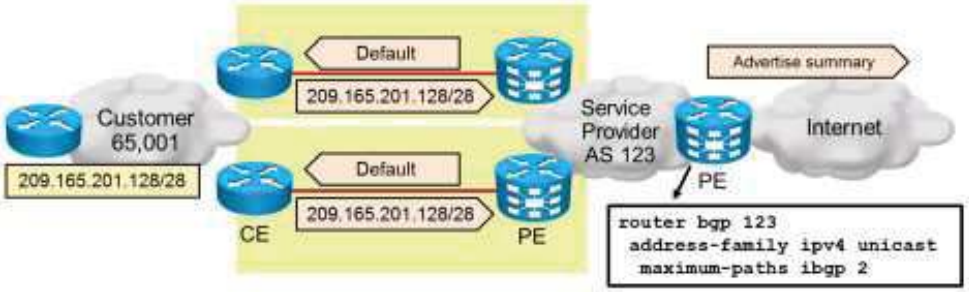
- The customer can divide its address space into several announcements. The CE router can send each announcement over one of its EBGP sessions with the ISP. For backup purposes, the customer should advertise the entire address space over all of its EBGP sessions. The service provider now uses the most explicit route rule, and as long as both links are up, traffic with destinations within one part of the customer address space is routed over one of the links, and traffic to the other part is routed over the other link.
- If the customer announces equivalent routes over both links as shown in the slide, the service provider routers use the closest connection with respect to the IGP of the service provider. If a service provider router has an equivalent distance to both connection points, the use of the **maximum-paths** (or BGP multipath) option causes load sharing.
- If the multiple links between the customer and the ISP network terminate in one single router on the customer side and one single router on the ISP side, the two routers must establish their EBGP session from loopback interface to loopback interface. Static routing is required for one router to get information on how to reach the loopback interface of the other router. The use of the **ebgp-multihop** option is also required because the address of the neighbor is not directly connected.

https://t.me/learningnets

Dual-Attached Customers Using BGP in a Load-Balancing Scenario (Cont.)

BGP Multipath:

- By default, BGP selects a single path as the best path and installs it in the IP routing table.
- If configured, a BGP router can select up to eight identical BGP routes as the best routes and install them in the IP routing table for load-sharing purposes.



By default, BGP route selection rules select one, and only one, route as the best. If there are two identical routes, the tiebreaker is either the most stable route or the router ID of the peer router that is advertising the route. However, when the **maximum-paths** router configuration command is used, the BGP route selection process will choose more than one route as best if they are identical. The routes are all installed in the routing table, and load sharing occurs. A router can use either only IBGP routes or only EBGP routes for load balancing but not both at the same time because this could lead to routing loops. BGP multipath can be enabled separately for IBGP and EBGP routes.

A BGP router running Cisco IOS XR Software can install up to eight BGP routes in the routing table. The actual type of load sharing is determined by Cisco Express Forwarding.

The Internet-facing PE router in the example is configured with IBGP multipath, where the maximum number of paths is set to 2. Because an IBGP session is established between all PE routers, IBGP multipath is enabled. The Internet-facing PE router then can use both customer-facing PE routers to send traffic to the customer.

Configuration of the PE router, running Cisco IOS and IOS XE Software, is as follows:

```
router bgp 123
 maximum-paths ibgp 2
```

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

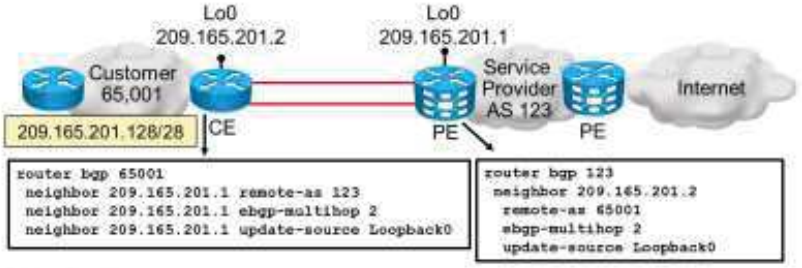
 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Dual-Attached Customers Using BGP in a Load-Balancing Scenario (Cont.)

EBGP Multihop:

- Used when parallel links between routers exist. In this case, loopback interfaces have to be used for BGP peering.
- Because of recursive lookup, load sharing toward a BGP destination always occurs if there are several equal-cost paths to the BGP next hop.
- By default, EBGP neighbors must be directly connected.
- The **ebgp-multihop** command declares an EBGP neighbor to be distant.



When two adjacent routers have multiple links between them, you can configure the EBGP session from loopback interface to loopback interface. In this case, you must use the **ebgp-multihop** option to make the BGP session go into the active state. There must be static or dynamic routing in use to provide both routers with information on how to reach the loopback interfaces of each other. Otherwise, their EBGP session does not complete establishment.

Routing to the loopback interface of the neighboring router is required to establish the EBGP session and is also used in the recursive lookup when the routes are installed by the router in its routing table. The two routes to the loopback interface of the neighboring router should be equivalent for load sharing to occur.

After configuration, one single EBGP session is established between the two routers. This session is used to exchange the routing information. There is only one BGP route to each destination, and it has a next hop that refers to the loopback interface of the other router.

The load balancing is an automatic as the result of the recursive route lookup of BGP next hops. The IP routing table will contain a single routing entry to reach the customer prefix but the Cisco Express Forwarding table (which contains the results of the recursive lookup) will have two equal-cost paths to reach the customer prefix.

By default, EBGP neighbors must be directly connected. A router verifies that an EBGP neighbor is reachable as directly connected over one of the router interfaces before the session goes into the active state. For an EBGP session, IP packets that carry the TCP segments with BGP information are also sent using a TTL value set to 1. This value means that they cannot be routed.

The **ebgp-multihop** neighbor configuration command changes this behavior. Although the neighbor is several hops away, the session goes into the active state, and packets start to be exchanged. The TTL value of the IP packets is set to a value larger than 1. If no value is specified on the command line, then 255 is used.

Use the **ebgp-multihop** command when you are establishing EBGP sessions between loopback interfaces for load-sharing purposes. You must take great care when using **ebgp-multihop** because proper packet forwarding relies on all of the intermediate routers along the path to the EBGP peer to make the correct forwarding decision. If the intermediate routers have a correct path to the EBGP peer but a wrong path to the final destination, the packet may get into a routing loop.

https://t.me/learningnets

Both routers in the example are configured with BGP peering between loopback interfaces. The **ebgp-mutlihop** option specifies that the peer is more than one hop away, and the **update-source** command specifies that the loopback interface should be used to originate BGP traffic. Configuration of static routes that instruct a router how to reach another router loopback interface is not shown in the example.

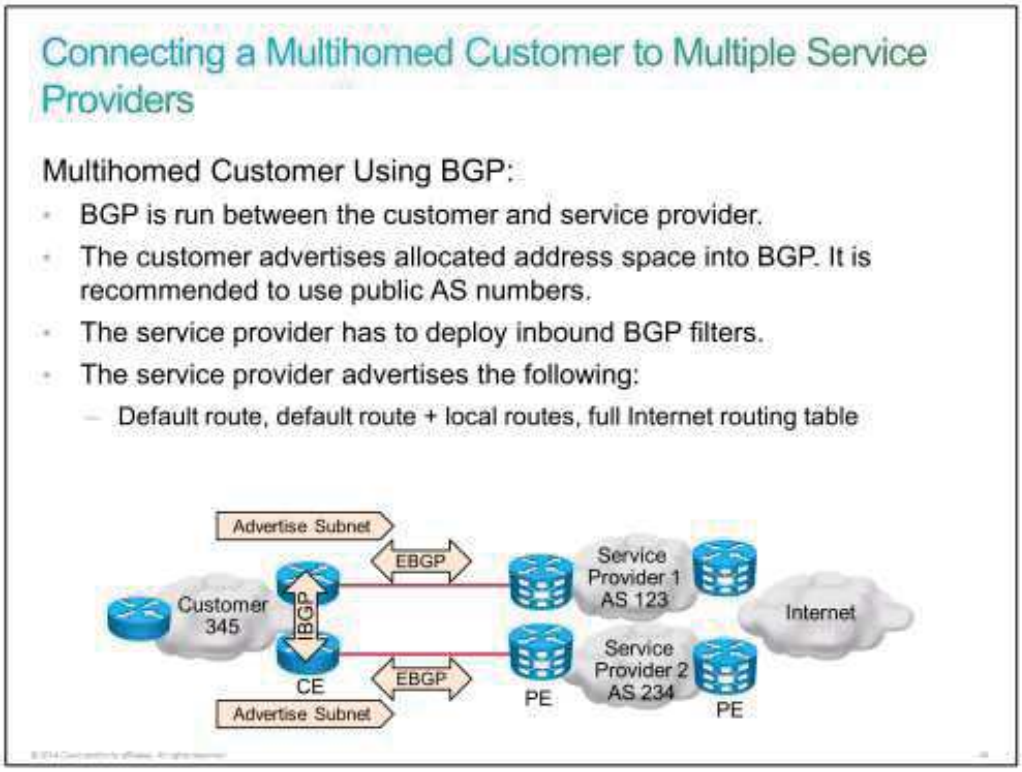
Note	For a complete command reference, refer to the Cisco IOS, IOS XE, and IOS XR Software Command Reference guides on www.cisco.com.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Connecting a Multihomed Customer to Multiple Service Providers

The highest level of redundancy to network failures is achieved in network designs that connect the customer network to two different P-networks. Customers use this option when the requirement for resilient Internet connectivity is very high. This requirement also involves duplication of equipment to make the customer network fully redundant.



Because the customer has to use PI address space, it is the responsibility of the customer to advertise the customer address space. The only routing option that can be used in this scenario is BGP because service providers run only BGP with customers. It is not enough to detect link failures or a failure in the remote router by link-level procedures. Failures that occur beyond the directly connected router must also be detected, and the only means of detecting these failures is by using a routing protocol. The only routing protocol that is suited for the Internet environment is BGP. Correctly configured, BGP manages rerouting in the following situations:

- Link failure between the customer network and the network of one of the service providers.
- Edge router failure on either the customer or the service provider side.
- Link failure or router failure within the customer network that causes the CE router to lose connectivity with the customer network core. This situation requires correct configuration of the route advertisement as described in an earlier topic.
- Link failure or router failure within the P-network that causes the PE router to lose connectivity with the rest of the Internet.

https://t.me/learningnets

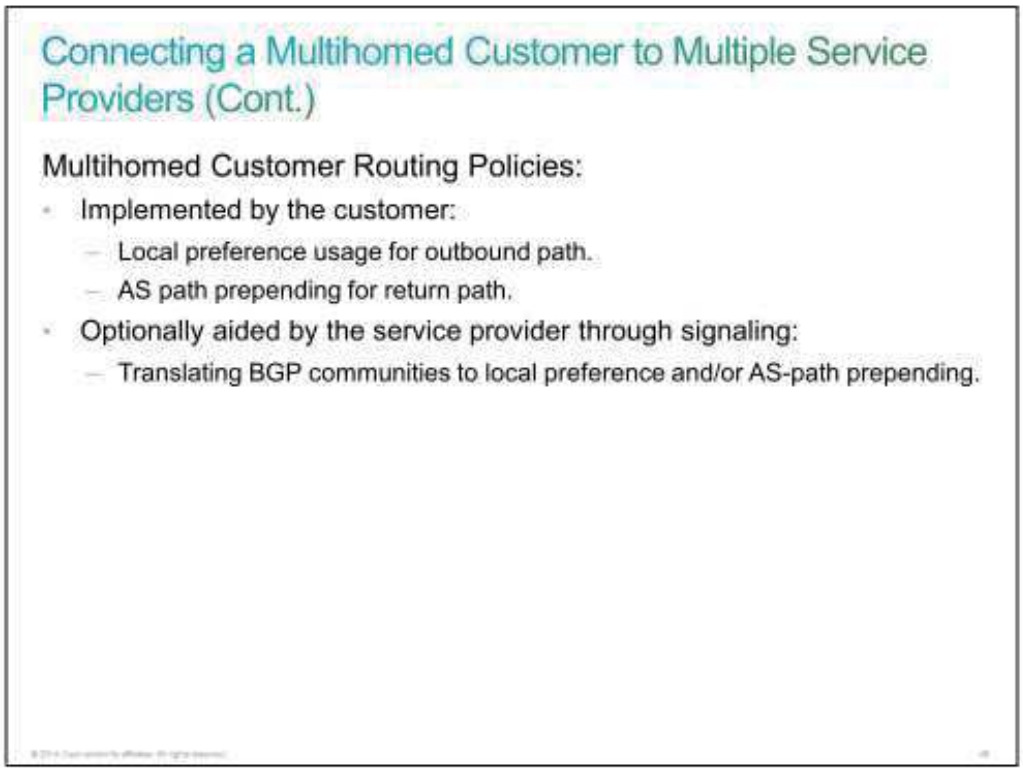
Multihomed customers have multiple permanent links to different service providers. The links should terminate in different edge routers in the customer network. Otherwise, one of the major advantages, resilience to router failure, is lost.

Multihomed customers should use BGP with two service providers. The customer should advertise its address space to two service providers. The route advertisement should be configured in both CE routers. The advertisement should be conditioned at the edge routers by the appropriate route policies leading toward the core of the customer network. This setup is analogous to that configured when you are connecting a dual-attached customer network to a single service provider.

The customer must use outgoing filters to prevent any route that is received from one of the ISPs from being propagated to the other. Otherwise, the customer network appears as a transit network between the two service providers.

Both service providers must apply filters on the incoming BGP information from the customer to protect themselves and the rest of the Internet from errors in the BGP configuration of the customer. Each of the service providers must accept routes from the customer that indicate networks within the customer address space only. AS-path filter lists should be implemented on the PE routers to allow incoming routes only if they have the correct AS-path attribute value. If the incoming filters on the ISP edge router accept customer routes, then the service provider should propagate those routes to the rest of the Internet.

Both service providers must provide the customer with at least some BGP routes. Depending on customer requirements, the volume of BGP routes that are provided by the ISP could range from the default route only to the default route with local routes as well and to the full Internet routing table.



Routing policies, which specify how traffic will be redistributed over both links to two service providers, can be implemented on the customer side. Optionally, routing policies also can be implemented by the service provider.

The customer can implement routing policies using the following:

- Changing local preference for the outbound path to influence path selection for outbound traffic. The customer increases local preference for routes received from the primary service provider to the preferred path through the primary ISP.

https://t.me/learningnets

- AS-path prepending to influence path selection for incoming traffic. The customer prepends the AS number to the routes advertised to the backup service provider to lengthen the AS path, thus making the path through the backup ISP less preferred.

Note	MED, as used in the previous lesson, cannot be used in the multihomed scenario. MED influences how traffic will flow into the local AS. However, MED is a nontransitive BGP attribute, which means that it is sent to the EBGP neighbors. These routers propagate MED inside their local AS, but they do not pass it onto the next AS (and to the Internet). This is why AS-path prepending has to be used instead of MED.

Routing policies also can be implemented on the service provider side. The service provider can set the local preference for incoming routing updates and AS-path prepending for outgoing updates (to the upstream AS) in two ways:

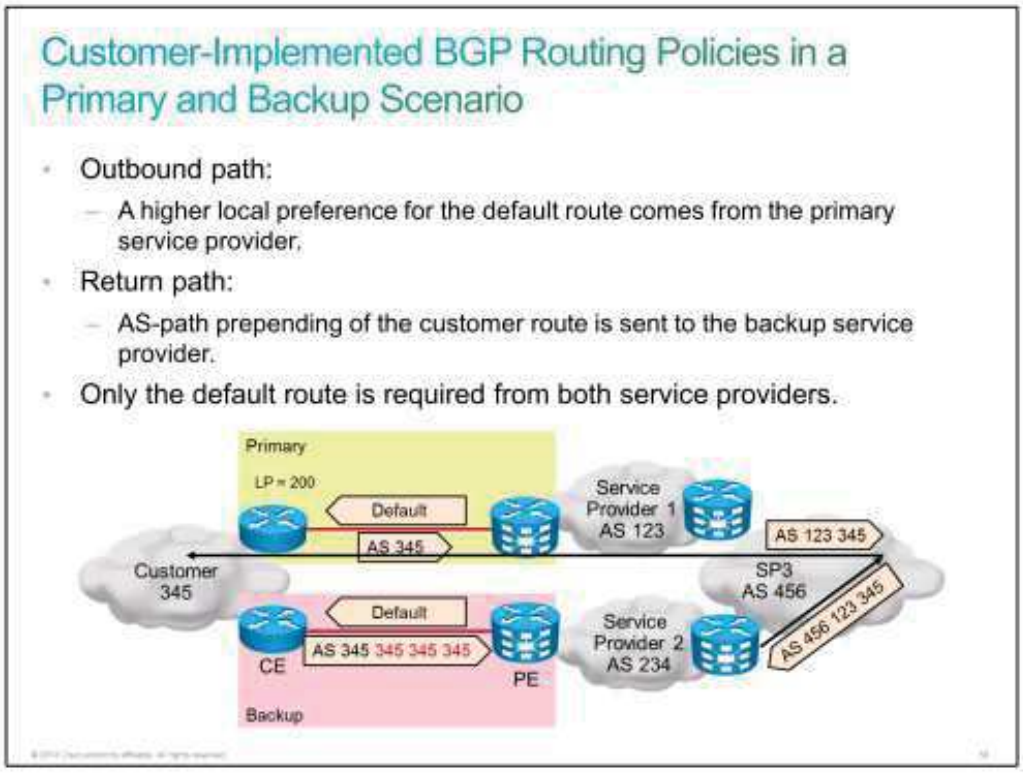
- **Manually**: The ISP changes the local preference and configures AS-path prepending manually for each customer route.
- **Through signaling**: The customer signals its needs using BGP communities. Each customer route is marked with BGP communities that are defined and published by the service provider. The service provider then matches those communities and sets the specific local preference and AS-path prepending.

Note	Manually implemented routing policies on the service provider side are not a scalable solution because the service provider has to manually configure route policies for each customer. Implementation of route policies on the service provider side using signaling is only discussed in the lesson.

https://t.me/learningnets

Customer-Implemented BGP Routing Policies in a Primary and Backup Scenario

In the primary and backup scenario, only the default route is required from each service provider. Routing policies are implemented by the customer.



The customer selects a primary service provider for outgoing traffic by setting a higher local preference (200) for routes received from the preferred ISP. Because the default local preference is set to 100 on the secondary CE router, the path through the primary service provider will be chosen.

The customer configures AS-path prepending of routes that are sent to the backup service provider to lengthen the AS path. Returning traffic originating from the Internet will therefore choose a shorter path through the primary service provider. In the example, the customer AS number is prepended to the AS path three times, thus making the customer four AS hops away from the backup service provider.

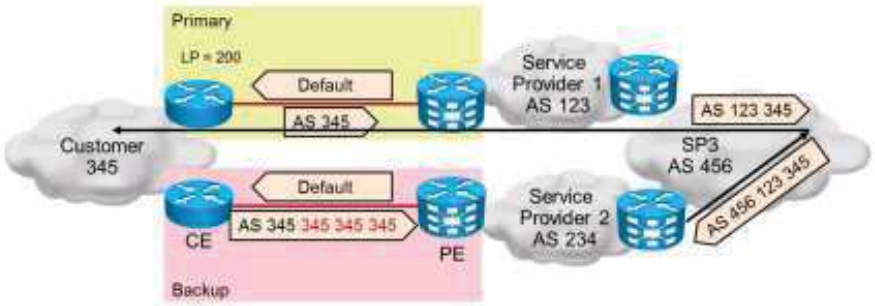
 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Note	A similar scenario is the primary or backup service provider with direct traffic to the secondary service provider. In this scenario, one service provider is used as the primary and another one is used as the backup. The only exception to this policy is traffic going to networks that are owned by the secondary service provider. In this scenario, the primary service provider sends only the default route to the customer, and the backup service provider sends the default route and service provider-owned routes. Routing policies are implemented as follows: higher local preference for the default route from the primary service provider, and AS-path prepending for routes sent to back up service providers. If the customer wanted to communicate with a customer that is connected to the secondary service provider, traffic would match a more specific route on the backup CE router, and traffic would be sent to the secondary service provider directly. For all other traffic, the default route on the primary CE would be used because it has a higher local preference than the default route on the secondary CE.

Customer-Implemented BGP Routing Policies in a Primary and Backup Scenario (Cont.)

- Outbound path:
  - A higher local preference for the default route comes from the primary service provider.
- Return path:
  - AS-path prepending of the customer route is sent to the backup service provider.
- Only the default route is required from both service providers.



The CE routers, which run Cisco IOS Software, are configured with an IBGP session with another CE router (not shown in the example) and an EBGP session with the PE router. The CE router advertises the customer network using the **network** command. The primary CE router sets the local preference of received routes from the primary service provider to 200 using a route map. This makes the primary router the preferred one. Recall that the default local preference on the backup CE router is 100. The secondary router advertises the customer network to the backup service provider with AS-path prepending. This makes the secondary router the least preferred one from the service provider point of view. In the example, the customer AS number would be prepended to the AS path three times, thus making the customer route four AS hops away from the secondary service provider. You have to make sure that the AS number is prepended enough times to make the customer route less preferred also from the backup P-network.

https://t.me/learningnets

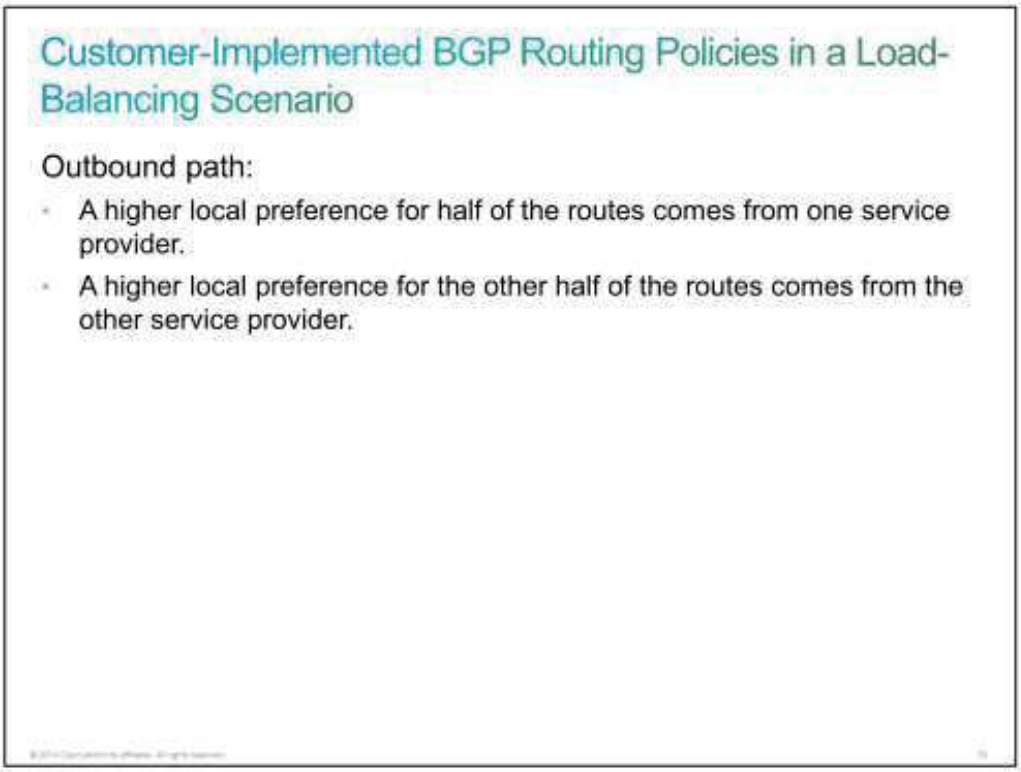
Note	For a complete command reference, refer to the Cisco IOS Software Command Reference guides on www.cisco.com.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Customer-Implemented BGP Routing Policies in a Load-Balancing Scenario

In this scenario, the customer connects to service providers with a load-balancing configuration. Both service providers will serve as the backup for one another.

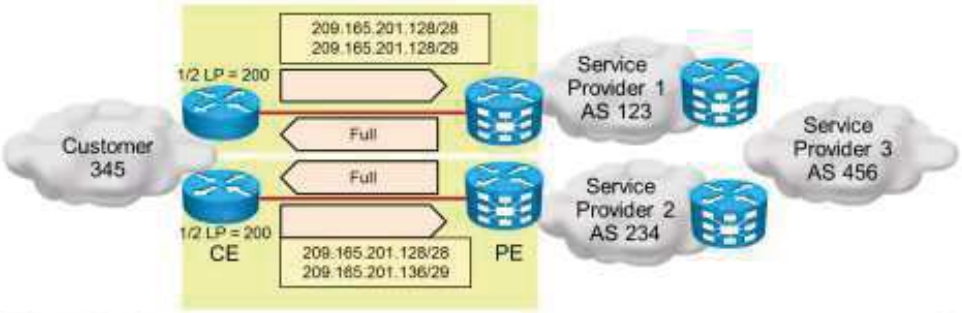


https://t.me/learningnets

Customer-Implemented BGP Routing Policies in a Load-Balancing Scenario (Cont.)

Return path:

- The customer announces portions of the customer address space to each service provider.
- A full routing table is required from both service providers.



In this scenario, complete routing tables should be advertised by each service provider. Load balancing for outgoing traffic can be achieved by setting a local preference to a higher value for half of the routes received from one service provider than for the other half of the routes received from the other service provider. Traffic statistics should be gathered to identify traffic patterns (such as how much traffic goes to specific destinations in the Internet). Based on this information, the customer can decide which routes need an increased local preference on each service provider.

For incoming traffic, load balancing can be configured by splitting the customer network into two subnets and advertising one subnet to one ISP and the other subnet to another ISP.

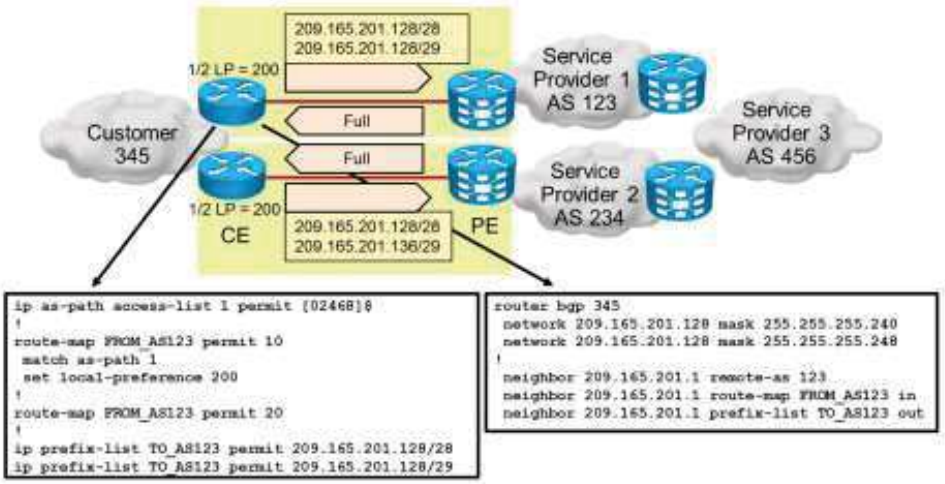
Two issues arise in this scenario:

- How do you equally split the complete Internet routing table into two halves?
- Will load balancing perform as expected? Will there be any asymmetric routing?

https://t.me/learningnets

Customer-Implemented BGP Routing Policies in a Load-Balancing Scenario (Cont.)

Configuration of CE:



The primary CE router, which runs Cisco IOS Software, is configured with an IBGP session with another CE router (not shown in the example) and an EBGP session with the primary PE router. The primary CE router sets the local preference to 200 for half of the networks received from the primary service provider. The example illustrates one of the ways that the complete Internet routing table can be split into two parts (not necessarily equal halves in the number of prefixes or in the resulting amount of load-balanced traffic). In the example, the prefixes that are received from the primary service provider will be assigned a local preference of 200 if they originate in an even-numbered AS (the AS-path attribute ends with 0, 2, 4, 6, or 8). Other prefixes will be assigned the default local preference of 100.

The load-balancing configuration for incoming traffic is achieved by splitting the customer address space into two halves and by advertising one half (209.165.201.128/29) to the primary service provider. Together, this will complete the customer network (209.165.201.128/28) for backup purposes.

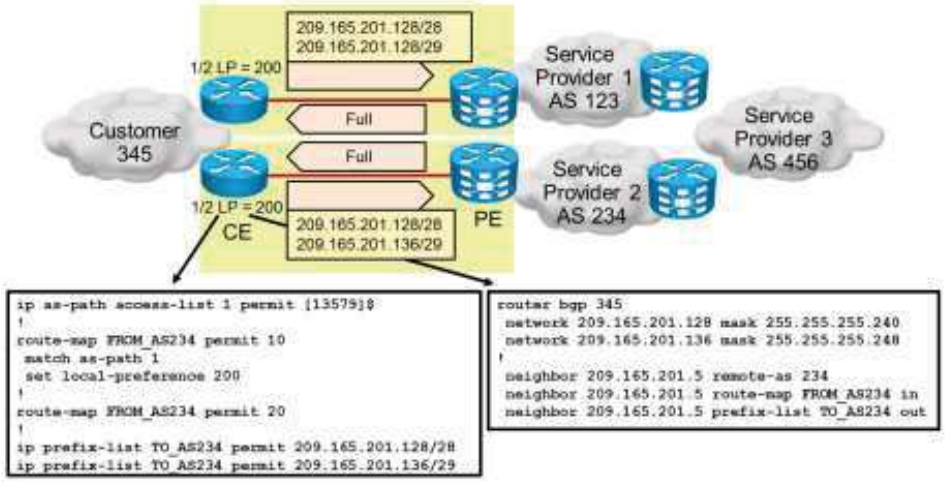
The configuration also shows how to configure filtering on the customer side to prevent the customer from becoming a transitive AS. Filtering with prefix lists allows only customer routes from the CE router to the PE router to be advertised.

Note	For a complete command reference, refer to the Cisco IOS Software Command Reference guides on www.cisco.com.

https://t.me/learningnets

Customer-Implemented BGP Routing Policies in a Load-Balancing Scenario (Cont.)

Configuration of CE (Cont.):



The secondary CE router, which runs Cisco IOS Software, is configured with an IBGP session with another CE router (not shown in the example) and an EBGP session with the secondary PE router. The secondary CE router sets the local preference to 200 for half of the networks that are received from the secondary service provider. In the example, the prefixes that are received from the secondary service provider will be assigned a local preference of 200 if they originate in an odd-numbered AS (the AS-path attribute ends with 1, 3, 5, 7, or 9). Other prefixes will be assigned the default local preference of 100.

Load-balancing configuration for incoming traffic is achieved by splitting the customer address space into two halves and by advertising one half (209.165.201.136/29) to the secondary service provider. Together, this will complete the customer network (209.165.201.128/28) for backup purposes.

The configuration also shows how to configure filtering on the customer side to prevent the customer from becoming a transitive AS. Filtering by using prefix-lists allows advertising of only customer routes from the CE router to the PE router.

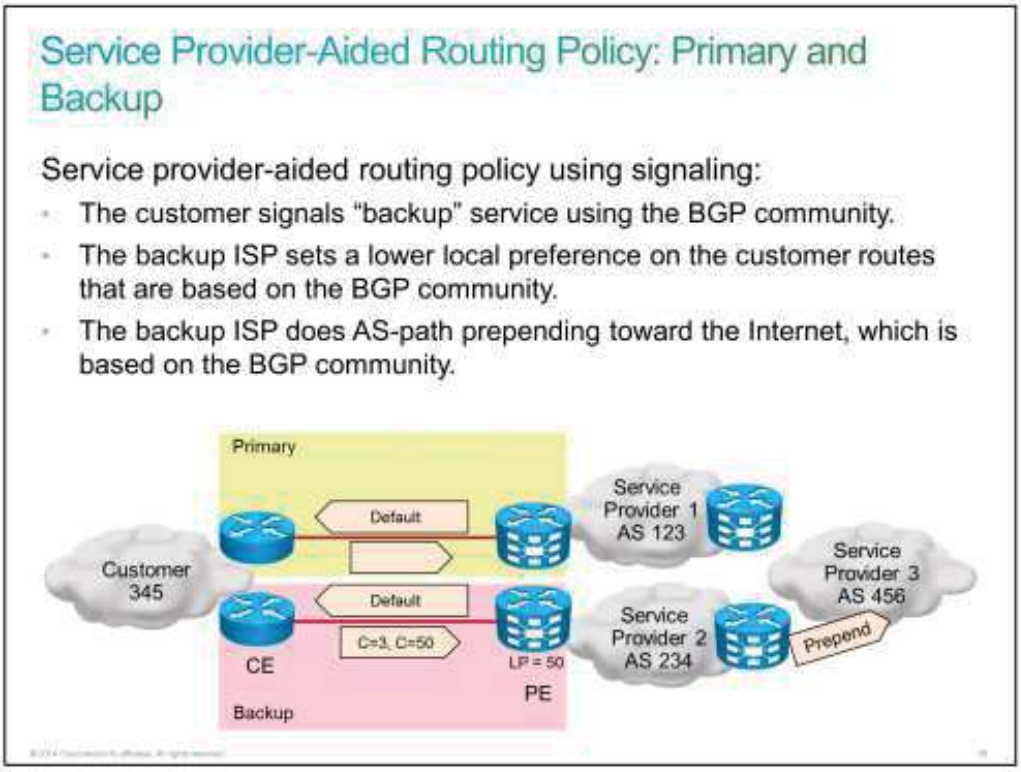
Note	For a complete command reference, refer to the Cisco IOS Software Command Reference guides on www.cisco.com.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Service Provider-Aided BGP Routing Policy in a Primary and Backup Scenario

The service provider uses a routing policy for the primary or backup distribution of traffic.



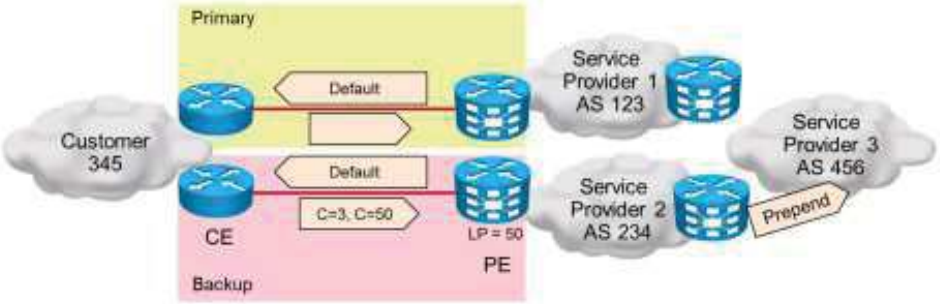
In this example, the customer signals backup service using BGP communities, which are defined and published by the backup service provider. These BGP communities are attached to the customer routes. The backup service provider then sets a higher local preference for customer routes and does AS-path prepending of routes that are advertised toward the Internet based on matched BGP communities.

https://t.me/learningnets

Service Provider-Aided Routing Policy: Primary and Backup (Cont.)

Backup service provider services:

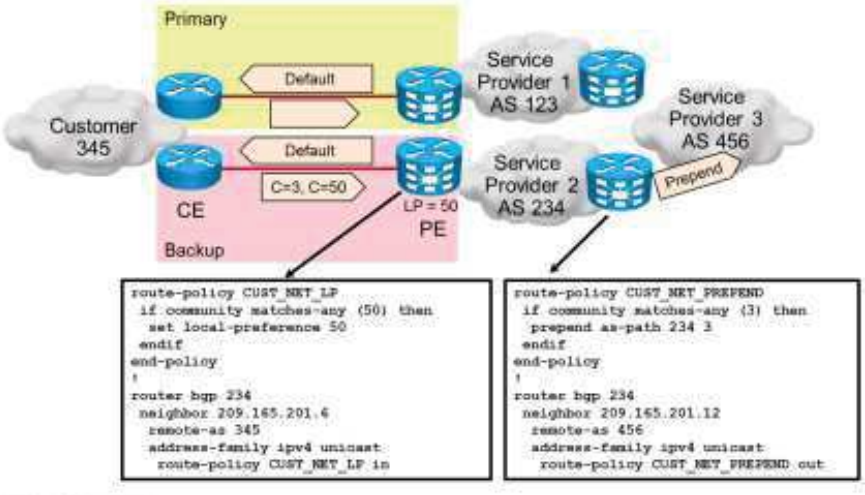
- C = 1: prepend AS number once
- C = 2: prepend AS number twice
- C = 3: prepend AS number three times
- C = 50: local preference equals 50
- C = 200: local preference equals 200



A customer can signal needed backup services to the backup service provider using BGP communities. These communities are defined by the service provider in a similar manner as in the example. If the customer appends the BGP community with a value of 3 to an advertised route, the service provider will prepend its own AS number three times to the customer route when advertising it further to the Internet. If the customer appends the BGP community with a value of 50 to an advertised route, the service provider will set the local preference to 50 when advertising the customer route inside AS number 234.

Service Provider-Aided Routing Policy: Primary and Backup (Cont.)

Configuration of PE:



 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

The figure shows configuration on the PE routers. The customer-facing PE router is configured with the CUST_NET_LP route policy that matches routes with BGP community 50. The router then sets the local preference to 50 to routes that are advertised to the IBGP neighbor. The Internet-facing PE router is configured with the CUST_NET_PREPEND route policy that matches routes with BGP community 3 and prepends AS number 243 three times to routes that are advertised to the Internet.

Configuration of the customer-facing PE router, running Cisco IOS and IOS XE Software, is as follows:

```
ip community-list 1 permit 50
!
route-map CUST_NET_LP permit 10
 match community 1
 set local-preference 50
!
router bgp 234
 neighbor 209.165.201.6 remote-as 345
 neighbor 209.165.201.6 route-map CUST_NET_LP in
```

Configuration of the Internet-facing PE router, running Cisco IOS and IOS XE Software, is as follows:

```
ip community-list 1 permit 3
!
route-map CUST_NET_PREPEND permit 10
 match community 1
 set as-path prepend  243 243 243
!
router bgp 234
 neighbor 209.165.201.6 remote-as 345
 neighbor 209.165.201.6 route-map CUST_NET_PREPEND out
```

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

Summary

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

Summary

- Static routing can be used with:
  - Single-attached customers using a single IP address.
  - Single-attached customers using multiple IP addresses.
  - Dual-attached customers using a primary and a backup path.
  - Dual-attached customers using load balancing.
- BGP can be used in a dual-attached scenario for conditional advertising on a CE router.
- BGP is used on PE routers by service providers in dual-attached scenarios.
- BGP must remove private AS numbers in dual-attached scenarios.
- BGP is run on both CE and PE routers in dual-attached scenarios.
- The local AS feature is used to mask the real AS number.

Summary (Cont.)

- BGP should be used with:
  - Dual-attached customers using a primary and a backup path.
  - Dual-attached customers using load balancing.
  - Multihomed customers.
  - Multihomed customers using a primary and a backup path.
  - Multihomed customers using load balancing.
- Some service providers allow signalization of BGP services using BGP communities to aid customers using primary and backup paths.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

References

For additional information, refer to these resources:

- Load Sharing with BGP in Single and Multihomed Environments: Sample Configurations
  http://www.cisco.com/en/US/tech/tk365/technologies_configuration_example09186a00800945bf.shtml
- Sample Configuration for BGP with Two Different Service Providers (Multihoming)
  http://www.cisco.com/en/US/tech/tk365/technologies_configuration_example09186a008009456d.shtml

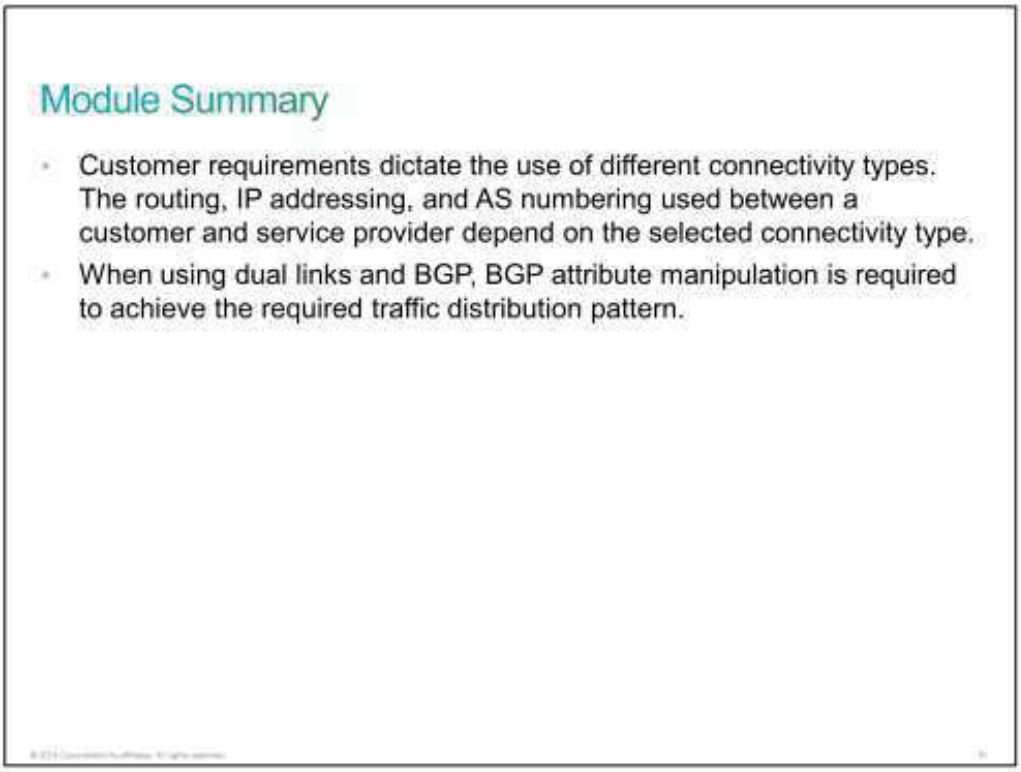
https://t.me/learningnets

https://t.me/learningnets

Lesson 3: Module Summary

Overview

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



This module covered different customer-to-provider connectivity types. The module described connectivity types and routing, IP addressing, and AS numbering options that are available in each connectivity type. The module also described how to configure customer and service provider devices for a required connectivity type.

https://t.me/learningnets

https://t.me/learningnets

Lesson 4: 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. A customer would like to connect to a service provider. Which two requirements should be considered before deciding on a type of connectivity? (Choose two.) (Source: Defining Customer-to-Provider Connectivity Requirements)

   A. Application availability
   B. Redundancy
   C. Cost of ownership
   D. Flexibility

2. Which two routing options could be used when connecting a customer to a service provider? (Choose two.) (Source: Defining Customer-to-Provider Connectivity Requirements)

   A. Static routing
   B. OSPF
   C. IS-IS
   D. BGP

3. Which two combinations are recommended when connecting a customer to a service provider? (Choose two.) (Source: Defining Customer-to-Provider Connectivity Requirements)

   A. Single-homed customer, PI address space, and static routing
   B. Dual-attached customer, PA address space, BGP routing, and public AS number
   C. Dual-attached customer, PA address space, BGP routing, and private AS number
   D. Multihomed customer, PI address space, BGP routing, and public AS number

https://t.me/learningnets

4. Customer networks that can be summarized in a service provider network should be tagged with no-export BGP community when redistributed into BGP. (Source: Connecting a Customer to a Service Provider)

   A. True
   B. False

5. You are a customer connected to a single service provider with two permanent links. You would like to achieve primary and backup traffic distribution. Which two BGP attributes do you have to manipulate? (Choose two.) (Source: Connecting a Customer to a Service Provider)

   A. Local preference for incoming traffic
   B. Local preference for outgoing traffic
   C. AS path for incoming traffic
   D. MED for incoming traffic

6. You are a customer connected to two service providers with one link to each provider. You would like to achieve primary and backup traffic distribution. Which two BGP attributes do you have to manipulate? (Choose two.) (Source: Connecting a Customer to a Service Provider)

   A. Local preference for incoming traffic
   B. Local preference for outgoing traffic
   C. AS path for incoming traffic
   D. MED for incoming traffic

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Module Self-Check Answers

Answer Key

1. B, D
2. A, D
3. C, D
4. A
5. B, D
6. B, C

https://t.me/learningnets

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Module 2: Scale Service Provider Network

Introduction

One of the major concerns in service provider networks is scalability, since it defines number of customers that can connect to the service provider. Two of the important aspects that a network engineer has to consider are BGP routing scaling and internal IPaddressing in the service provider core network. The service provider also has to be able to monitor how traffic flows from or to a customer in order to be able to optimize routing policies.

This module first describes how routes are propagated inside a typical service provider network. The module then discusses how to scale IP addressing and BGP routing. The module also describes BGP route reflectors and BGP confederations as effective means to achieve BGP scalability. The module concludes with BGP policy accounting, which enables the service providers to measure traffic that is sent to or received from different BGP peers.

Upon completing this module, you will be able to:

- Describe common routing and addressing scalability issues in service provider networks
- Describe the function of route reflectors and confederations in a BGP environment, and list the steps needed to implement BGP route reflectors

https://t.me/learningnets

https://t.me/learningnets

Lesson 1: Scaling BGP in Service Provider Networks

Overview

Properly scaling IP addressing and BGP is a common area of concern to all service providers and can be the difference between a successful and a problematic BGP implementation. Next-generation service P-networks are complex and must meet the administrative policy and routing demands of the internal network, different customers, and other providers. Therefore, proper scaling is crucial to the success of the network. Interactions between IGP and the BGP can be quite complex, especially when network administrators are supporting internal routing, customer connectivity, and transit traffic (and the administrative policies that match). Furthermore, the large number of prefixes that are required to support complete Internet routing requires administrators to fully characterize IGP and BGP interactions for internal networks and customers.

The lesson first describes how routes propagate inside the service provider core network. The lesson then discusses scaling issues in service provider networks, and actions that can be taken to scale BGP routing and IP addressing inside the service provider core network.

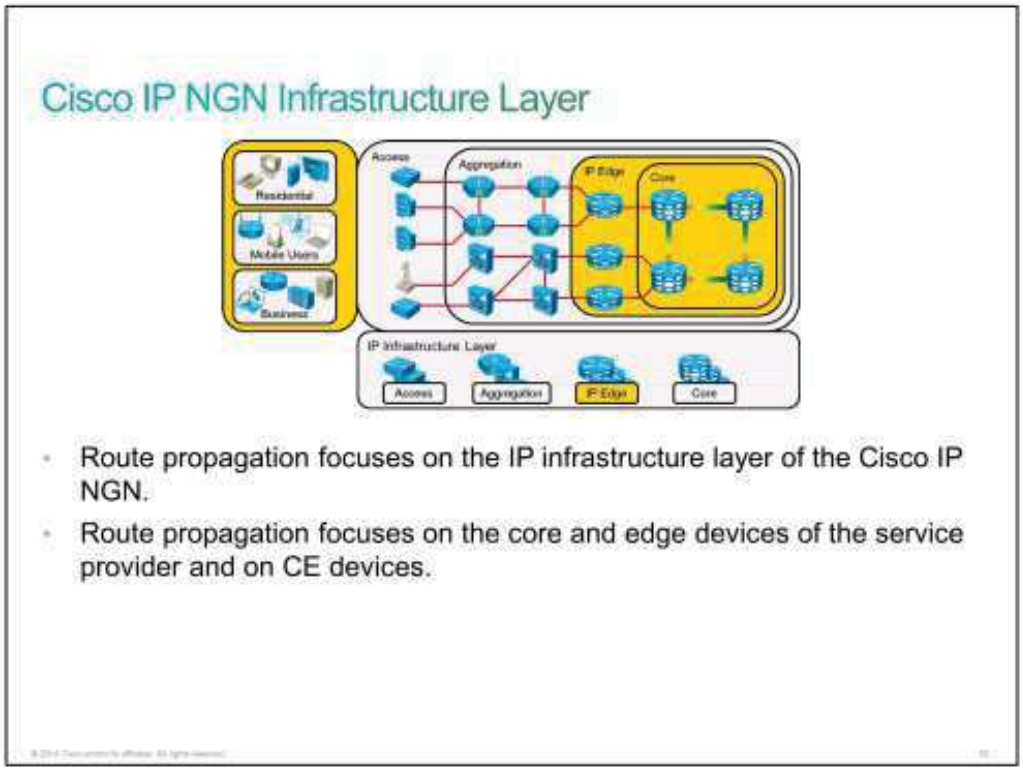
Upon completing this lesson, you will be able to:

- Show the Cisco IP NGN infrastructure layer
- Describe the common interior (OSPF and IS-IS) and exterior (BGP) routing protocols used in service provider networks
- Describe BGP and IGP route propagation in service provider networks
- Describe BGP route exchange with upstream service providers
- Describe how customer routes are learned using BGP or static routing
- Describe the BGP next-hop-self option
- Describe proper scaling of BGP in service provider networks
- Describe proper IP address scaling in service provider networks
- Describe the BGP policy accounting feature and configuration

https://t.me/learningnets

Cisco IP NGN Infrastructure Layer

Route propagation describes how customer routes are propagated across the service P-network and then to the Internet, and how internal routes in the service P-network are propagated inside the service provider core network.

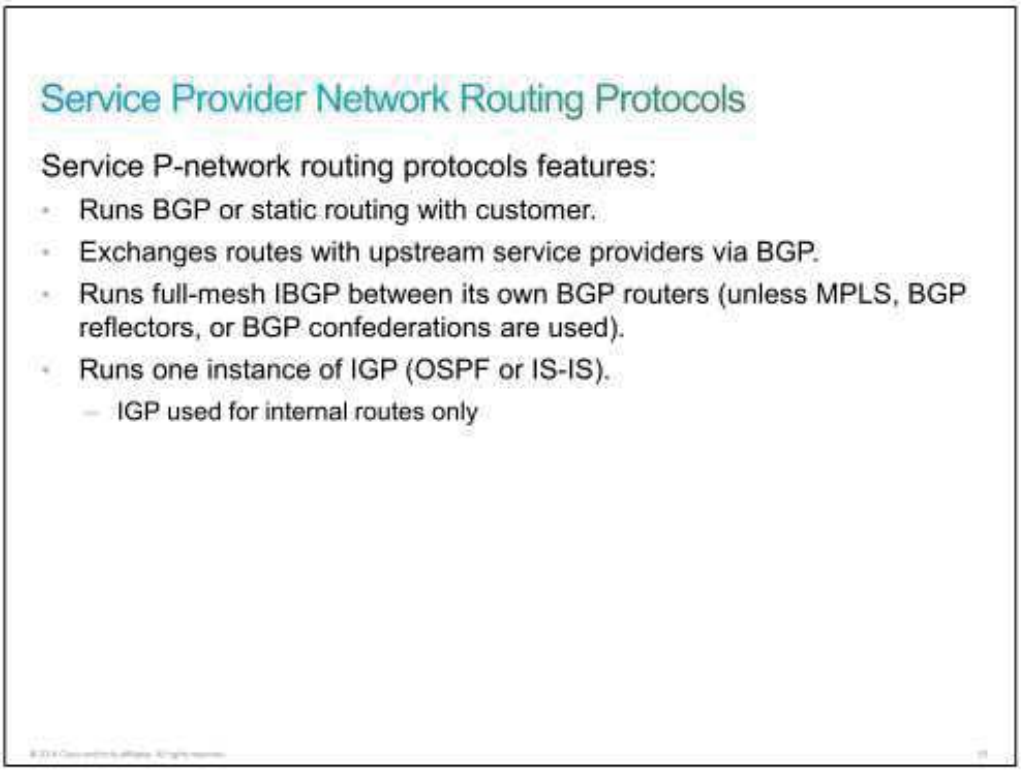


Route propagation is a part of the IP infrastructure layer of the Cisco IP Next-Generation Network. It focuses on customer premises equipment and the core and edge devices of the service provider.

https://t.me/learningnets

Service Provider Network Routing Protocols

The common service P-network runs EBGP or static routing with customers. EBGP is always used as the routing protocol between different service providers.



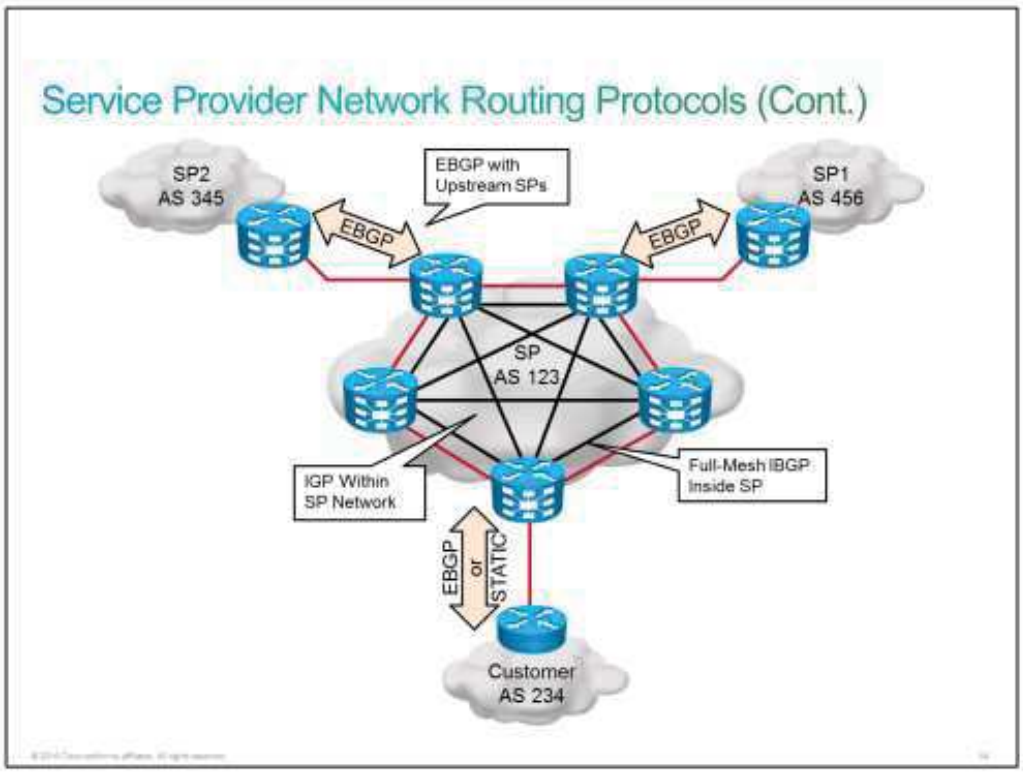
IBGP is required in the P-network because all EBGP-speaking routers in an AS must exchange external routes via IBGP. Also, non-EBGP-speaking routers must take part in the IBGP exchange if they are in a transit path and forward packets based on destination IP addresses. Full mesh of IBGP is required between service provider BGP routers, unless MPLS is used to switch packets, or BGP reflectors or BGP confederations are used. If MPLS is deployed in the service provider core network, then BGP can be removed from the service provider core network. With MPLS enabled, traffic across the service provider core network can be MPLS-switched rather than IP-routed. The core P routers can use the MPLS label to switch the traffic towards the edge PE router. The IBGP full mesh requirement within the AS can be removed. Only the PE routers need to have IBGP peering among each other when MPLS is deployed in the core network. BGP reflectors and BGP confederations will be discussed in the next lesson.

The service P-network also runs an IGP. The protocols of choice are OSPF and IS-IS. The IGP is used for two purposes:

- It provides IP connectivity between all IBGP speakers so that TCP sessions for IBGP can be established between BGP-speaking routers. Remember that full mesh of IBGP is needed and that IBGP neighbors are often more than one hop away.
- It provides optimal routing to the BGP next-hop address.

A single IGP should be used within the entire AS. This setup facilitates effective packet forwarding from the ingress router to egress routers. The IGP is configured to carry internal routes only, including internal links and loopback addresses of the routers. For performance and scalability reasons, no customer routes or external routes should be injected into the IGP.

https://t.me/learningnets



The typical service P-network consists of a network core that connects various edge devices. Some of the edge devices connect customers, while others connect to other service providers.

The edge devices that connect to other service providers use EBGP to exchange routing information. The edge devices that connect customers use either static routing or EBGP.

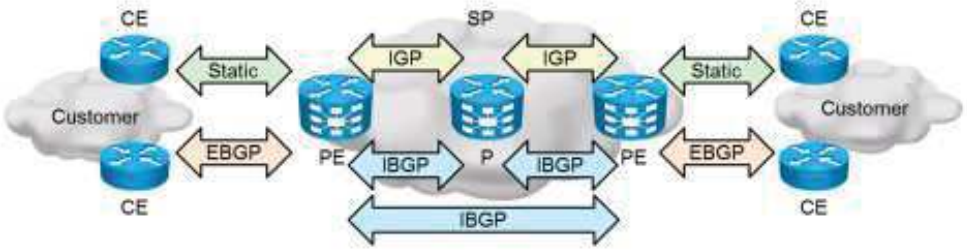
Unless MPLS is configured on the service provider backbone, routers in a transit path are also required to have complete routing information. Therefore, these routers take part in the IBGP routing exchange.

An IGP is also required within the service P-network. The IGP is used to carry internal routes, including the loopback interface addresses of IBGP-speaking routers. The IGP provides reachability information to establish IBGP sessions and to perform the recursive routing lookup for the BGP next hop.

 © 2014 Cisco Systems, Inc.

https://t.me/learningnets

Service Provider Network Routing Protocols (Cont.)



- PE routers use EBGP or static routing with CE routers.
- PE and P routers use full-mesh IBGP routing.
- The provider core IGP is a single instance of IS-IS or OSPF and is used only within the service provider core network.
- Optimal routing between PEs is desired.

CE routers connect via IP to PE routers. In many cases, the PE routers use static routing to customer networks. The PE routers advertise static routes to the rest of the service P-network and to other autonomous systems using BGP. PE and P routers use full-mesh IBGP routing to exchange BGP routing information.

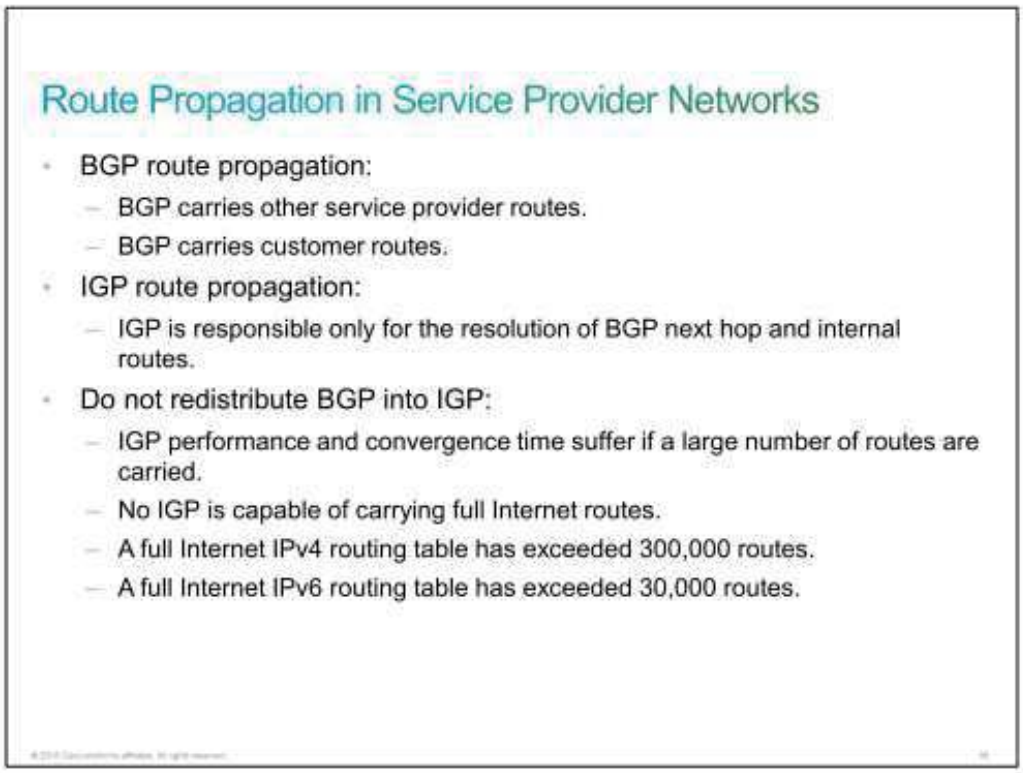
Service providers use BGP routing with the customer when redundancy requires the use of a routing protocol.

The service provider backbone typically uses a single instance of either IS-IS or OSPF as its IGP. The IGP is used within the provider backbone only. The provider backbone exchanges no IGP routing information with customer routers or with routers in other autonomous systems.

https://t.me/learningnets

Route Propagation in Service Provider Networks

It is important to avoid sending any unnecessary routing information in the IGP. The IGP performs best if it carries as few routes as possible. Performance and convergence time will suffer if the IGP carries a large number of routes, so the design goal should be to minimize the volume of routing information that is carried by the IGP. Note that the number of route flaps is reduced as the number of routes is reduced.



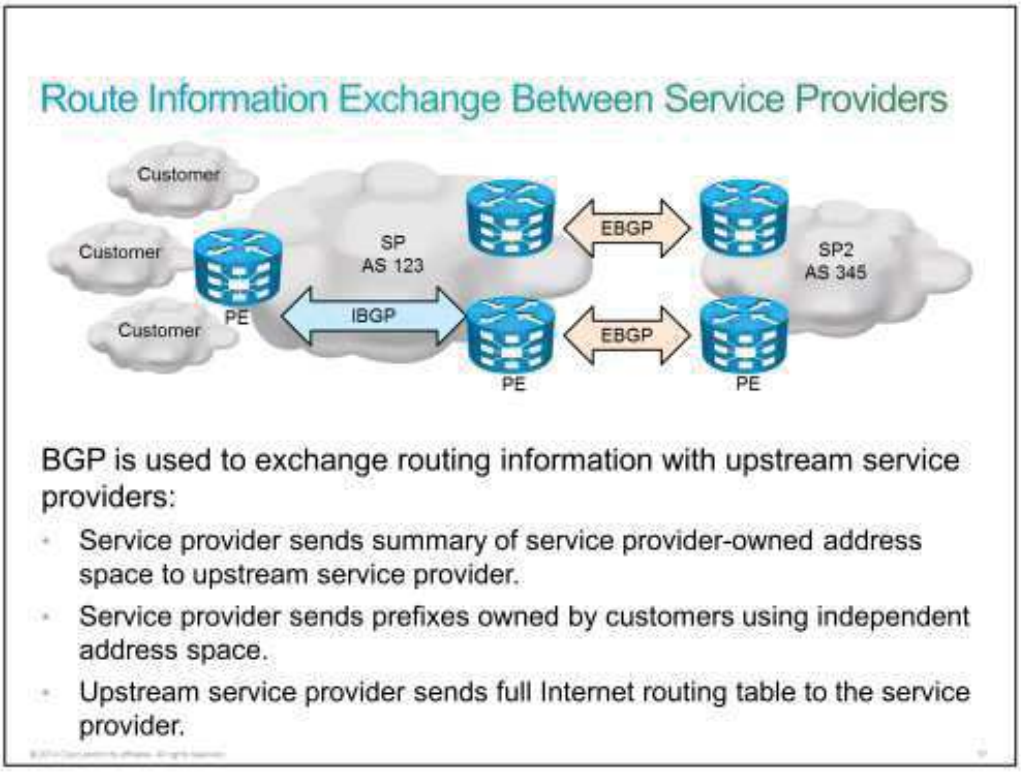
Optimally, the IGP should contain only information about BGP next hops and routes that are internal to the service P-network, enabling the establishment of IBGP sessions. All other routing information should be carried in BGP, which is designed to scale for large volumes of routing information. Customer routes and the routes from other service providers should be carried in BGP. These routes should not be propagated from BGP into the provider IGP.

BGP scales to a much larger volume of routing information because of the inherent qualities of the design of BGP. Potentially, the BGP routers of the service provider can receive the full Internet routing table, which has exceeded 300,000 routes for IPv4, and 30,000 for IPv6. You should therefore never redistribute the routing information that has been received by BGP into the IGP, because no IGP is capable of carrying several hundreds of thousands of routes.

https://t.me/learningnets

Route Information Exchange Between Service Providers

PE routers use BGP to exchange routing information with other service P-networks for redundancy and scalability reasons.



The figure shows how routes are exchanged between service providers.

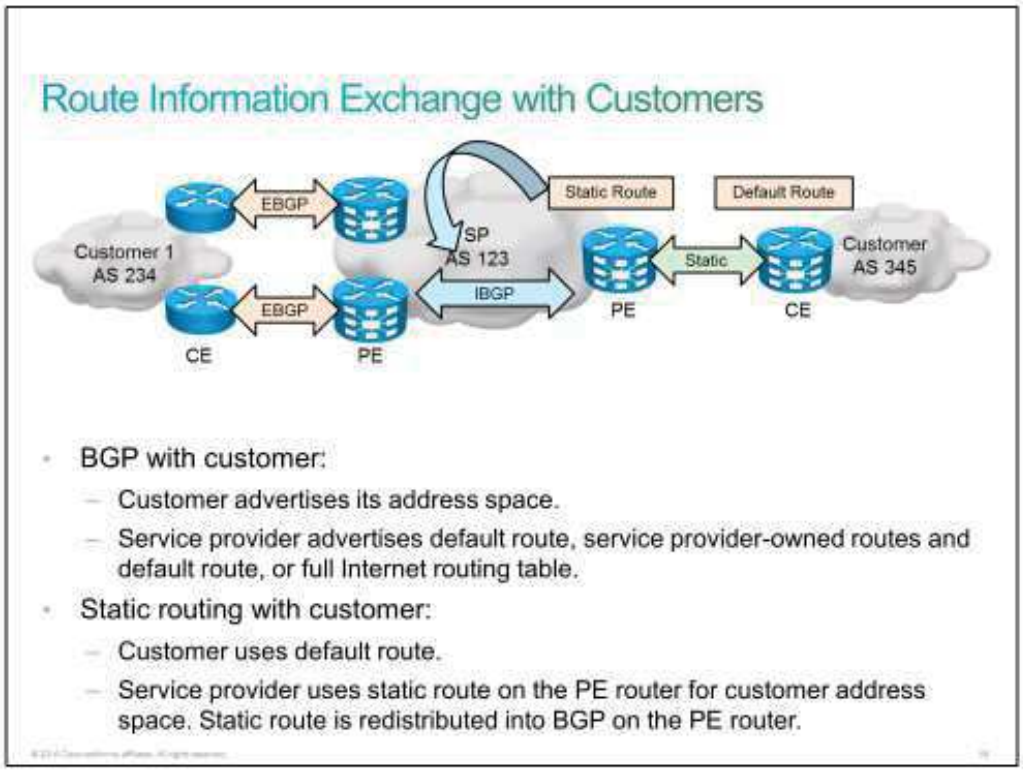
Static routing with other service providers is generally not a viable solution, due to the dynamic routing requirements of the service provider environment.

Routing information from customers is received at PE routers using EBGP and then propagated using IBGP to the rest of the service P-network. At another PE router, the routing information is further propagated to a different service provider using EBGP with other autonomous systems. The service provider sends a summary of service provider-owned space to upstream service providers, along with prefixes owned by the customer using provider-independent address space. The upstream service provider sends the full Internet routing table to the service provider, which allows the service provider to implement a primary/backup scenario or a load-balancing scenario when peering with multiple upstream service providers.

https://t.me/learningnets

Route Information Exchange with Customers

The first option, which is used when redundancy is required by the customer, is running BGP between the customer and the service provider.



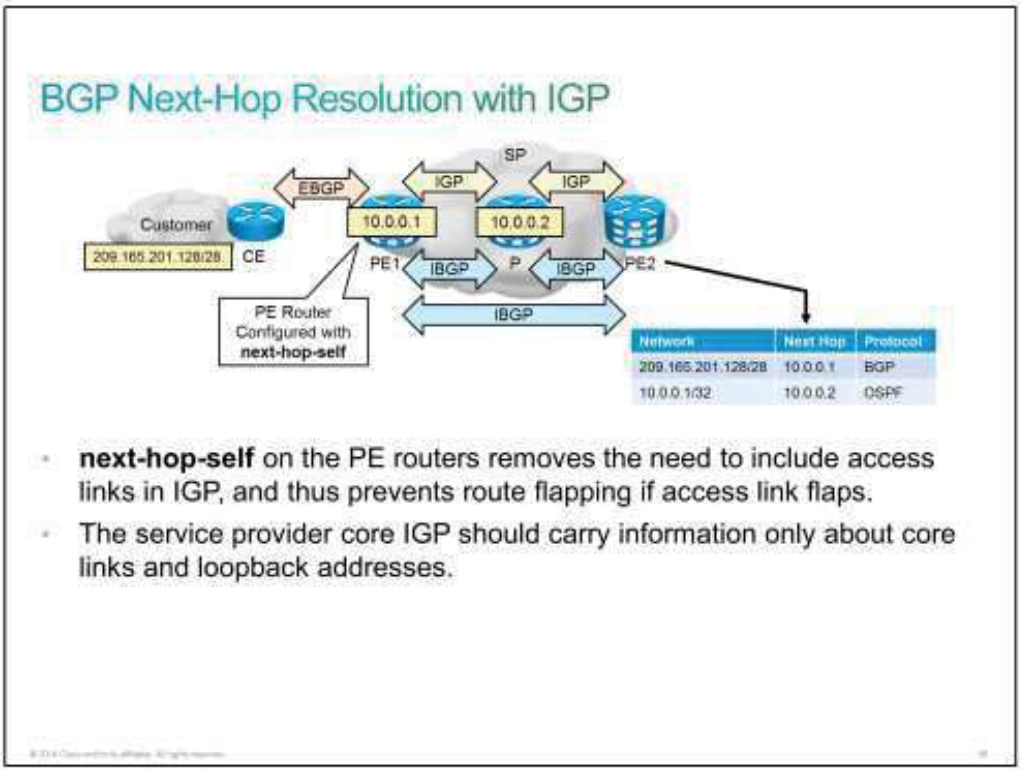
The figure shows how customers exchange routes with the service provider. In this case, the customer advertises its address space to the service provider. The service provider advertises to the customer either the default route, service provider-owned routes and the default route, or the full Internet routing table, based on the customer requirements.

When redundancy is not required by the customer, static routing is used. In this case, the customer configures a static default route that points to the PE router of the service provider. The service provider configures a static route for the customer network that points to the CE router. The PE router redistributes static routes into BGP. The service P-network then uses BGP to propagate the information to the rest of the service P-network using IBGP. The service provider also advertises customer routing information to other autonomous systems using BGP.

https://t.me/learningnets

BGP Next-Hop Resolution with IGP

The IGP used in the service provider core should carry information only about core links and loopback addresses. The service provider should use BGP to carry all other information.



The figure shows why IGP is needed in the service provider core network.

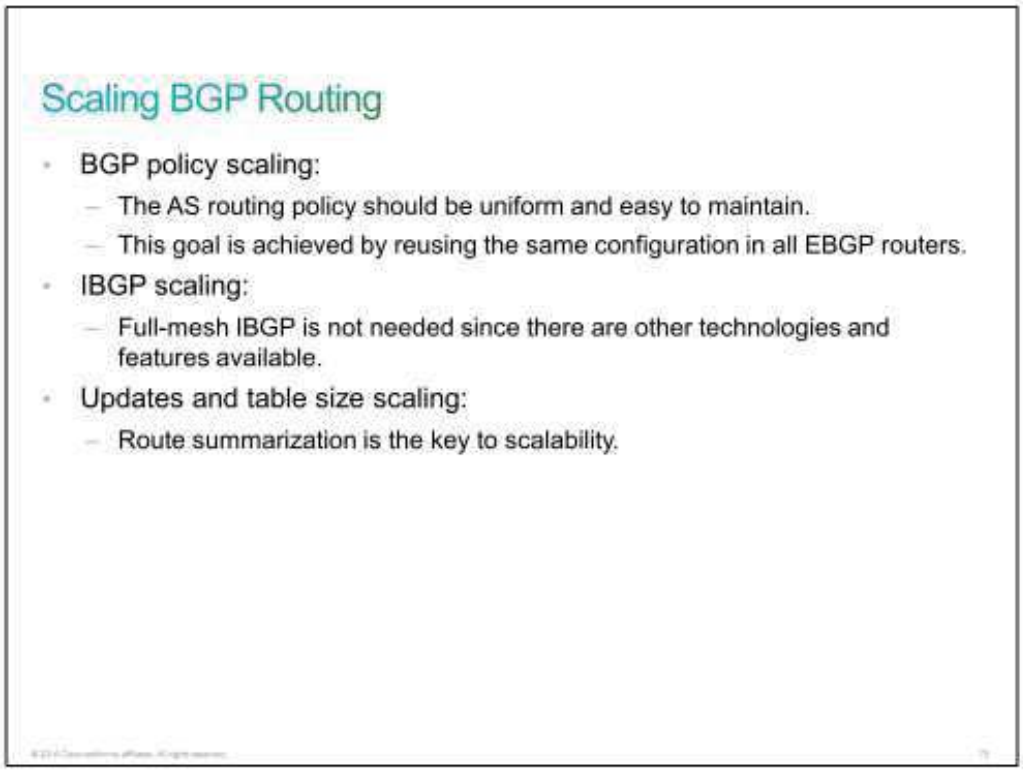
Use the BGP **next-hop-self** command on the PE routers when BGP routing is exchanged with the customer or other service providers. Using the **next-hop-self** command results in the BGP next hop being set to the address of the PE router and not to the access link address of the customer. The IGP can then be relieved of the burden of carrying information about the access link. The benefit of not carrying customer link information is that a flapping access link will not disturb the service provider IGP.

As shown in the routing table on the PE2 router, the next hop for the customer network points to the address of the PE1 router (usually the IP address of the loopback interface). Because the PE1 and PE2 routers are not directly connected, an IGP routing protocol is needed that carries information about core links and loopbacks across the service P-network.

https://t.me/learningnets

Scaling BGP Routing

The proper scaling of BGP in service P-networks includes: BGP policy scaling, IBGP mesh scaling, and updates and table size scaling.



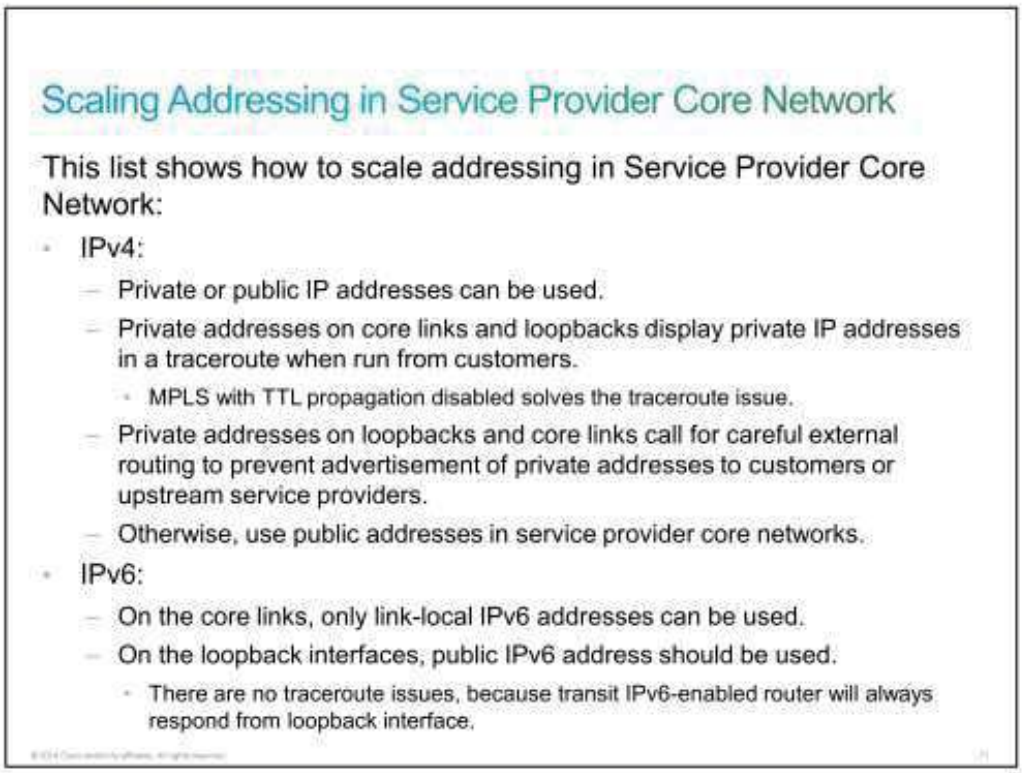
The task of scaling BGP actually involves three different and independent scaling tasks:

- **BGP policy scaling:** The AS routing policy should be uniform and easy to maintain. Different PEs in the same AS should not use different policies, since that would advertise different routes to neighboring autonomous systems. Regardless of which router is currently active, the same routing policy should be in place. Administratively, replication of the same routing policies requires the same configuration lines in several PE routers.
- **IBGP mesh scaling:** All BGP-speaking routers must be updated with consistent IBGP information. In the traditional BGP approach, ensuring consistent routing information was achieved by establishing a full mesh of IBGP sessions between all routers within the AS. An IBGP full mesh is not scalable, and several tools are now available to achieve the same results without the full mesh.
- **Updates and table size scaling:** The number of routes in the routing table and the number of updates that are sent and received represent the third scaling task. Route summarization is the key to this scalability.

https://t.me/learningnets

Scaling Addressing in Service Provider Core Networks

Internal IP addressing in the service provider core network can be simplified to reduce the use of public addresses and to simplify configuration.



For IPv4, public or private IP addresses can be used on core links. However, using private addresses has some drawbacks. Private addresses on the provider internal links will cause trouble for the traceroute application. When the **traceroute** command is executed from the customer, the customer will see private IP addresses in the traceroute printout. Using MPLS without TTL propagation in the service P-network can easily overcome the traceroute problem with private addresses. If these functions are used, the provider network will appear as a single hop to the traceroute application. The intermediate routers will be invisible and thus can use private addresses.

Using private addresses on the service P router loopback interfaces is possible. However, you must take care not to advertise any private addresses to any other AS.

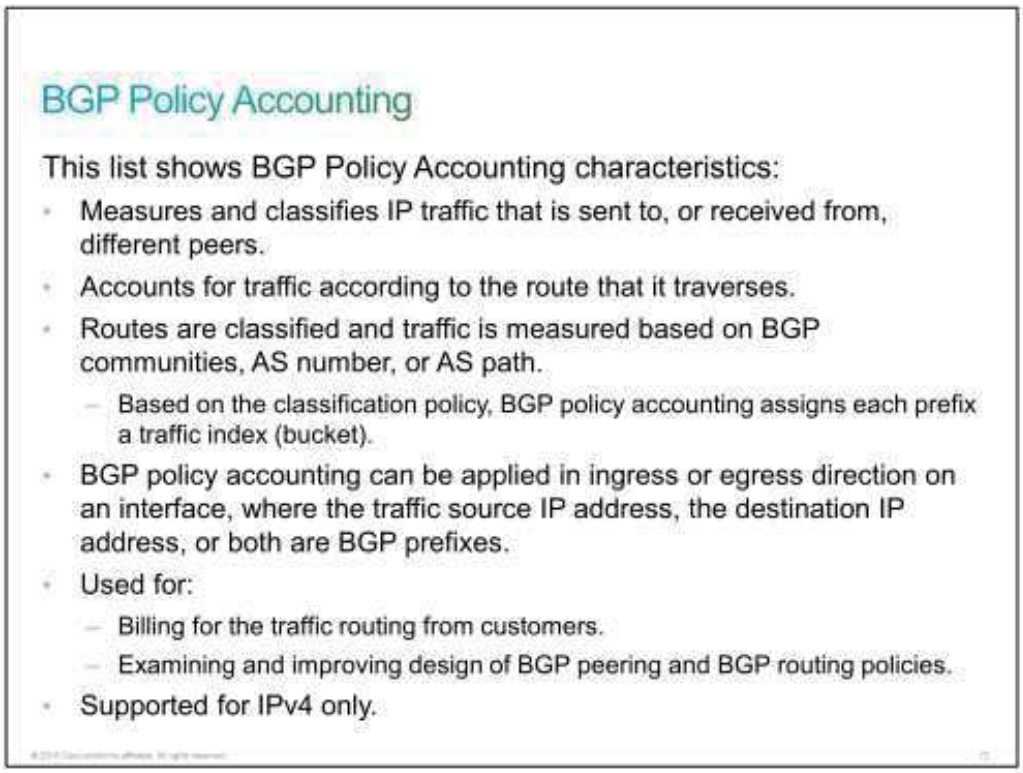
A rule of safety is to prevent the announcement of any private addresses by using prefix-lists that are applied on outgoing updates to external neighbors. The same prefix-list mechanism can also be used on the provider edge routers to prevent accepting private addresses from any other AS if the other AS, by mistake, announces private addresses.

For IPv6, only link-local IP addresses can be used on the core links. On the loopback interfaces, public IPv6 addresses should be used. Even though the link-local addresses are used on the core links, there are no problems with traceroute. The IPv6 enabled router will always respond to traceroute using loopback interface, if transit interfaces use only link-local addressing.

https://t.me/learningnets

BGP Policy Accounting

As network administrators learn to manage and scale larger and larger networks, they must also be able to account for the usage of a growing customer base. How can you ensure that customers are being charged correctly for their network utilization? How can you ensure that they are receiving the services for which they have contracted? BGP policy accounting addresses these concerns.



BGP policy accounting using AS numbers can be used to improve the design of network circuit peering and transit agreements between service providers.

BGP policy accounting measures and classifies IP traffic that is sent to, or received from, different peers. Policy accounting is enabled on an interface in the ingress or egress direction, and counters based on parameters such as community-list, AS number, or AS path are assigned to identify the IP traffic. Using BGP policy accounting, you can account for traffic according to the route that it traverses. Service providers can identify and account for all traffic by customer and can bill accordingly.

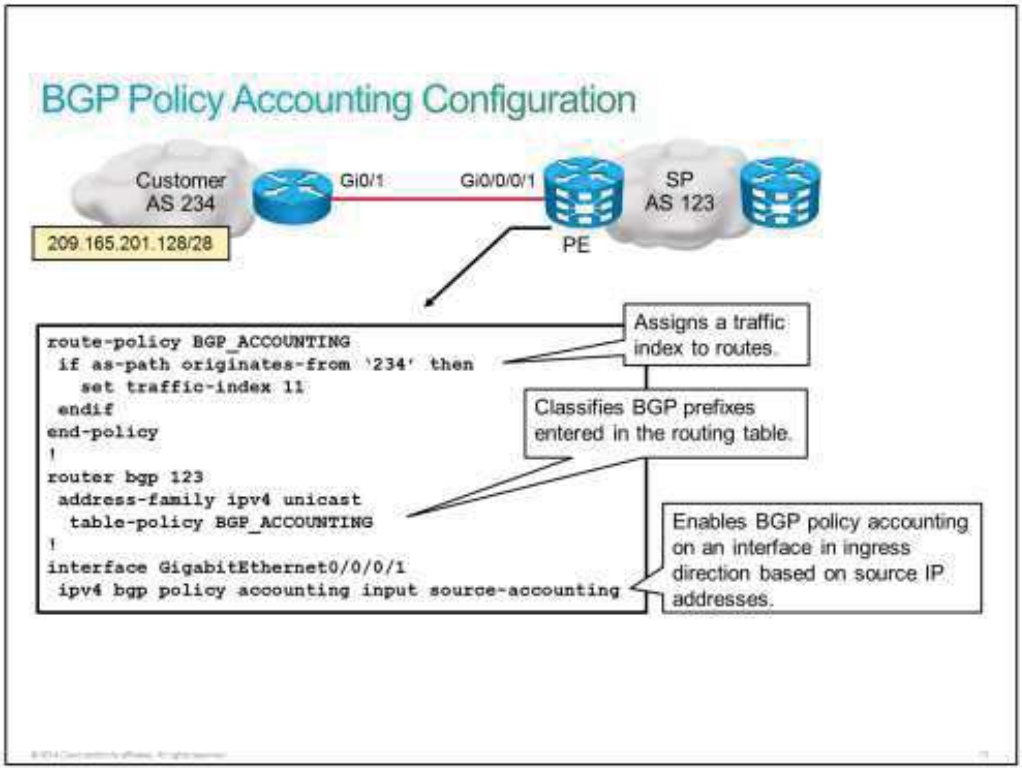
Based on the specified routing policy, BGP policy accounting assigns each prefix a traffic index (bucket) associated with an interface. BGP prefixes are downloaded from the RIB to the FIB along with the traffic index.

There are a total of 63 (1 to 63) traffic indexes (bucket numbers) that can be assigned for BGP prefixes. Internally, there is an accounting table that is associated with the traffic indexes to be created for each input (ingress) and output (egress) interface. The traffic indexes allow you to account for the IP traffic, where the source IP address, the destination IP address, or both are BGP prefixes.

BGP policy accounting is supported for IPv4 only.

https://t.me/learningnets

BGP Policy Accounting Configuration



The example shows the configuration of BGP policy accounting.

The PE router, which runs Cisco IOS XR Software, is configured with BGP. First, configure a route policy, which matches routes originating from the customer address space and assigns a traffic index of 11 to them. Then, enable the router to classify BGP prefixes that are entered into the routing table by applying the route policy under BGP router configuration mode. Use the **table-policy** command to apply the route policy. Finally, enable BGP policy accounting under the interface configuration mode using the **ipv4 bgp policy accounting input source-accounting** command. The **input** keyword specifies that traffic will be accounted for in the inbound direction. The **source-accounting** keyword specifies that accounting will be based on source IP addresses. This means that if the source IP address is a prefix that is known to the router via BGP, traffic will be accounted for.

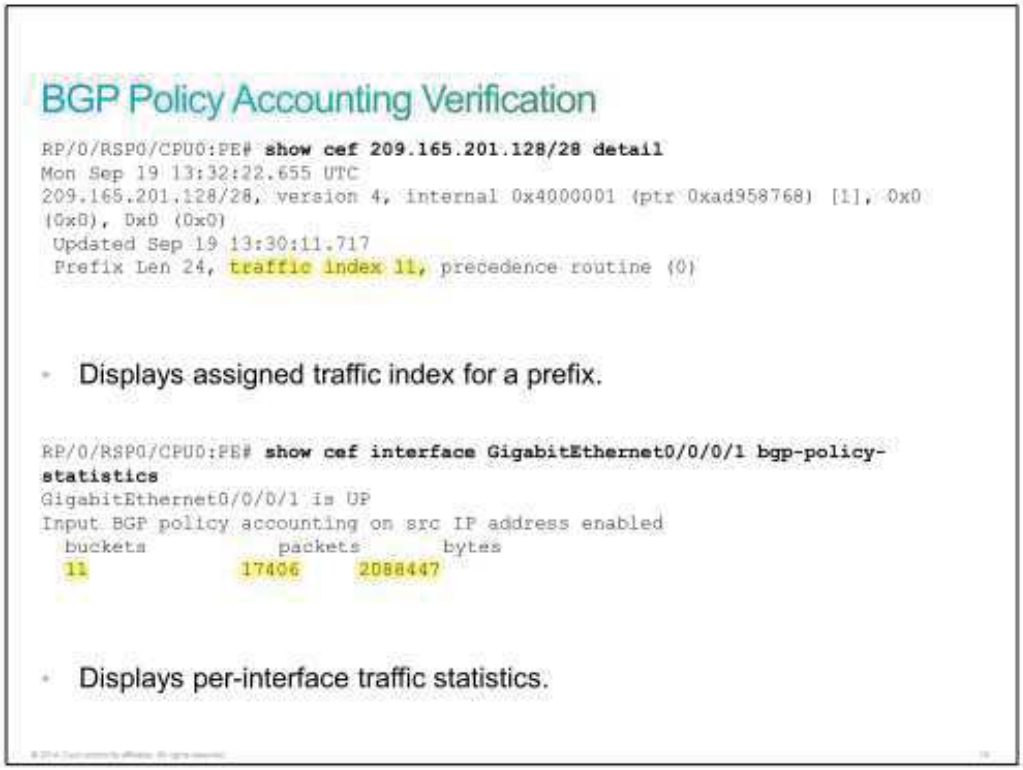
Configuration of the customer-facing PE router, running Cisco IOS and IOS XE Software, is as follows:

```
ip as-path access-list 1 permit _234$
!
route-map BGP_ACCOUNTING permit 10
 match as-path 1
 set traffic-index 11
!
router bgp 123
 table-map BGP_ACCOUNTING
!
interface GigabitEthernet0/0/1
 bgp-policy accounting input source
```

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

https://t.me/learningnets

BGP Policy Accounting Verification



To verify if routes are assigned a traffic index, use the **show cef detail** command. The output should show you which index is assigned to the specified prefix. To verify per-interface traffic statistics, use the **show cef bgp-policy-statistics** command. The output should show you all configured traffic indexes (buckets), and the amount of traffic in packets and bytes that falls into an individual bucket. In the example, you see that 17,406 packets have been received from the customer subnet.

https://t.me/learningnets

Summary

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

Summary

- Route propagation focuses on the IP infrastructure layer of the Cisco IP NGN.
- Service providers most commonly use integrated IS-IS and OSPF as interior gateway protocols and BGP as the exterior gateway protocol.
- BGP is used to carry customer routes while IGPs are used to carry service provider internal prefix reachability information.
- BGP allows ISP clients to acquire information about all or some networks reachable through the ISP.
- Static routing or BGP can be used by the ISP to direct traffic going to the customers to the correct links.

Summary (Cont.)

- Next-hop-self can be used to avoid redistributing transit segments into IGP on iBGP neighbors.
- When BGP networks grow, various actions must be taken to make them scalable, for iBGP scalability use route reflectors or confederations.
- When IP networks grow, several aspects of addressing need to be considered to reduce sizes of routing tables and to avoid consuming too many addresses.
- BGP accounting feature can be used when an overview of BGP's use of resources or detailed statistical analysis are required.

https://t.me/learningnets

https://t.me/learningnets

Lesson 2: Introducing BGP Route Reflectors and Confederations

Overview

In standard BGP implementations, all BGP routers within an AS must be fully meshed so that all external routing information can be distributed among the other routers that reside within the AS. Therefore, within an AS, all routers must establish TCP sessions with all other BGP routers. As the AS grows, scalability challenges arise because of an ever-increasing number of TCP sessions and demands for router CPU and memory resources.

There are two approaches available to address IBGP scalability in service provider networks. The first approach is the use of route reflectors, which modify the BGP split-horizon rule. The second approach is the use of BGP confederations, which allow creation of smaller autonomous systems inside one AS, thus transforming IBGP sessions into intraconfederation EBGP sessions. BGP route reflectors are more common and recommended approach to address IBGP scalability in Cisco Next-Generation Networks service provider networks.

The lesson first describes BGP route reflectors. The lesson then continues on how to design and implement route reflectors in the service provider network. The lesson concludes with overview of BGP confederations.

Upon completing this lesson, you will be able to:

- Explain the need for BGP route reflectors or BGP confederations in BGP transit core networks
- Describe the IBGP split-horizon rule
- Describe route propagation in a route reflector-enabled network
- Describe using redundant route reflectors
- Describe route reflector clusters
- Describe the originator-ID BGP attribute

https://t.me/learningnets

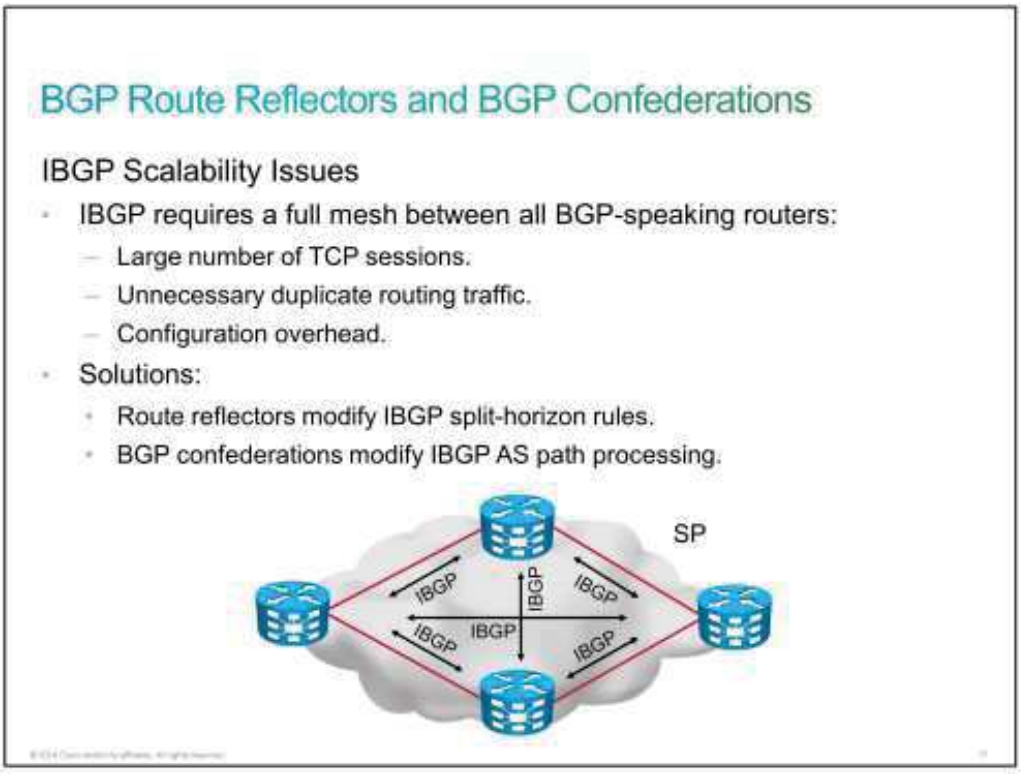
- Describe network design with BGP route reflectors
- Describe hierarchical route reflectors
- Explain the configuration of route reflectors in service provider networks
- Describe BGP confederations
- Describe AS path propagation within BGP confederations
- Describe intraconfederation EBGP session properties

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

BGP Route Reflectors and BGP Confederations

Classic IBGP split-horizon rules specify that updates that are received on an EBGP session should be forwarded on all IBGP and EBGP sessions, but updates that are received on an IBGP session should be forwarded only to all EBGP sessions. This rule requires a BGP boundary router to be able to send routing updates to all other BGP-speaking routers in its own AS directly through a separate IBGP session to each of them.



The primary reason for the IBGP split-horizon rule is to avoid routing information loops within the AS. If the information that is received through an IBGP session is forwarded on other IBGP sessions, the information might come back to the originator and be forwarded again in a never-ending loop. The originator would not detect the loop because no BGP attributes are changed on IBGP sessions.

The general design rule in classic IBGP is to have a full mesh of IBGP sessions. But a full mesh of IBGP sessions between *n* number of routers would require (n * (n – 1)) / 2 IBGP sessions. For example, a router with an AS that contains 10 routers would require (10 * (10 – 1)) / 2 = 45 IBGP sessions. Imagine the number of sessions (and the associated router configuration) that would be required for a single AS containing 500 routers.

Every IBGP session uses a single TCP session to another IBGP peer. An update that must be sent to all IBGP peers must be sent on each of the individual TCP sessions. If a router is attached to the rest of the network over just a single link, this single link has to carry all TCP/IP packets for all IBGP sessions. This requirement results in multiplication of the update over the single link.

Two different solutions are available to achieve greater scalability when you are faced with the full-mesh rules of IBGP autonomous systems:

- Route reflectors modify the classic IBGP split-horizon rule and allow a particular router to forward incoming IBGP updates to an outgoing IBGP session under certain conditions. This router becomes a concentration router, or a route reflector.

https://t.me/learningnets

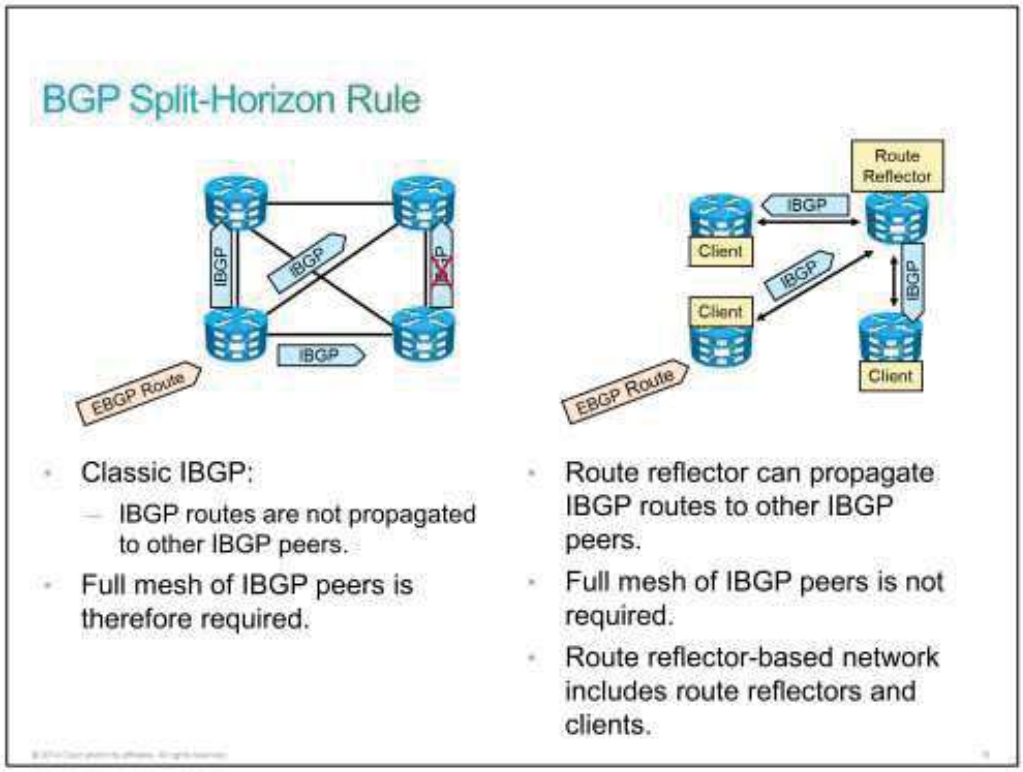
- BGP confederations introduce the concept of a number of smaller autonomous systems within the original AS. The small autonomous systems exchange BGP updates between them using intraconfederation EBGP sessions.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

BGP Split-Horizon Rule

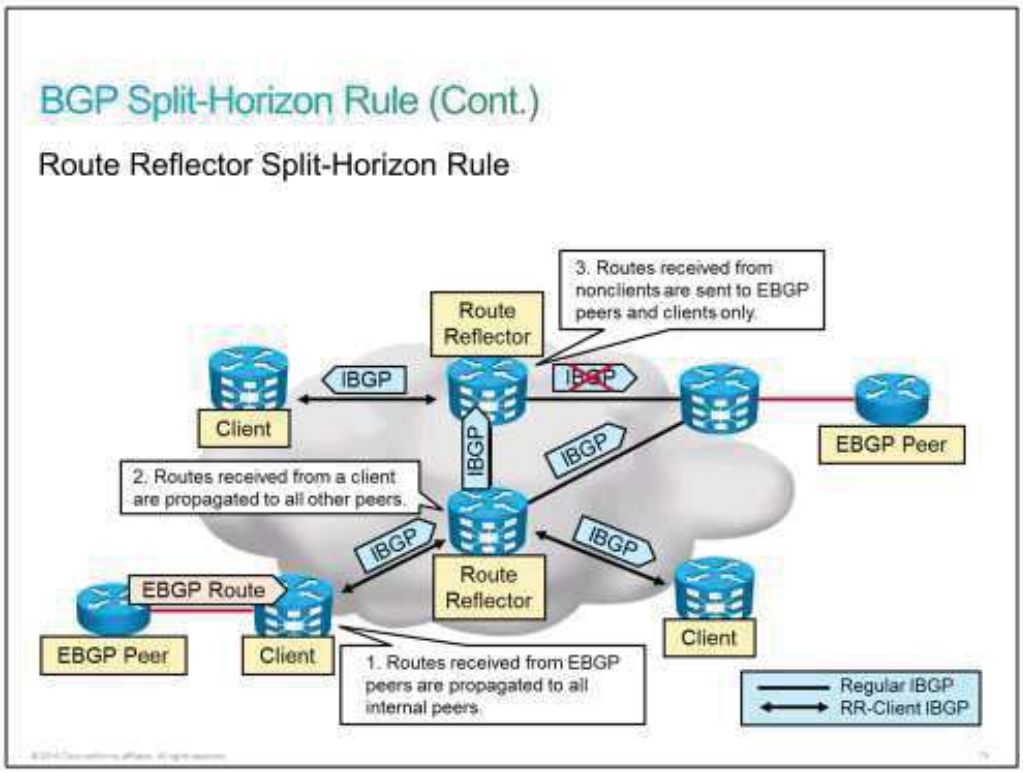
In classic IBGP, the IBGP split-horizon rules mandate that updates received on EBGP sessions should be forwarded on all IBGP and EBGP sessions, but updates received on an IBGP session should be forwarded only on all EBGP sessions. This requires the BGP edge or boundary router to send updates to all other BGP enabled routers in its own AS directly through individual IBGP sessions to each BGP router. To allow every router to update every other router, a full mesh of IBGP sessions is required.



The IBGP route reflector design reduces the need for a full mesh. The router that is configured as a route reflector, under certain conditions, will relay updates that are received through an IBGP session to another IBGP session. This capability requires modifications of the classic IBGP split-horizon rules.

A route reflector-based network includes route reflectors that act as concentrators for IBGP sessions and clients.

https://t.me/learningnets



When you implement a route-reflector-based IBGP network, the BGP routers are divided into route reflectors (which implement modified split-horizon rules) and clients (which behave like traditional IBGP routers).

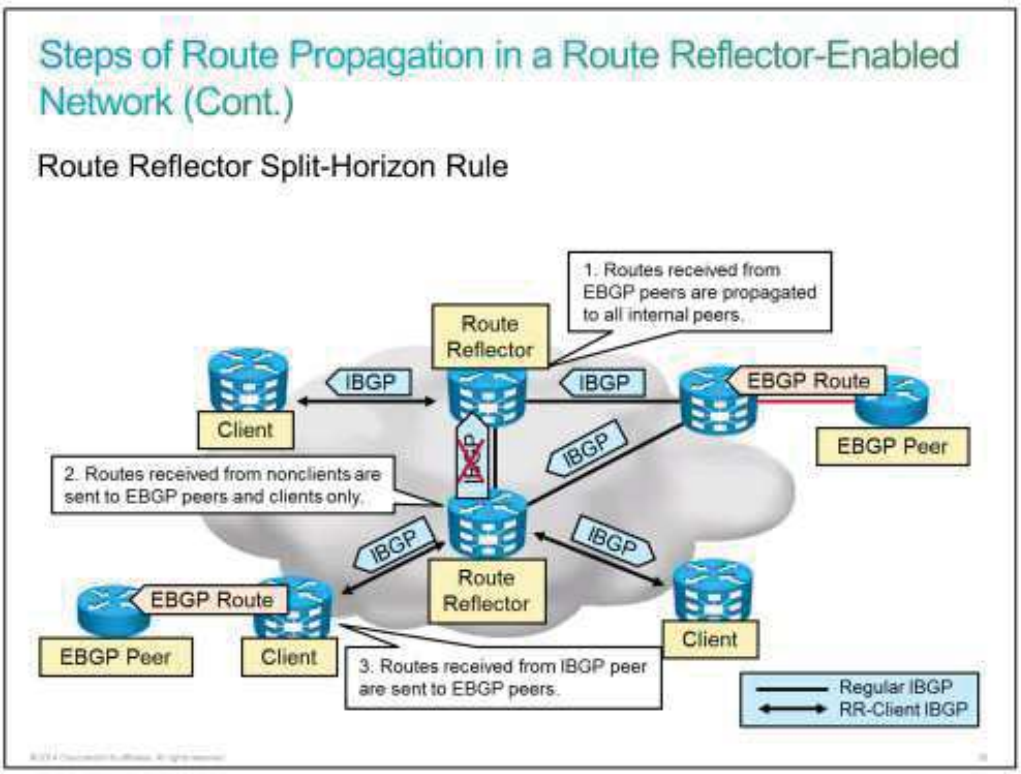
https://t.me/learningnets

Steps of Route Propagation in a Route Reflector-Enabled Network

Route reflector clients are excluded from the full mesh. They can have any number of EBGP sessions but may have only one IBGP session, the session with their route reflector. The following steps describe route propagation in a route reflector-enabled network:

1. Clients conform to the classic IBGP split-horizon rules and forward a received route from EBGP on their IBGP neighbor sessions.
2. The route reflector conforms to the route reflector split-horizon rules and recognizes that it has an IBGP session to a client. When the IBGP update is received from the client, the route reflector forwards the update to all other BGP neighbors, therefore alleviating the IBGP full-mesh requirement for its clients.
3. If a route reflector receives a route from nonclients (from another route reflector, in the example), the route is sent only to EBGP peers and clients. This is the reason why the route is not sent to the classic BGP neighbor in the example.

Forwarding of an IBGP update in a route reflector does not change the next-hop attribute or any other common BGP attribute. This feature means that the client will use the optimum route by means of recursive routing, regardless of the way that it has received the BGP route.



https://t.me/learningnets

Example: Route Propagation in Route Reflector-Enabled Network

Another example of route propagation in a route reflector-enabled network:

- Nonclients conform to the classic IBGP split-horizon rules and forward a received route from EBGP on their IBGP neighbor sessions.
- The bottom route reflector that receives a route from a nonclient sends the route to EBGP peers and clients only. This is the reason why the IBGP update is not sent to the top route reflector as well.
- The client receives the IBGP update and sends it to the EBGP peers.

The table presents detailed IBGP split-horizon rules that are modified by the introduction of BGP route reflectors. For purposes of definition, a route reflector is a BGP speaker that can advertise IBGP learned routers to another IBGP peer and, hence, can reflect routes. IBGP peers of the route reflector fall under two categories: clients and nonclients. The route reflector and its clients form a cluster. All IBGP peers of the route reflector that are not part of the cluster are nonclients. A "classic" IBGP router is a router that does not support route reflector functionality.

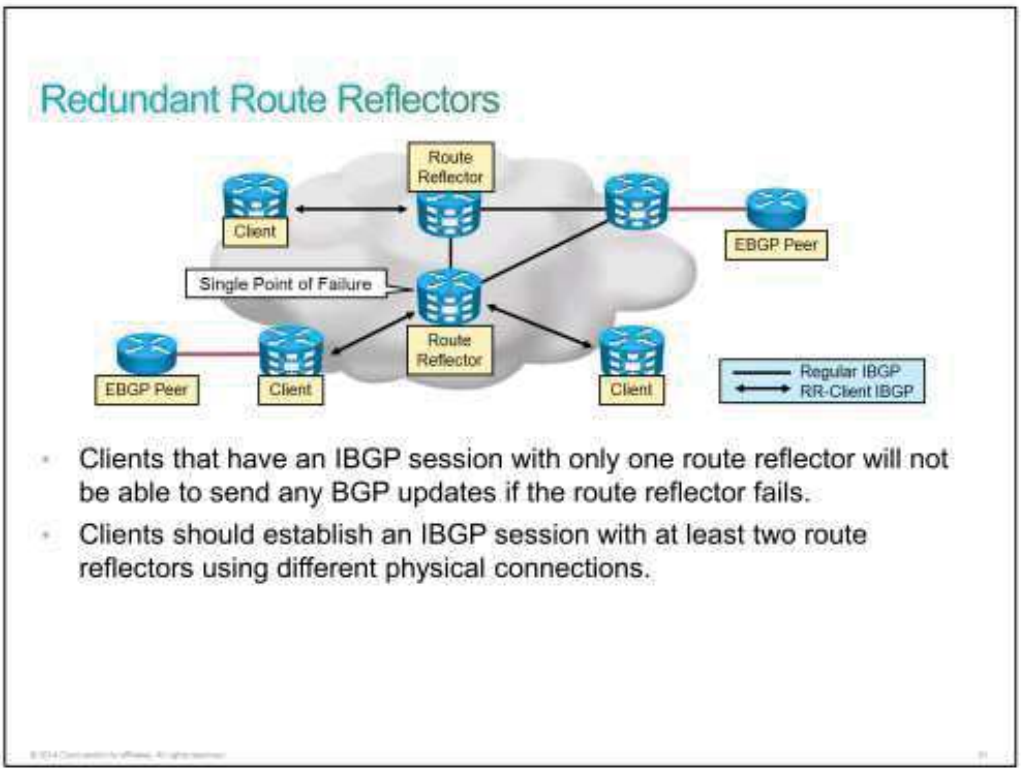
Type of Receiving Router	Incoming Update From	Is Forwarded To
Classic (nonclient)	EBGP peer	All peers (IBGP and EBGP)
	IBGP peer	EBGP peers
Route reflector	EBGP peer	All peers (IBGP and EBGP)
	Nonclient IBGP peer	EBGP peers and clients
	Client IBGP peer	All peers but the sender
Client	EBGP peer	All peers (IBGP and EBGP)
	IBGP peer	EBGP peers

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Redundant Route Reflectors

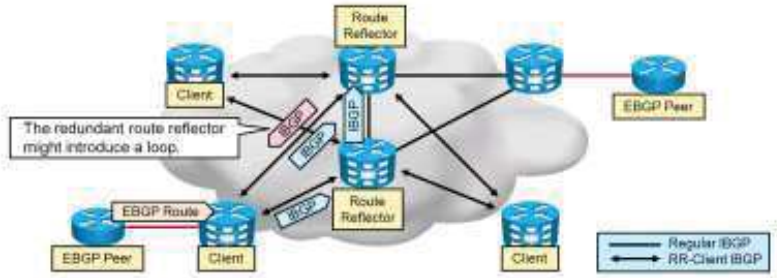
Clients may have any number of EBGP peers but may have IBGP sessions only with their route reflector or reflectors. If the reflector fails, its client can no longer send BGP updates to, or receive them from, the rest of the AS. The route reflector is, therefore, a single point of failure.



To avoid introducing a single point of failure into the network, the route reflector functionality must be as redundant as the physical network. If a client will still be physically attached to the network after its route reflector has failed, the client should have a redundant route reflector. Thus, in all highly available networks, route reflectors must be redundant.

https://t.me/learningnets

Redundant Route Reflectors (Cont.)



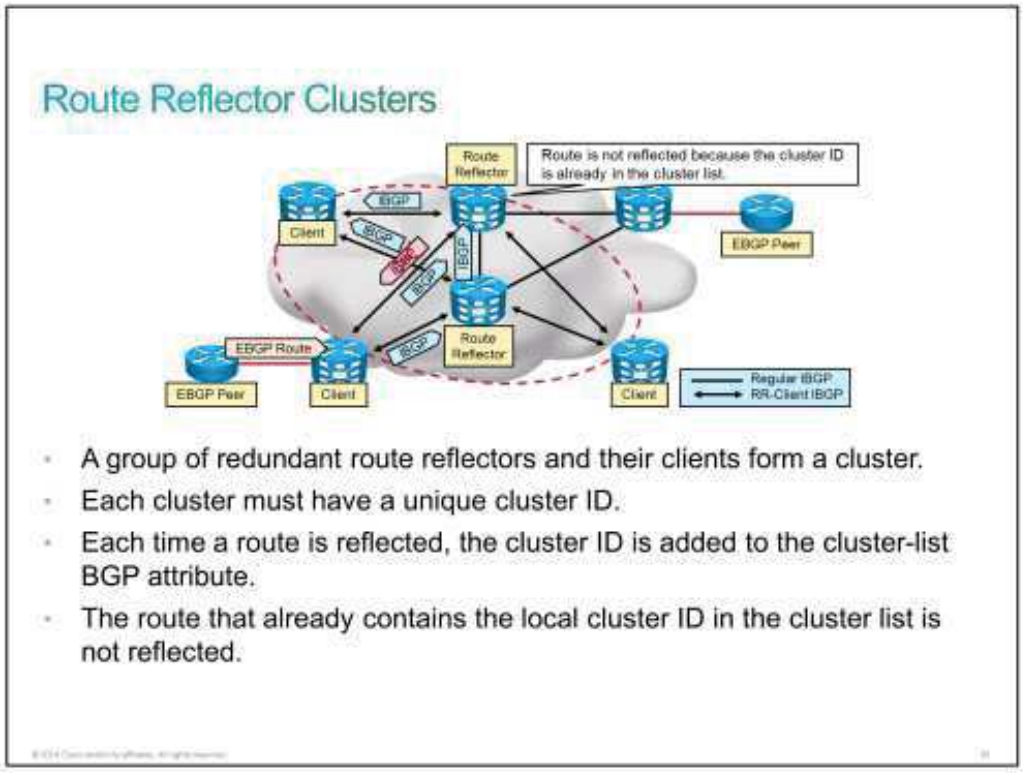
- Redundant reflectors solve the high-availability requirement.
- The concept of clusters is introduced to prevent IBGP routing loops between route reflectors.

A client may have IBGP sessions to more than one route reflector to avoid a single point of failure. Each client will receive the same route from both of its reflectors. Both route reflectors will receive the same IBGP update from their client, and they will both reflect the update to the rest of the clients. Additionally, both route reflectors will get updates from the full mesh and reflect those updates to their clients. As a result, each client will get two copies of all routes. Under certain circumstances, particularly when you use weights on IBGP sessions to influence BGP route selection, improper route reflection can result in an IBGP routing loop that is impossible to detect. Additional BGP attributes are thus necessary to prevent these routing loops.

https://t.me/learningnets

Route Reflector Clusters

A router that is acting as a route reflector client does not require any specific configuration. It simply has fewer IBGP sessions than it would have if it were part of the full mesh. But improperly configuring the client to also be a reflector could easily cause a loop. An IBGP route coming in from one of the real reflectors to the client could be forwarded by the client, erroneously acting as reflector, to the other reflector.



Route reflector clusters prevent IBGP routing loops in redundant route reflector designs.

The role of the network designer is to properly identify which route reflectors and their clients will form a cluster. The designer assigns to the cluster a cluster ID number that is unique within the AS.

Note	The cluster ID number must be configured in the route reflectors. The clients should not be configured with this information.

A route reflector router can reflect routes only within a single cluster. A route reflector can, however, participate in another cluster but only as a client. A client can function as a client only to a route reflector belonging to the same cluster.

When a route is reflected, the reflector creates the cluster-list attribute and attaches it to the route if it does not already exist. It then sets its cluster ID number in the cluster list or adds its cluster ID number to an already existing cluster-list attribute. If the route, for any reason, is ever reflected back to the same reflector, it will recognize its cluster ID number in the cluster list and not forward it again. The first route reflector that reflects the route also sets an additional BGP attribute, called originator-ID. This attribute carries the router ID of the originator of the route.

https://t.me/learningnets

Note	The cluster-list and originator-ID attributes are nontransitive optional BGP attributes, allowing routers that do not support route reflector functionality to coexist with route reflectors and their clients in the same AS.

Based on cluster-list and originator-ID attributes, routers can implement two loop-prevention mechanisms:

- Any router that receives an IBGP update with the originator-ID attribute set to its own BGP router ID will ignore that update (reflected route was propagated back to its originator).
- Any route reflector that receives an IBGP update with its cluster ID already in the cluster list will ignore that update (route reflector loop).

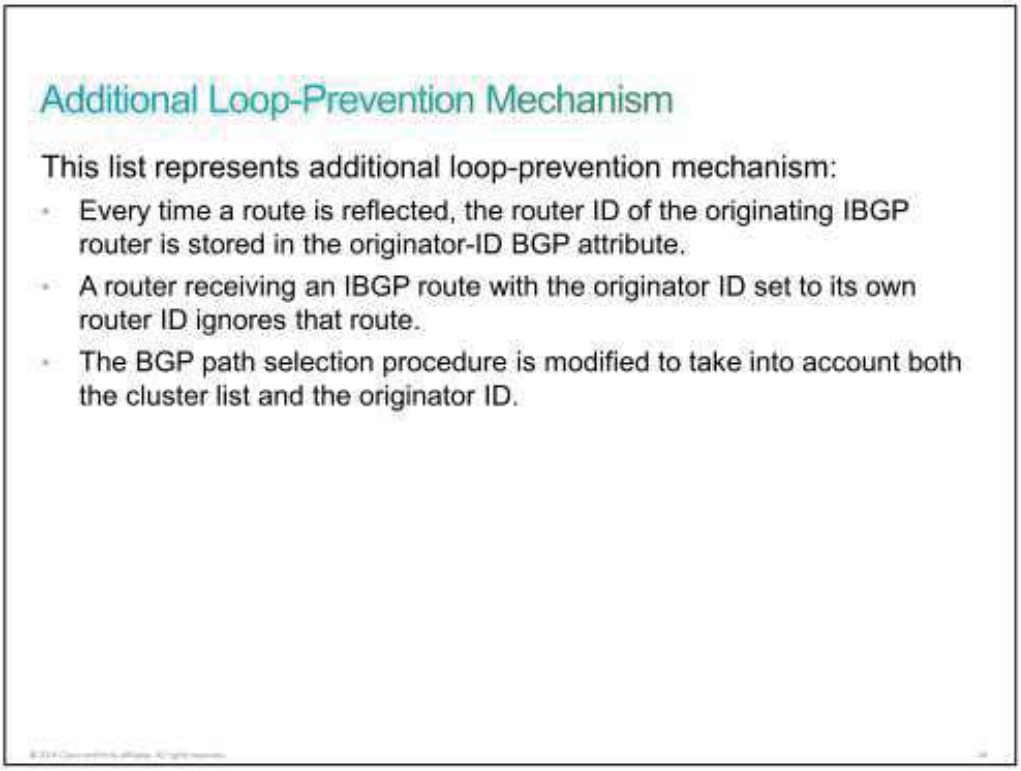
In the example, the client in the cluster forwards the received EBGP update to both reflectors. The route reflectors forward the update into the IBGP full mesh. This behavior means that they send the update to each other as well. But when a route reflector receives a BGP update from another route reflector, it recognizes their common cluster ID number in the cluster-list attribute. Therefore, the newly received route update is ignored.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Additional Loop-Prevention Mechanism

The second BGP update loop prevention mechanism is using the originator-ID attribute.



When a route is reflected, the route reflector sets the originator-ID BGP attribute (a nontransitive optional BGP attribute) to the router ID of the peer from which it received the route. Any router that receives a route with its own router ID in the originator-ID attribute silently ignores that route.

BGP path selection rules have been modified to select the best route in scenarios where a router might receive reflected and nonreflected routes or several reflected routes:

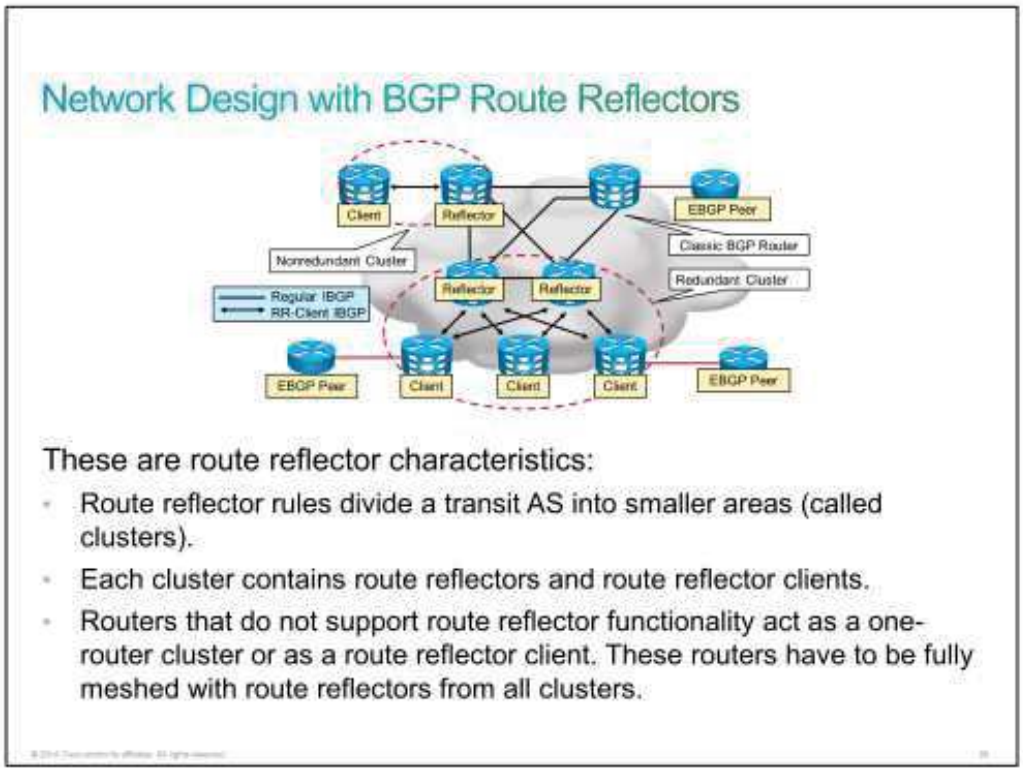
- The traditional BGP path selection parameters, such as weight, local preference, origin, and MED, are compared first.
- If these parameters are equal, the routes that are received from EBGP neighbors are preferred over routes that are received from IBGP neighbors.
- When a router receives two IBGP routes, the nonreflected routes (routes with no originator-ID attribute) are preferred over reflected routes.
- The reflected routes with shorter cluster lists are preferred over routes with longer cluster lists.

If the additional route-reflector-oriented selection criteria do not yield a decision, the rest of the traditional BGP path selection rules are followed.

https://t.me/learningnets

Network Design with BGP Route Reflectors

Implementing route reflectors within the transit AS will create smaller areas (clusters) of routers. A cluster consists of route reflector routers, either redundant or nonredundant, and the client routers that are connected to them.



In designing the implementation of route reflectors within a transit AS, identify a group of peripheral routers that are physically connected to the same backbone router or routers. Consider the peripheral routers as clients and the backbone routers as route reflectors. Then, consider that this group of routers forms a cluster. In the Cisco IP Next-Generation Networks, BGP route reflectors are implemented in the IP Infrastructure layer on core service provider devices.

If a router does not support the route reflector functionality (which is usually not the case anymore), the router can either act as a route reflector client or as a standalone (one router cluster) router. Standalone routers have to be fully meshed with all IBGP peers (or route reflectors representing a cluster) in order to receive all routing updates.

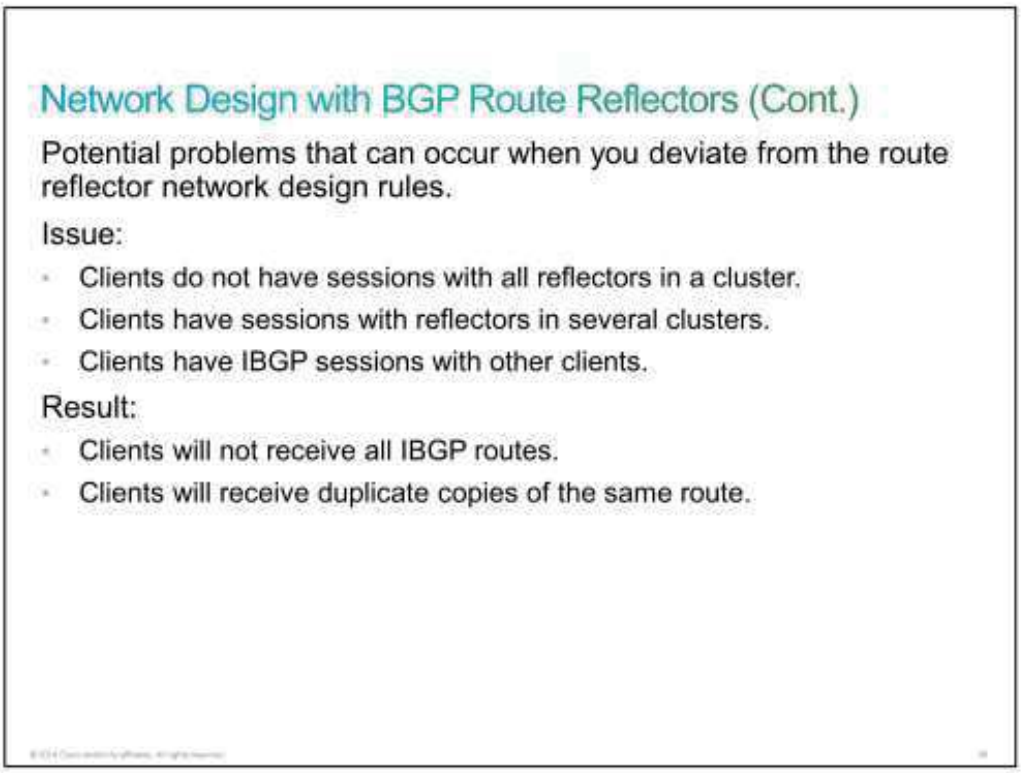
In the example, there are two clusters and one standalone BGP routers. In the area that is labeled "redundant cluster," the three client routers and the two route reflector routers make up the cluster. Each of the three client routers has an IBGP session with the two route reflectors and only with those two route reflectors.

In the nonredundant area, the client router has a single physical connection to a route reflector router. These two routers form a nonredundant cluster. The router that is designated as the route reflector in the cluster is already a single point of failure in this physical design, because a failure of this router will prevent the clients in the cluster from reaching the rest of the network. Therefore, there is no new single point of failure that is introduced when the router is configured as the only route reflector in this cluster. The client has a single IBGP session to the route reflector.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

The standalone router that is shown is not configured as a route reflector, nor is it a client to any other route reflector. This other router serves as an example of where a non-route-reflector router participates in the full mesh. This router has a full mesh of IBGP sessions with all route reflectors, each reflector representing its cluster.



Two nontransitive optional BGP attributes, originator-ID and cluster-list, are both used to prevent fatal loops of information. The use of these two attributes makes a network fairly insensitive to poor configuration. However, for optimal performance, you must have an optimal configuration. Here are some of the problems that could occur if you deviate from route reflector network design rules:

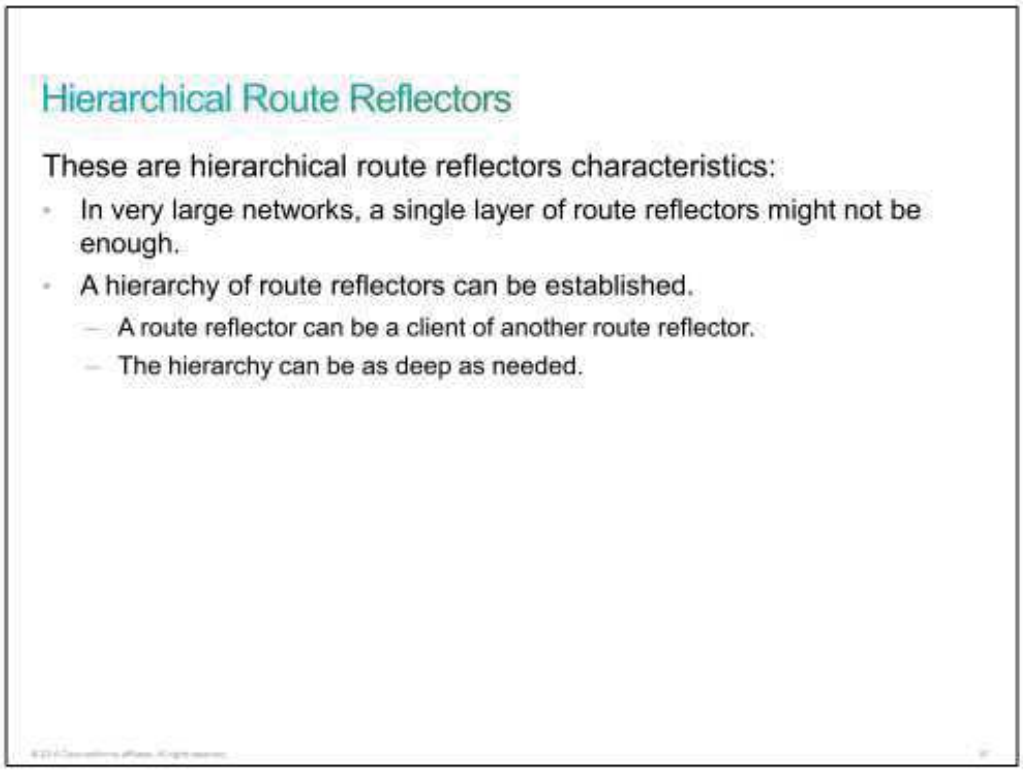
- If route reflectors are not connected with IBGP sessions in a full mesh, some clusters will not have all the routes.
- If a client has IBGP sessions with some route reflectors in a cluster, but not with all of them, the client might miss some BGP routes.
- If a client has IBGP sessions to route reflectors that belong to different clusters, the BGP update from the client will be forwarded by the client into the full mesh with different cluster IDs in the cluster-list attribute. When the BGP update enters the mesh, it will reach the other route reflector, which will, unnecessarily, accept the route as valid and forward it into its cluster. This situation, in turn, causes unnecessary duplication of updates to the clients.

If a client has IBGP sessions to other clients in the same cluster, those clients will receive unnecessary duplications of updates.

https://t.me/learningnets

Hierarchical Route Reflectors

Network designers can build route reflector clusters in hierarchies. With hierarchies, a router that serves as a route reflector in one cluster can act as a client in another cluster.

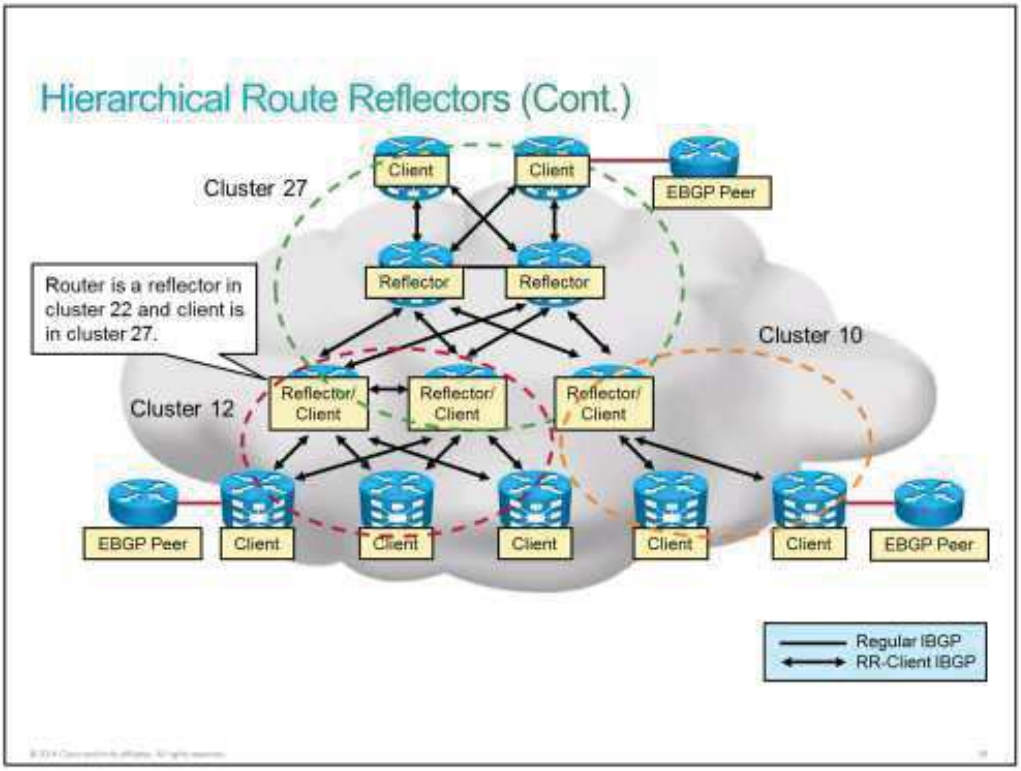


In configuring an IBGP session on a route reflector, the designer must configure the session to reach a client in order for the route reflector IBGP split-horizon rules to start working. All other IBGP sessions that are configured on the route reflector are a part of the full mesh. Also, the designer must configure the cluster ID on the route reflector.

A router that is configured to be a route reflector will still have ordinary IBGP sessions that are part of the full mesh. If these sessions are reduced in number and only a few remain, and the remaining ones reach a second level of route reflectors, a hierarchy of route reflectors is created. When a designer builds a first level of clusters, the remaining full mesh is smaller than when all routers belonged to the full mesh. But if the remaining full mesh is still large enough, the designer can build an additional level of route reflectors.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets



The figure shows an example of hierarchical route reflectors.

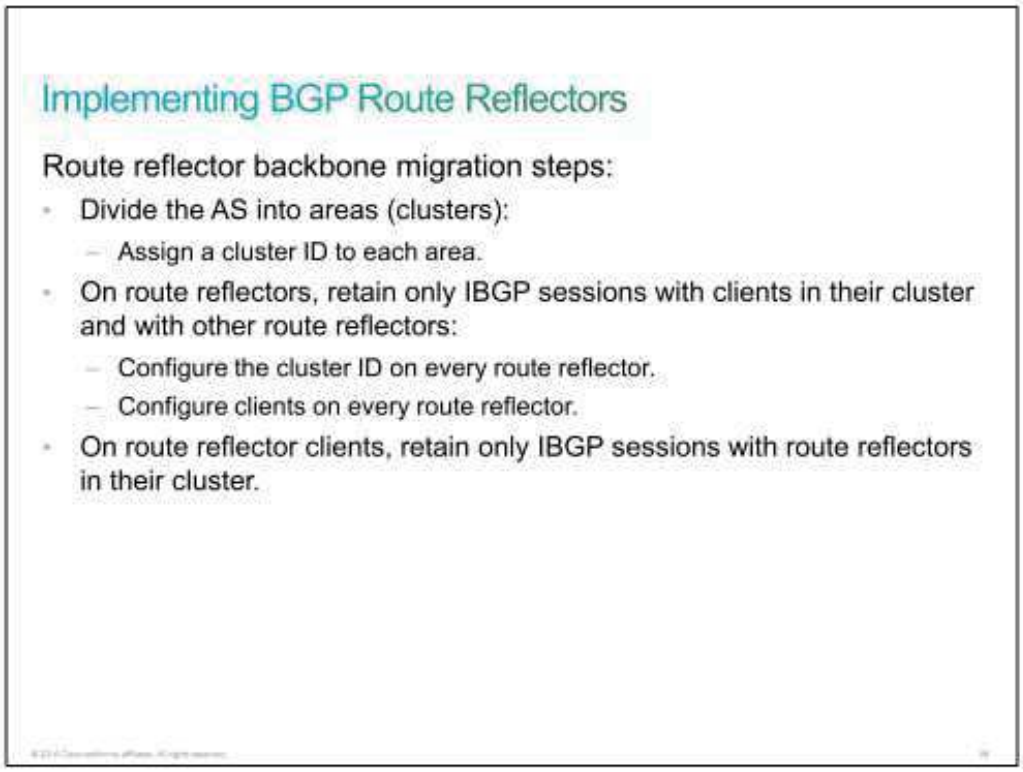
The first level of hierarchy reduced the original full mesh of twelve routers (all routers in the service provider network) to a full mesh of seven routers (the lower three route reflectors and the upper two route reflectors and two clients). The second level of route reflector clusters was built by creating cluster 27. This second step further reduced the full mesh of seven routers to a full mesh consisting of only two routers (upper route reflectors). Only the two route reflectors in cluster 27 should be connected in a full mesh.

When a client in the lowest level receives an EBGP update, it will forward the update on all configured IBGP sessions to a route reflector. The route reflector recognizes BGP updates that are received from configured clients and will forward these updates to all other clients that use normal IBGP sessions. The update, sent on a normal IBGP session, will be a second-level client update to the second-level route reflector. The second-level route reflector will recognize that the update was received from a client, and will forward it to all other clients and into the full mesh.

https://t.me/learningnets

Implementing BGP Route Reflectors

The physical topology of the AS serves as a guide to designing clusters. You should not introduce any additional single points of failure when you are deploying route reflectors.



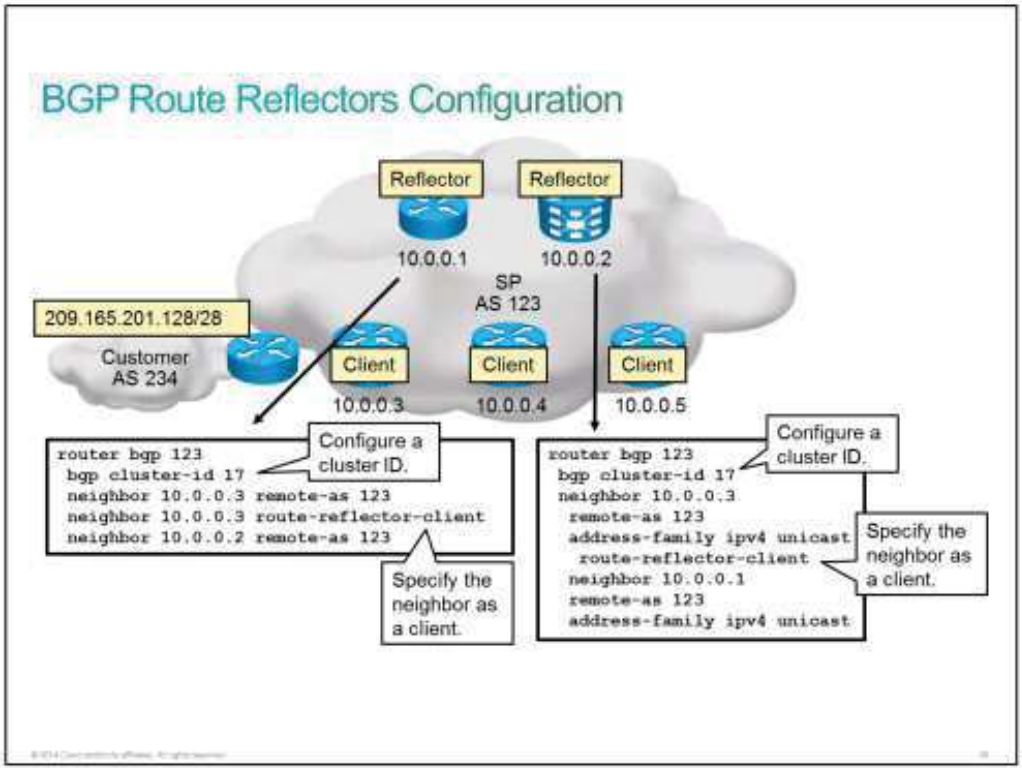
If the physical topology is redundant, a good practice is to have redundant route reflectors. If the physical topology is not redundant, introducing a nonredundant cluster does not add a single point of failure because the network was already nonredundant.

The following planning and preparation steps are required before you migrate from a full mesh of IBGP sessions to a route reflector design:

1. Identify a group of peripheral routers that are physically connected to the same set of backbone routers. Consider the peripheral routers as clients and the backbone routers as route reflectors. Let the routers form a cluster. Make sure that no router belongs to two different clusters, because this setup would represent an illegal configuration. Create a numbering plan that indicates how numbers are assigned to the clusters in the network. The plan must make sure to uniquely identify each of the clusters within the AS.
2. On the route reflectors, configure the cluster ID and configure the clients as route reflector clients.
3. On the route reflector clients, remove all IBGP sessions to neighbors that are not route reflectors in their cluster.

https://t.me/learningnets

BGP Route Reflectors Configuration



The figure shows an example of the configuration of two route reflectors. Routers at 10.0.0.1 and 10.0.0.2 are configured as redundant route reflectors and are configured with each other as a classic IBGP peer. The router with the 10.0.0.2 IP address runs Cisco IOS XR Software.

To configure the cluster ID, enter the BGP router configuration mode and set the cluster ID using the **bgp cluster-ID** command, followed by the cluster ID. The cluster ID can be entered either as an IP address or value. The range is 1 to 4294967295.

To configure an IBGP peer as a route reflector client, enter the neighbor configuration mode using the **neighbor** command and then specify that peer as a client, using the **route-reflector-client** command under the proper address-family configuration mode.

The router with the 10.0.0.1 IP address runs Cisco IOS XE Software. The configuration is similar to the configuration of the Cisco IOS XR Software.

Configuration for IPv6 is similar, with IPv6 addresses specified as neighbors. On the Cisco IOS XR Software routers, you have to select the IPv6 unicast address family before configuring route reflector clients. On the Cisco IOS and IOS XR Software routers, you have to enter the IPv6 address family before specifying a neighbor as a route reflector client, and you have to activate neighbors for the IPv6 address family.

Note	Because BGP is multiprotocol, it should in theory be able to transfer IPv4 and IPv6 prefixes using either IPv4 or IPv6 as the transport protocol. However, on the Cisco IOS XR Software router, IPv4 prefixes can only be transferred using IPv4, and IPv6 prefixes can only be reliably transferred using IPv6.

https://t.me/learningnets

BGP Route Reflectors Verification

BGP Route Reflectors Verification

```
RP/0/RSP0/CPU0:P# show bgp neighbors 10.0.0.3
BGP neighbor is 10.0.0.3
 Remote AS 64500, local AS 64500, internal link
 Remote router ID 10.0.0.3
 Cluster ID 17
  BGP state = Established, up for 00:00:07
<… output omitted …>
For Address Family: IPv4 Unicast
  BGP neighbor version 17
  Update group: 0.1 Filter-group: 0.1 No Refresh request being processed
  Route-Reflector Client
  NEXT_HOP is always this router
<… output omitted …>
```

- Displays information about the BGP session with the neighbor.

The **show bgp neighbors** command, issued on the route reflector router, indicates that the neighbor is a route reflector client.

BGP Route Reflectors Verification (Cont.)

```
RP/0/RSP0/CPU0:P# show bgp 209.165.201.128
BGP routing table entry for 209.165.201.128/28
<… output omitted …>
234, (Received from a RR-client)
   10.0.0.3 (metric 2) from 10.0.0.3 (10.0.0.3)
        Origin IGP, metric 0, localpref 100, valid, internal, best, group-best
        Received Path ID 0, Local Path ID 1, version 3
```

- Displays routes received from the client as seen on the reflector

```
RP/0/RSP0/CPU0:PE# show bgp 209.165.201.128
BGP routing table entry for 209.165.201.128/28
<… output omitted …>
234
   10.0.0.1 (metric 2) from 10.0.0.1(10.0.0.1)
        Origin IGP, metric 0, localpref 100, valid, internal
        Received Path ID 0, Local Path ID 0, version 0
        Originator: 10.0.0.1, Cluster list: 0.0.0.17
```

- Displays reflected routes as seen on the client

The **show bgp** command displays entries in the BGP table. When this command is issued on the route reflector for a route received from a client, the printout shows that the route has been received from a client.

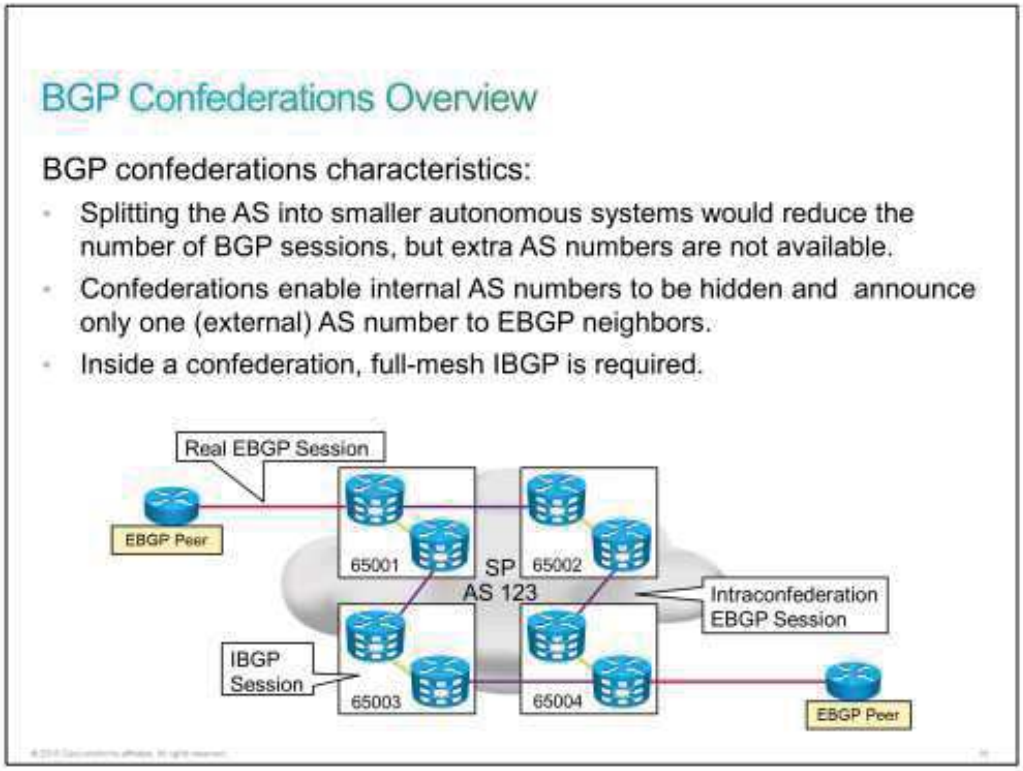
https://t.me/learningnets

When this command is issued on the route reflector client for a route that has been reflected, the printout shows originator-ID and cluster-ID list as BGP attributes.

https://t.me/learningnets

BGP Confederations Overview

A large number of routers in a large transit AS would traditionally introduce a complex full-mesh structure of IBGP sessions. By splitting the AS into a number of small autonomous systems, you can provide each one of the small systems with a fairly simple IBGP structure. Interconnections between these autonomous systems could then be made using EBGP, which allows for arbitrary topologies.



Splitting an AS into smaller autonomous systems requires many official AS numbers, which are a scarce resource. However, by introducing the BGP confederation, you can enable a large AS to be partitioned into a number of smaller autonomous systems (called member autonomous systems), where each is internal to the larger AS. The AS numbers of each member AS that is used within the confederation are never visible from outside the confederation itself. This invisibility allows private AS numbers (in the range 64512 to 65535) to be assigned to autonomous systems inside a confederation to identify a member AS, without the need to coordinate AS number assignments with an official AS delegation authority.

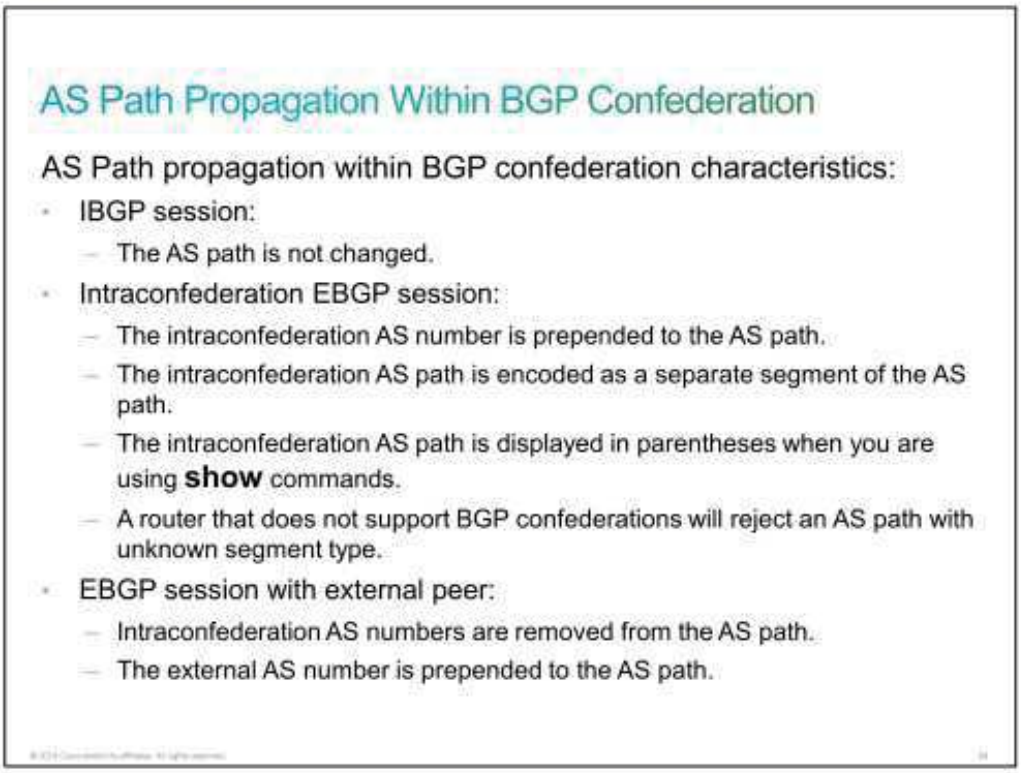
Within a member AS, the classic IBGP rules apply. Therefore, all BGP routers inside the member AS must still maintain a full mesh of BGP sessions.

EBGP sessions are established between member autonomous systems inside a confederation. These EBGP sessions behave slightly differently from classic EBGP sessions, and are therefore named intraconfederation EBGP sessions to differentiate them from true EBGP sessions.

https://t.me/learningnets

AS Path Propagation Within BGP Confederation

The mandatory well-known BGP attribute AS-path is modified on EBGP and intraconfederation EBGP sessions. The sender prepends its own AS number to the AS path whenever an EBGP update is sent. When a BGP update traverses the Internet, every AS that it passes through is recorded in the AS path. If the update, for any reason, comes back to an AS in which it has already been, the receiving router recognizes its own AS in the AS path and silently ignores the update. This mechanism prevents information loops and allows arbitrary topology when you are interconnecting autonomous systems.



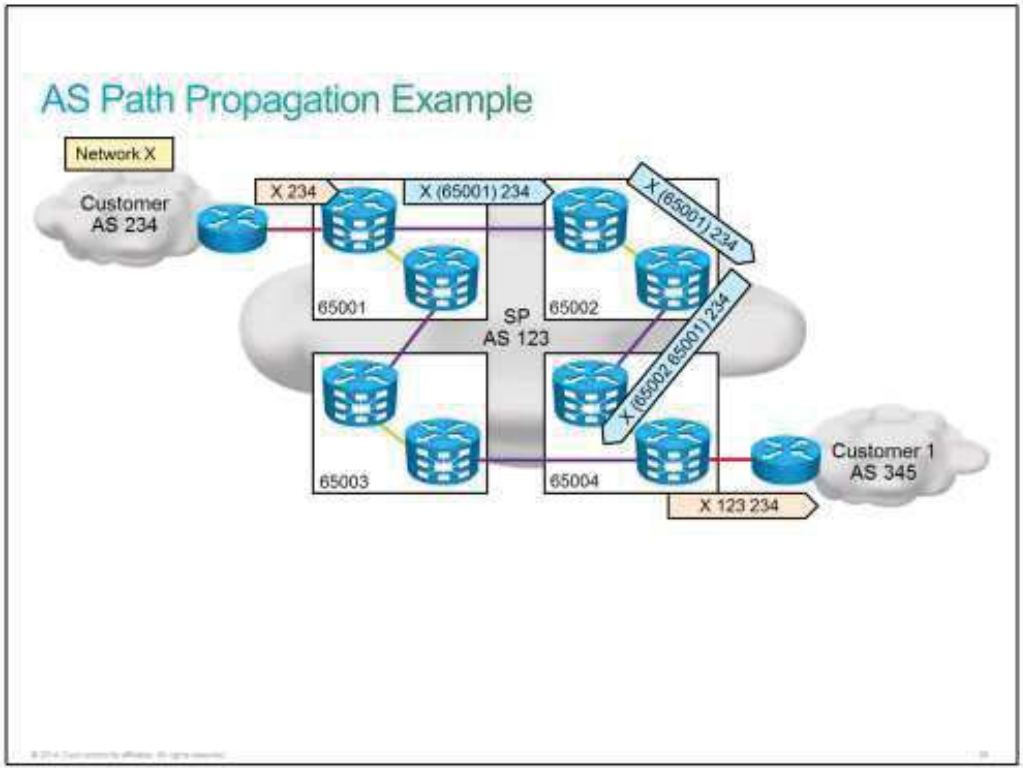
IBGP sessions do not modify the AS-path attribute, so the topology within each AS is limited to the full mesh, and the propagation of BGP updates across multiple IBGP sessions is prohibited.

When a router sends a BGP update over an intraconfederation EBGP session, it prepends the member-AS number to the AS path. This information is maintained by the routers within the confederation and prevents routing information loops inside the confederation.

When a router sends a BGP update over a true EBGP session to an AS that is outside of the confederation, it removes the part of the AS path that describes the member-AS numbers and prepends the official AS number to the AS path. As a result, the confederation appears as one single AS to the outside world.

https://t.me/learningnets

AS Path Propagation Example



The figure shows an example of BGP confederation. Network X originates inside AS 234 and is announced by the edge router in AS 234 over a true EBGP session to the ingress router in the confederation. The edge router in AS 234 determines that the edge router that it is communicating with resides in AS 123. AS 234 prepends its assigned AS number in the AS path (which was previously empty, because network X originated in AS 234). When the EBGP update arrives at the confederation member AS 65001, the AS-path attribute has been set to 234.

The member AS 65001 has intraconfederation EBGP sessions to member AS 65002 and to member AS 65003. The router in AS 65001 prepends its own AS number to the AS path. When doing this, it signals that this part of the AS path describes the intraconfederation AS path. When printed out, the intraconfederation part of the AS path is displayed within parentheses. Therefore, when member AS 65001 sends the update to member AS 65002 and member AS 65003, the route to network X has an AS-path attribute set to (65001) 12.

Within a member AS, the router sends the update using classic IBGP. The router does not modify the AS path when transmitting it over an IBGP session.

Member AS 65002 and member AS 65003 both prepend their AS numbers, so member AS 65004 receives the update about the route to X via two different paths, one with AS path (65002 65001) 234 and the other with AS path (65003 65001) 234.

Member AS 65004 selects one of the alternatives as the best BGP route. It then forwards this update on the intraconfederation EBGP session. This update could introduce a loop, but if the update were ever to be forwarded all the way back up to member AS 65001, the loop would be detected and member AS 65001 would silently ignore the update.

Member AS 65004 also forwards the update about network X on a true EBGP session to AS 345. When it does, it removes all the parenthetical information in the AS path and replaces it with the official AS number of the confederation, 123. AS 345 thus receives the update about network X with an AS path that is set to 123 234.

 ©2014 Cisco Systems, Inc.

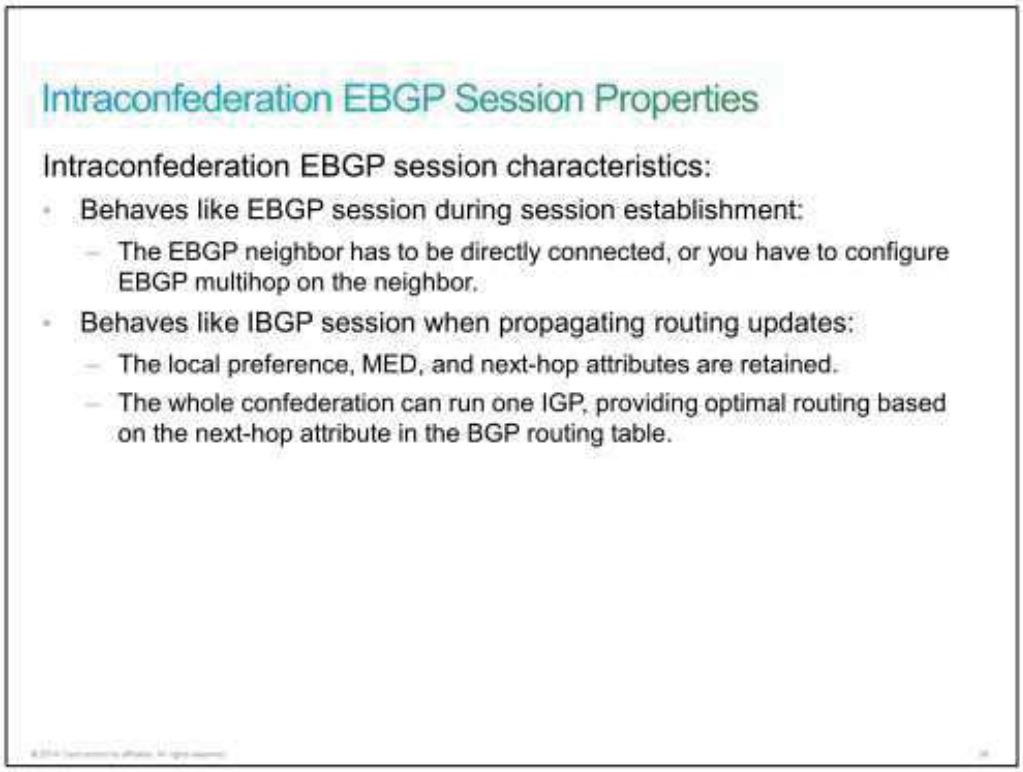
https://t.me/learningnets

If no other policy is configured, the routers in AS 345 select the best path that is based on the length of the AS path. AS 345 will see the route to X with an AS path length of two. When AS 345 forwards packets that are destined for network X into the confederation (AS 123), member AS 65004 must make a forwarding decision. This decision process inside the confederation will use both confederate AS paths for loop avoidance, but not for choosing the shortest AS path within the confederation. The multiple-step BGP path selection process will treat both AS paths as equal and have to use the other attributes to select the preferred path. All else being equal, the BGP decision process chooses normal EBGP routes over confederation EBGP routes, and confederation EBGP over IBGP routes.

https://t.me/learningnets

Intraconfederation EBGP Session Properties

Intraconfederation EBGP sessions, while having EBGP-like properties (for example, updating the AS-path attribute when propagating BGP routes), still run inside a real AS and therefore have to share some properties with IBGP sessions to achieve the same end results. Similar to IBGP sessions, the BGP attributes of local preference, MED, and next-hop are not changed in updates that are propagated across intraconfederation EBGP sessions. All routers in all member autonomous systems inside the confederation consequently use the same next-hop address when they perform recursive routing. Because all intraconfederation routers use the same next-hop address, the entire confederation should use the same IGP to resolve the BGP next-hop address. The IGP information should not be limited by the member AS boundary.



On the other hand, intraconfederation EBGP sessions behave exactly like EBGP sessions when they are established. EBGP sessions are normally opened between directly connected interfaces. However, because all routers within the confederation run the same IGP and exchange internal routing information, they are easily able to open multihop sessions. Resiliency of BGP sessions and consequent stability of BGP routing are introduced into the network if the intraconfederation EBGP sessions are established between loopback interfaces, just like IBGP sessions normally are. When intraconfederation EBGP sessions are opened between loopback interfaces, the **ebgp-multihop** qualifier must be given to the session. Otherwise, the EBGP session will never leave the idle state.

https://t.me/learningnets

Summary

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

Summary

- BGP reflectors are one of two ways to increase IBGP network scalability.
- The IBGP split-horizon rule prevents loops in a fully meshed IBGP cloud.
- A single route reflect is a single point of failure, therefore redundant configuration is essential.
- Route reflectors can be combined into clusters for redundancy.
- The originator-ID BGP attribute is used to prevent routes from being reflected.
- BGP route reflectors allow constructions of flexible network designs.
- BGP route reflector clusters can be nested to many levels.
- BGP route reflectors are configured on reflectors only; clients do not know they are clients.

Summary (Cont.)

- IGP confederations allow the creation of AS domains within AS domains.
- The intraconfederation AS number is prepended to the AS path and is displayed in parentheses.
- BGP intraconfederation interneighbor peering sessions combine the properties of both eBGP and IBGP peering sessions at the same time.

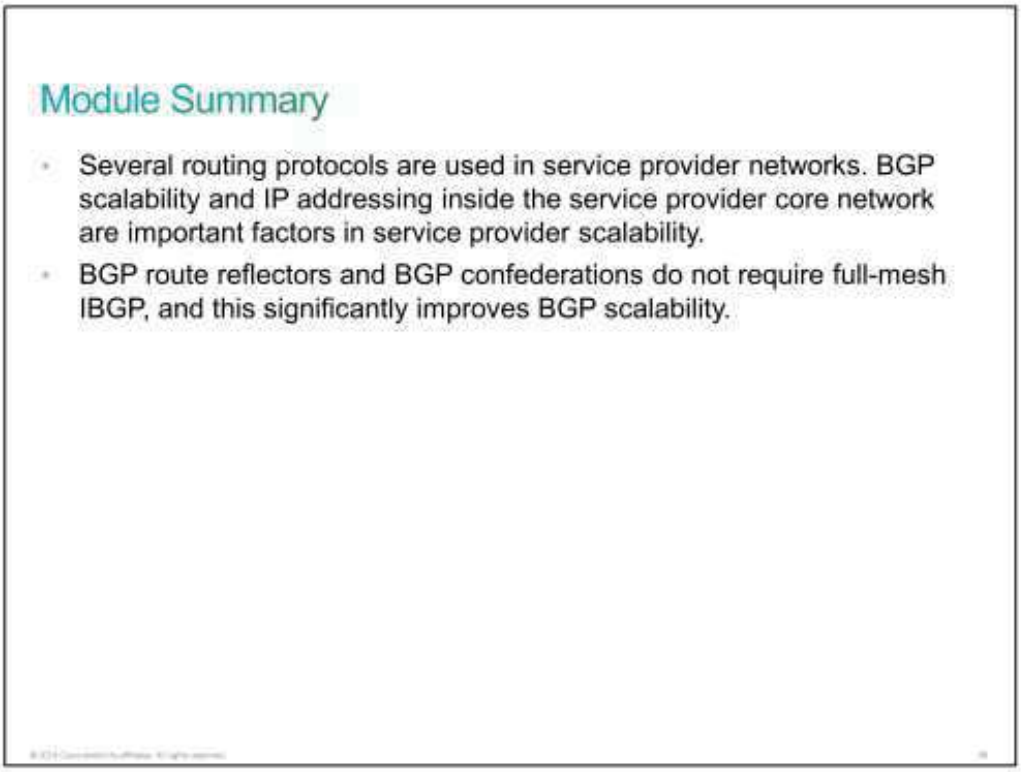
https://t.me/learningnets

https://t.me/learningnets

Lesson 3: Module Summary

Overview

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



This module first described how routes are propagated inside a typical service provider network. The module also discussed how to scale IP addressing and BGP routing. The module also described BGP route reflectors and BGP confederations as effective means to achieve BGP scalability. The module ended with a discussion of BGP policy accounting, which enables the service providers to measure traffic that is sent to, or received from, different BGP peers.

https://t.me/learningnets

https://t.me/learningnets

Lesson 4: 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 routing protocols might be used in a typical service provider network? (Choose two.) (Source: Scaling BGP in Service Provider Networks)

   A. EIGRP as IGP inside the service provider core network
   B. EBGP between BGP speakers inside the service provider network
   C. IS-IS as IGP inside the service provider core network
   D. IBGP between BGP speakers inside the service provider network

2. There is no need to include access links between customers and the service provider in IGP. (Source: Scaling BGP in Service Provider Networks)

   A. True
   B. False

3. Which three options can be used to scale IP addressing in service provider networks? (Choose three.) (Source: Scaling BGP in Service Provider Networks)

   A. Private IPv4 addresses on core links and loopback interfaces with MPLS switching
   B. Private IPv4 addresses on core links and loopback interfaces without MPLS switching
   C. Public IPv4 addresses on core links and loopback interfaces
   D. Link-local IPv6 addresses on core links and public IPv6 addresses on loopback interfaces

4. A total of 64 traffic indexes (buckets) are available for BGP policy accounting. (Source: Scaling BGP in Service Provider Networks)

   A. True
   B. False

https://t.me/learningnets

5. On the route reflector, routes received from a nonclient are sent to clients, EBGP peers and other nonclients. (Source: Introducing BGP Route Reflectors and Confederations)

   A. True
   B. False

6. Which two mechanisms are used to prevent routing loops in networks with redundant route reflectors? (Choose two.) (Source: Introducing BGP Route Reflectors and Confederations)

   A. Route-reflector-ID BGP attribute
   B. Cluster-ID list BGP attribute
   C. Originator-ID list BGP attribute
   D. Originator-ID BGP attribute

7. Place the steps that are needed to migrate a network to a route reflector–based design into the correct sequence. (Source: Introducing BGP Route Reflectors and Confederations)

Configure clients on route reflectors.	Second step
On clients, retain only IBGP sessions with route reflectors.	First step
Divide the AS into clusters and designate route reflectors.	Fourth step
Configure a cluster ID on route reflectors.	Third step

8. Which two Cisco IOS XR Software commands are needed to establish an intraconfederation EBGP session between loopback interfaces? (Choose two.) (Source: Introducing BGP Route Reflectors and Confederations)

   A. **update-source**
   B. **ttl-security**
   C. **ebgp-multihop**
   D. **send-community-ebgp**

 © 2014 Cisco Systems, Inc.

https://t.me/learningnets

Module Self-Check Answers

Answer Key

1. C, D
2. A
3. A, C, D
4. B
5. B
6. B, C
7. Configure a cluster ID on route reflectors. — Second step
   Divide the AS into clusters and designate route reflectors. — First step
   On clients, retain only IBGP sessions with route reflectors. — Fourth step
   Configure clients on route reflectors. — Third step
8. A, C

https://t.me/learningnets

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Module 3: Secure and Optimize BGP

Introduction

BGP is designed for reliability and scalability. As such, it has become the standard protocol that is used to carry many prefixes in the Internet today. BGP also has a great amount of flexibility regarding administrative policy controls, route selection, performance tuning, and scalability features. However, because BGP is used across the Internet, it is prone to various network attacks. Cisco IP Next-Generation Network administrators should take care to protect BGP and the infrastructure of the service providers from attacks. Administrators also need to ensure scalability, quick convergence, and low CPU utilization on devices across the network.

This module introduces advanced BGP configuration tools that are designed to secure BGP. Then the module discusses features that are available to improve BGP convergence and CPU utilization. Finally, the module introduces options that are available to increase BGP scalability.

Upon completing this module, you will be able to:

- Implement BGP security and optimization options
- Improve convergence in BGP networks using different features
- Configure BGP to limit the number of prefixes that are received from a neighbor and describe how to use BGP peer and configuration templates and how to use route dampening to minimize the impact of unstable routes

https://t.me/learningnets

https://t.me/learningnets

Lesson 1: Implementing Advanced BGP Operations

Overview

As organizations increase their web presence and reliance on the Internet for revenue, the need for reliable and geographically diverse Internet connectivity has become more common. These needs are often met through multihomed configurations that require BGP for connectivity to the BGP-speaking routers of a service provider. However, introducing BGP routing into organizations introduces additional risks that are present due to threats to BGP. Therefore, network administrators should have a thorough insight into BGP threats and countermeasures that are available to secure BGP. Furthermore, network administrators should know which BGP optimization options are available to optimize BGP operations.

This lesson first describes BGP security options. The lesson then discusses how to improve BGP operations using BGP optimization options.

Upon completing this lesson, you will be able to:

- Show the Cisco IP NGN infrastructure layer
- Describe threats against the service provider BGP process and DDoS attacks
- Describe countermeasures that are available to protect against BGP attacks
- Describe the BGP maximum prefixes limiting feature and configuration
- Describe BGP neighbor authentication
- Describe BGP TTL security check
- Describe control plane policing
- Explain BGP neighbor authentication, TTL security, and control plane policing configuration
- Describe remote-triggered black-hole filtering
- Describe destination-based RTBH and its implementation
- Describe source-based RTBH and its implementation

https://t.me/learningnets

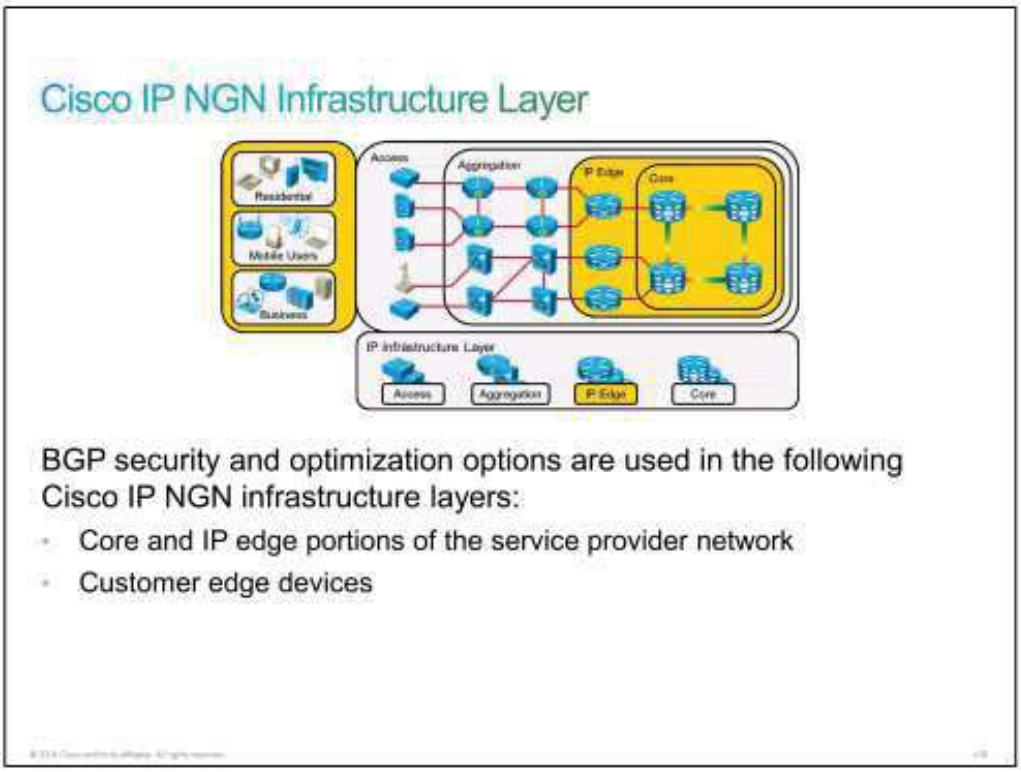
- Describe Cisco NSF
- Describe Cisco NSR
- Explain BGP NSF and NSR configuration
- Describe BGP process restart

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Cisco IP NGN Infrastructure Layer

BGP security and optimization options are used in the infrastructure layer of the Cisco IP Next-Generation Network (NGN).

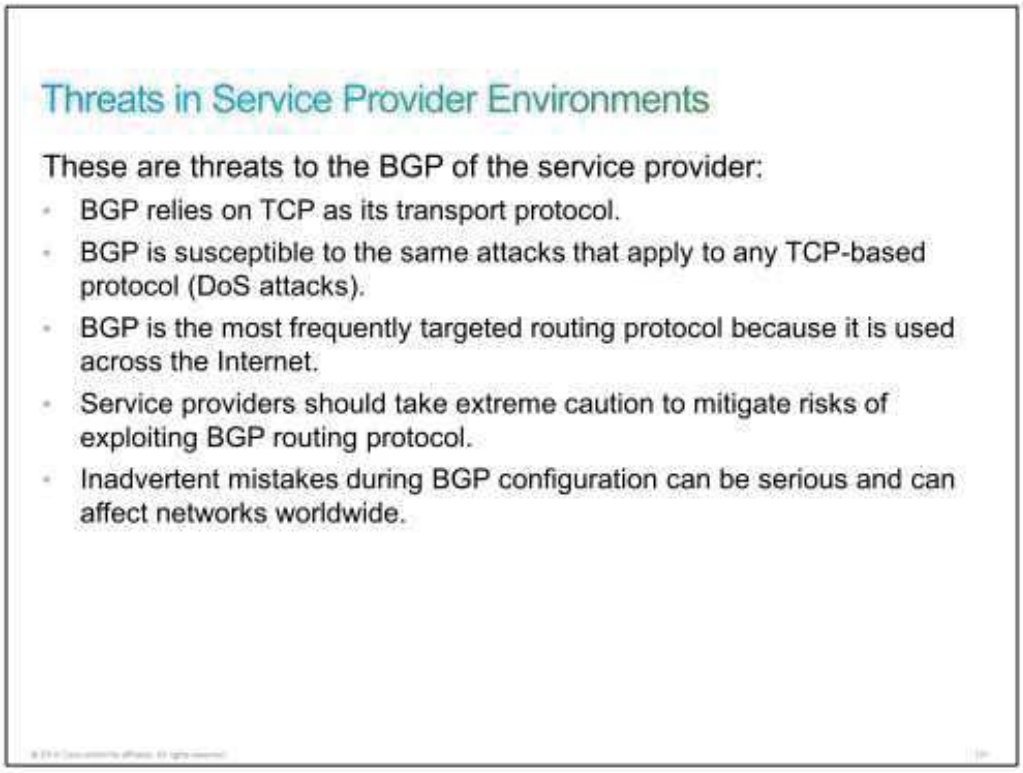


BGP security and optimization options mainly focus on the core and IP edge devices of the service provider. Some options also are used on the customer edge devices.

https://t.me/learningnets

Threats in Service Provider Environments

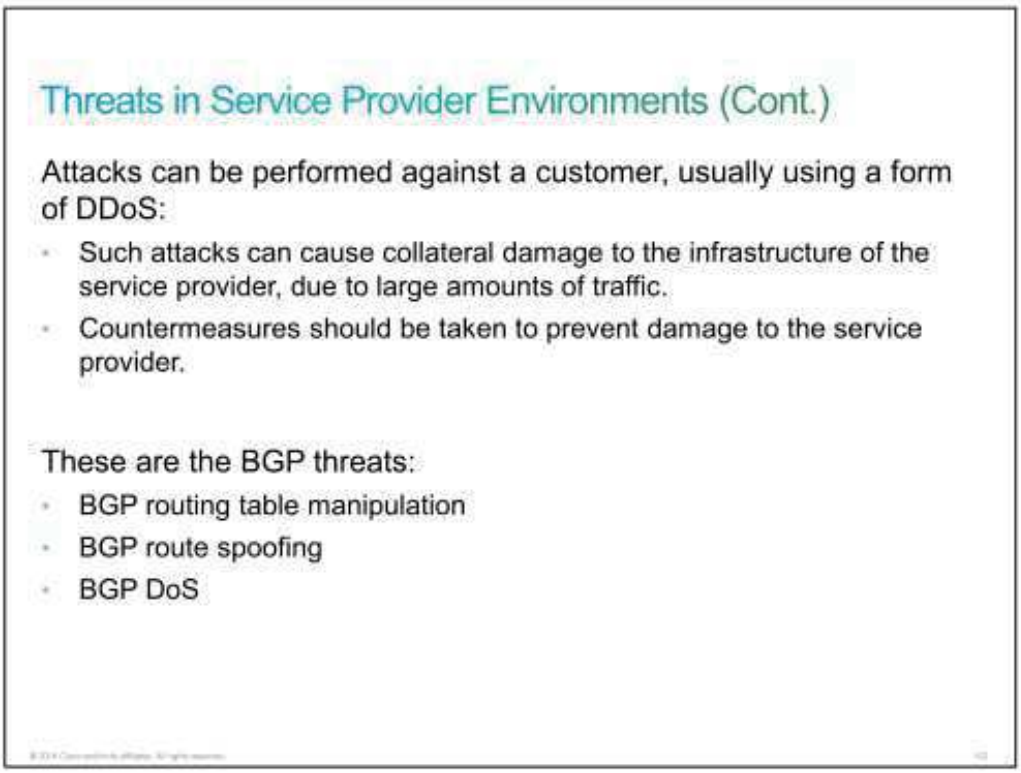
Service providers use BGP to exchange routing information with customers and upstream service providers. Because BGP relies on TCP, it is susceptible to the same attacks that apply to any TCP-based protocol. An example of this kind of attack is a DoS attack using SYN flooding. BGP is also used across the Internet and is therefore the most targeted routing protocol for network attacks. Therefore, various threats exist to service provider environments, and network administrators should take proper measures to counter these threats.



Network administrators in service provider networks must understand many important aspects of BGP as a protocol in order to assess where it may be susceptible to various forms of attack and where it must be protected using available countermeasures. Problems with BGP operations can also occur due to inadvertent mistakes during BGP configuration, and can affect networks worldwide because BGP routes are advertised across the world.

Not only attacks that are aimed directly at the service provider cause damage. When a customer is under a DDoS attack, the vast amount of traffic that is generated can have devastating consequences for the service provider infrastructure because all of the devices in the service provider must process and forward traffic. Appropriate measures should be taken to prevent collateral damage to the service provider in the case of a DDoS attack on a customer, even if the customer who is the intended DDoS victim must be "sacrificed."

https://t.me/learningnets



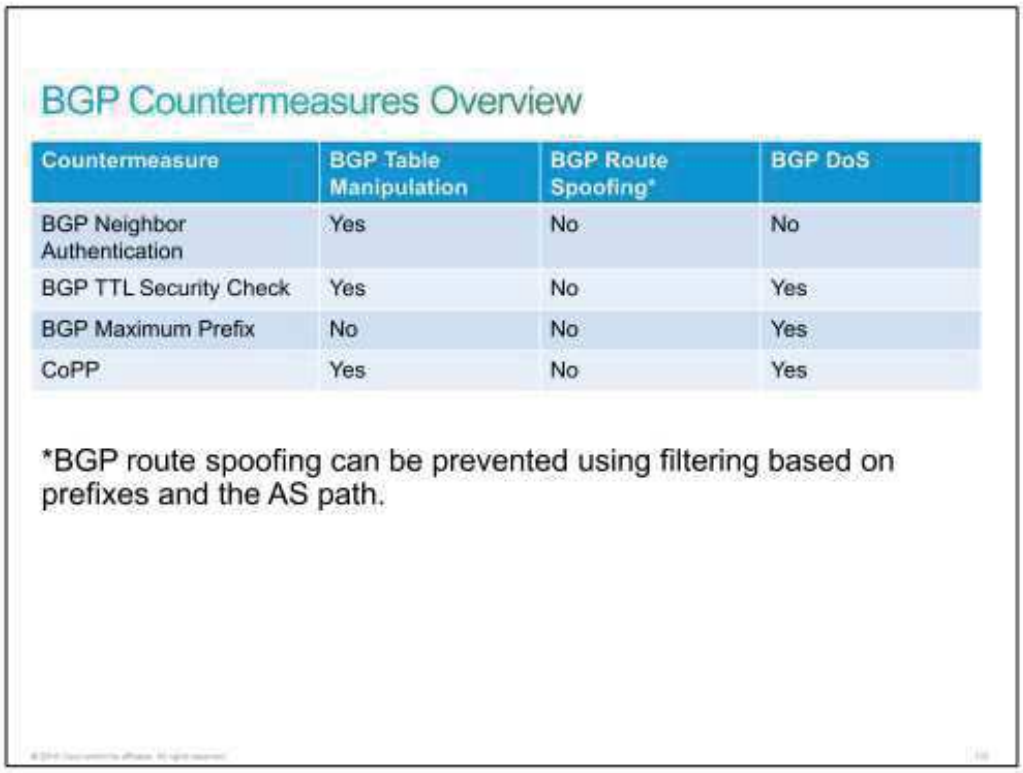
Threats that are specific to BGP include the following:

- **BGP routing table manipulation**: This scenario occurs when a malicious device alters the contents of the BGP routing table, which can, among other consequences, prevent traffic from reaching its intended destination without acknowledgment or notification.
- **BGP route spoofing**: This scenario occurs when a rogue BGP peer maliciously announces the prefixes of a victim in order to reroute some or all traffic to itself for untoward purposes (for example, to view contents of traffic that the router would otherwise not be able to read).
- **BGP DoS**: This scenario occurs when a malicious host sends unexpected or undesirable BGP traffic to a victim in an attempt to expend all available BGP or CPU resources, which results in a lack of resources for valid BGP traffic processing.

https://t.me/learningnets

BGP Countermeasures Overview

BGP has many countermeasures that are available to protect against attacks.



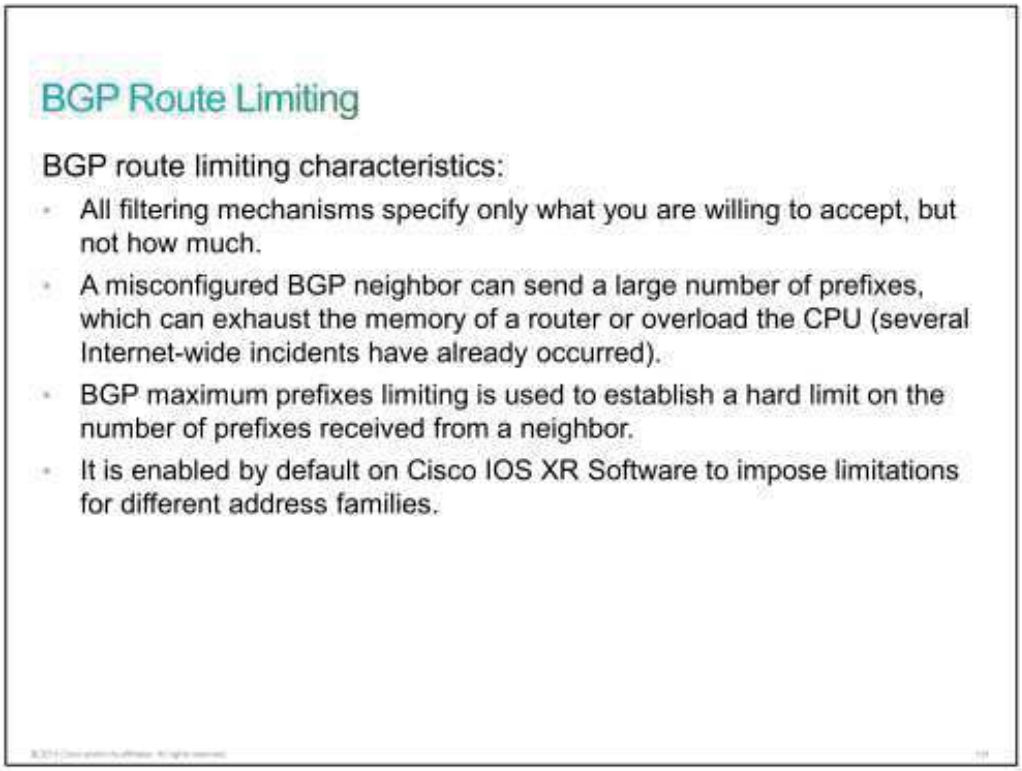
The table in the figure shows which countermeasures are available to protect against BGP threats.

Note	BGP route spoofing can be prevented using filtering based on prefixes and the AS path, and will not be discussed here. Filtering ensures that that network traffic is sent over the intended paths.

https://t.me/learningnets

BGP Route Limiting

Incoming route policies, which are applied on BGP routers, indicate the BGP attribute values that a route should have in order to be accepted. Route policies can be applied that match routes based on the network number or the BGP attributes that are attached to a route. The most commonly applied filtering used in route policies is one that matches the contents of the AS path attribute.



A service provider with a multihomed customer may use route policies to ensure that the routes that are received from the customer originate within the AS of the customer.

Improperly configured route policies in a customer router may accidentally cause a large number of Internet routes to be received by the customer. Even worse, a faulty configuration may cause prefixes to be advertised by the customer as though the routes originated inside the customer AS. This situation results in a BGP table in the service provider router that lists many of the possible destination networks on the Internet as reachable in the customer AS (with the AS path containing only a single entry, the customer AS). The BGP route selection in the service provider network would select those routes as the best (based on the AS path length) and direct much of the provider traffic that is intended for the Internet to the customer network.

A route policy that is based on an AS path in the service provider router will not prevent this accident. The routes that are sent by the customer have the anticipated AS path value. A route policy that is based on customer prefixes and that distinctly identifies and permits each of the network numbers that the customer may advertise will prevent the accident. However, such a routing policy is difficult to maintain.

A scalable solution to the need for limiting the number of routes (prefixes) that are received from a BGP neighbor is to use BGP maximum prefixes limiting. This feature enables you to specify how many routes a router can receive from a neighbor.

https://t.me/learningnets

BGP Route Limiting (Cont.)

BGP router terminates peering when a number of maximum prefixes is exceeded.

A router can be configured to:

- Generate a logging message when a specified percentage of the maximum prefixes is reached
- Re-establish BGP peering after a specified time (from 1 to 65535 minutes)
- Generate a logging message when the maximum prefix limit is exceeded, instead of terminating BGP peering

The BGP maximum prefix limiting feature is available on Cisco IOS, IOS XE, and IOS XR Software platforms. However, in Cisco IOS XR Software the maximum number of prefixes that are accepted from a peer for a given address family is enabled by default. This limitation safeguards the router from resource depletion that is caused by misconfiguration, either locally or on the remote neighbor. The default limits can be overridden through configuration of the **maximum-prefix** *limit* command for the peer for the appropriate address family. The following default limits are used if the user does not configure the maximum number of prefixes for the address family:

- 512K (524,288) prefixes for IPv4 unicast
- 128K (131,072) prefixes for IPv4 multicast
- 128K (131,072) prefixes for IPv6 unicast
- 512K (524,288) prefixes for VPNv4 unicast

Note	In Cisco IOS XR Software BGP also imposes maximum limits on the number of neighbors that can be configured on the router. The maximum limit is set to 4000 by default.

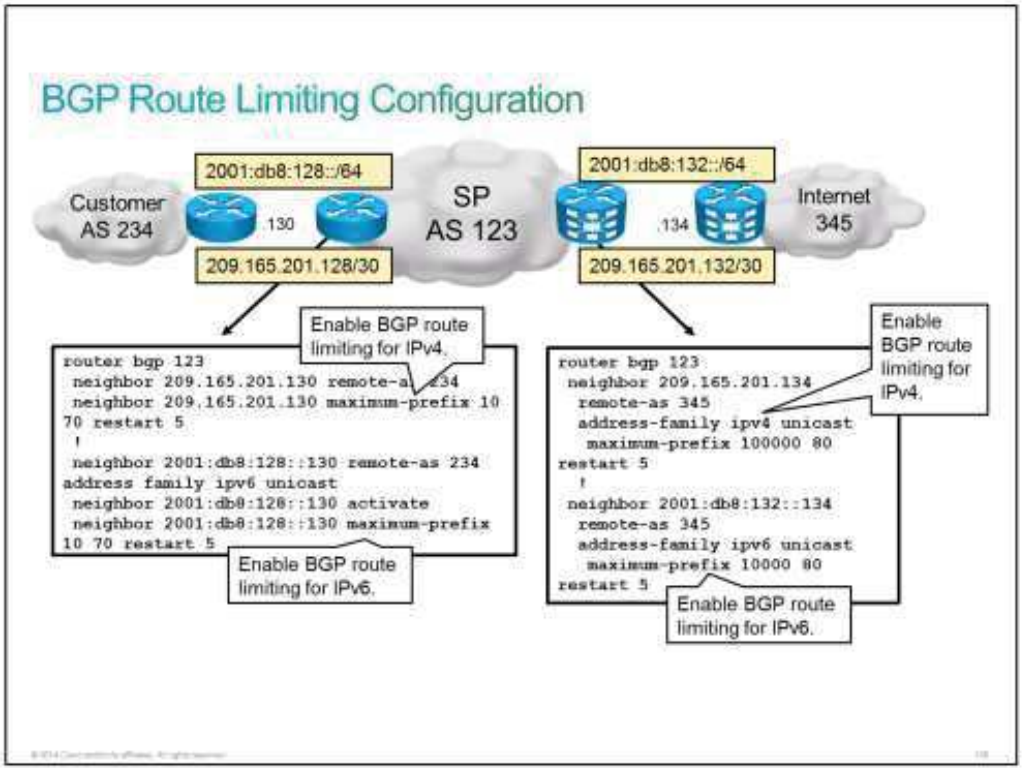
A cease notification message is sent to the neighbor, and peering with the neighbor is terminated, when the number of prefixes that is received from the peer for a given address family exceeds the maximum limit (either set by default or configured by the user) for that address family.

A router can be also configured to generate a logging message when a configured percentage of the maximum prefixes is reached. The router can also be configured to automatically re-establish a peering session that has been disabled because the maximum prefix limit has been exceeded. In addition, the router can be configured to generate and display a warning when a configured percentage of the maximum prefixes has been reached, instead of ceasing the peering with a neighbor.

 © 2014 Cisco Systems, Inc.

https://t.me/learningnets

BGP Route Limiting Configuration



The figure shows an example of how to configure a maximum number of prefixes that can be received by a neighbor.

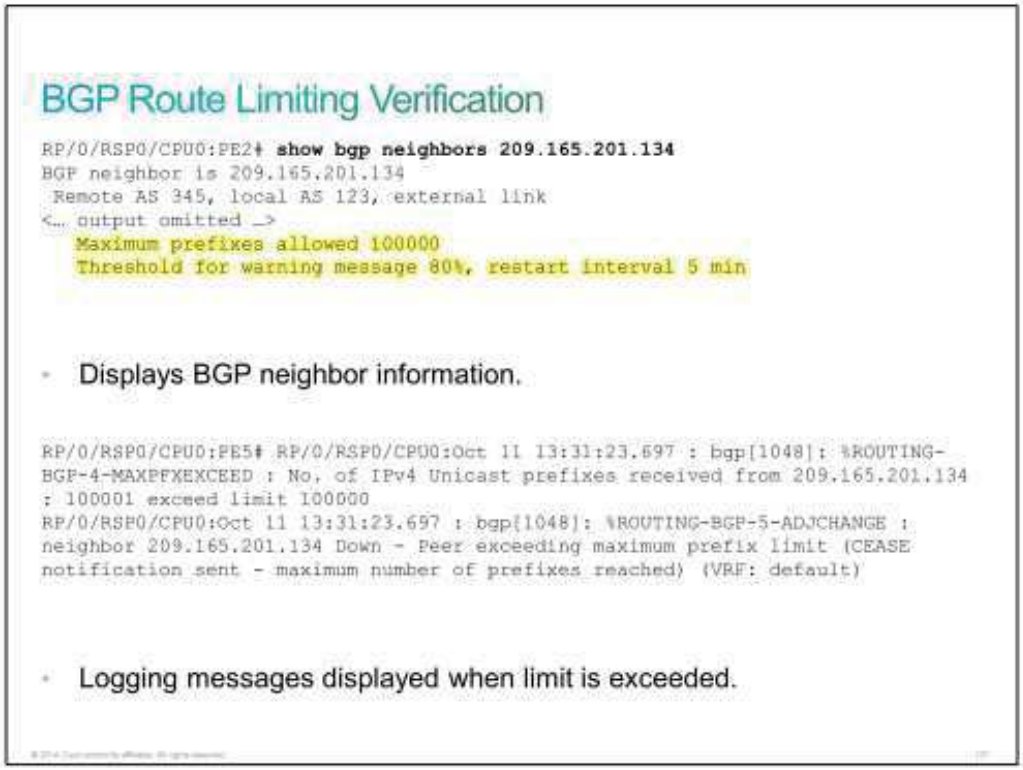
On the Cisco IOS XR Software platform, first enter BGP router mode. Then enter neighbor configuration submode and enter the proper neighbor address family configuration mode using the **address-family** command. Then enable BGP route limiting using the **maximum-prefix** command, followed by maximum number of routes, optional percentage when a logging message will be displayed, and optional restart interval when a peering will be enabled automatically. In the example, the maximum number of IPv4 prefixes is set to 100,000 and a logging message will be generated when 80 percent of the maximum limit is reached. A peering that is torn down will be automatically re-established after 5 minutes. In the example, the limit for IPv6 routes also is configured and is set to 10000 routes. To enable the router to generate a logging message instead of terminating the BGP session, use the **warning-only** keyword.

On the Cisco IOS and IOS XE Software platforms, configuration is similar. To enable route limiting for IPv4, use the **neighbor maximum-prefix** command under BGP router configuration mode. To configure route limiting for IPv6, first enter IPv6 address family, and then use the same command. Do not forget to activate a neighbor to transport IPv6 prefixes using the **neighbor activate** command under the proper address family configuration mode.

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

BGP Route Limiting Verification



To verify the configured maximum number of prefixes that can be received from a neighbor, use the **show bgp neighbor** command, followed by the IP address of the neighbor. The printout in the figure displays the configured parameters for the BGP maximum prefixes feature.

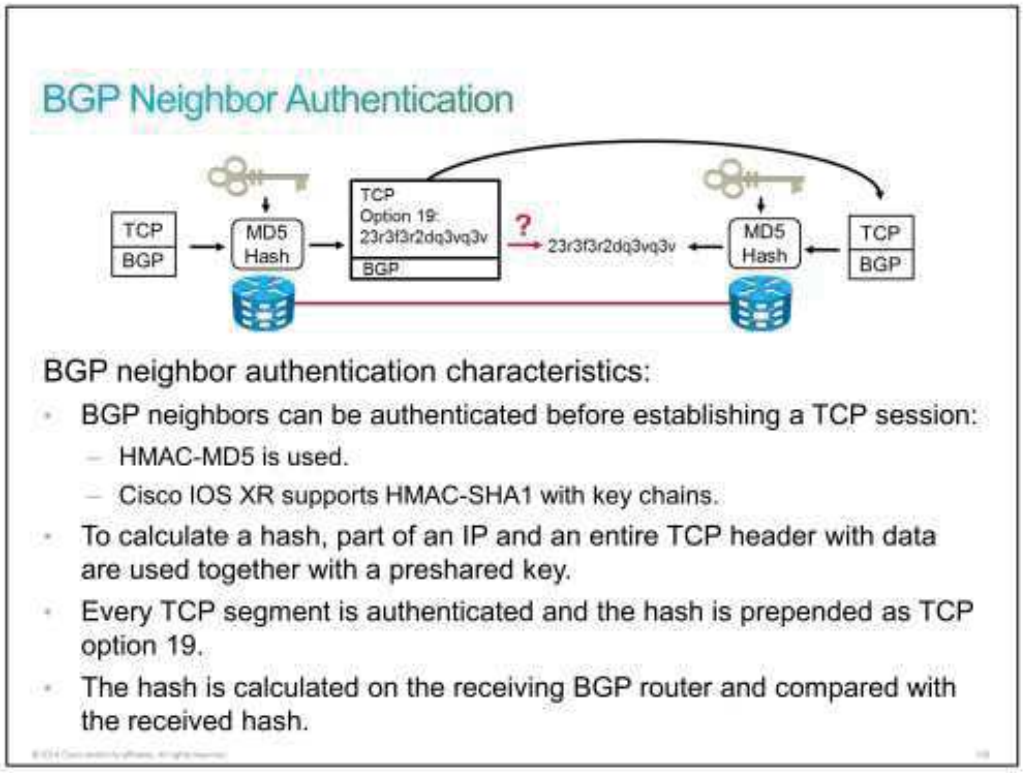
The figure also displays a logging message, which is displayed when a configured number of prefixes is reached and peering is stopped.

To force the neighbor to re-establish the BGP session, a network administrator must issue the **clear bgp** command for the neighbor. An exception to this practice is if the network administrator has specified the **restart** keyword in the configuration, in which case the router tries to re-establish the BGP session automatically after the expiration of the configured restart timeout interval.

https://t.me/learningnets

BGP Neighbor Authentication

One attack scenario is the route information manipulation attack. BGP neighbor sessions are established between two peers and then routes are exchanged between them. Administrators can enable the MD5-based neighbor authentication mechanism to ensure that only authorized peers can establish this BGP neighbor relationship, and that the routing information exchanged between these two devices has not been altered in transit.



The BGP neighbor authentication process is illustrated in the figure. Because BGP neighbor authentication is a symmetric technique, it must be enabled on both sides of the peering session at the same time. As shown in the figure, neighbor authentication using MD5 creates an MD5 hash for each packet that is sent as part of a BGP session. Specifically, the sending router uses portions of the IP and TCP headers, TCP payload (including the BGP route advertisements), and a shared secret to generate the MD5 hash. The MD5 hash is then stored in TCP option 19, which is created specifically for this purpose by RFC 2385. The receiving BGP neighbor uses the same algorithm and shared secret to compute its own version of the MD5 hash. It then compares its own version with the one it received. If the MD5 hash values are not identical, the receiving BGP neighbor discards the packet. In any other circumstance, the packet is accepted and processed by BGP.

Note	Because the MD5 hash also is calculated over IP and TCP headers, BGP authentication will not work over devices performing NAT or PAT. BGP authentication will also not work with firewalls that randomize TCP sequence numbers. Some firewalls may also remove the TCP option 19 from the TCP header, thus failing the authentication check on the receiving router.

https://t.me/learningnets

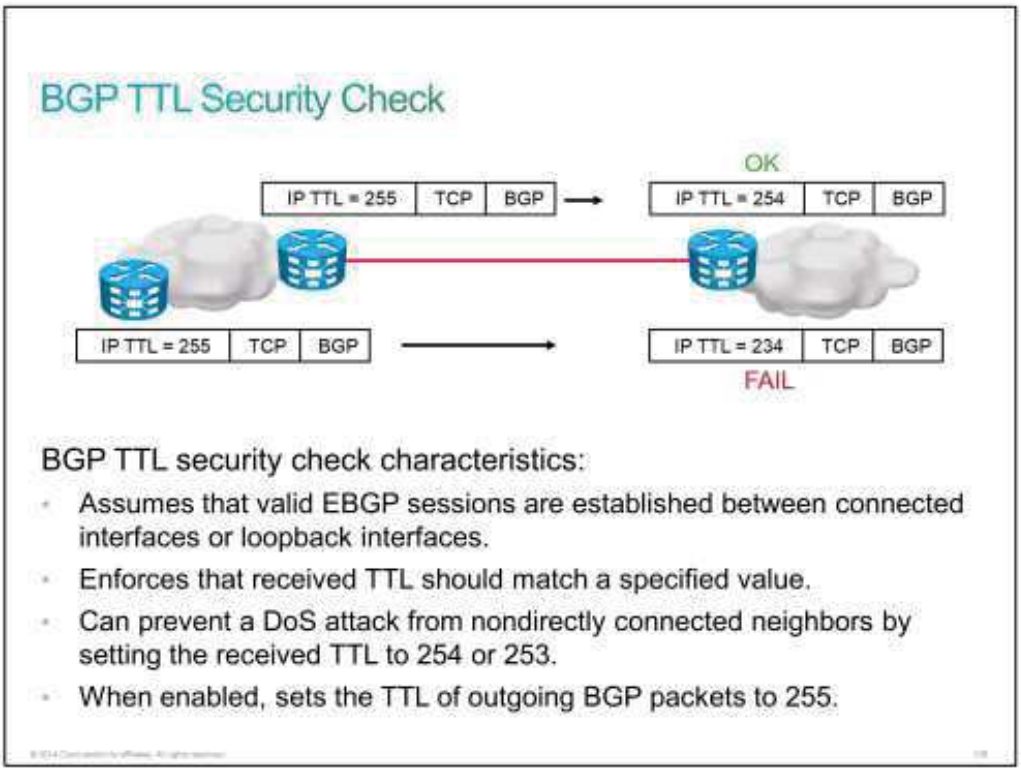
BGP neighbor authentication can be added to existing BGP sessions. The shared secret may also be changed on existing sessions without terminating and re-establishing these sessions, as long as both sides of the BGP session are modified within the BGP session timeout window. The timeout window has a default of 180 seconds.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

BGP TTL Security Check

Another BGP attack scenario is a DoS attack against the BGP process. The BGP TTL security check is designed to protect the BGP process from these kinds of CPU-utilization-based attacks and route manipulation attempts.



By default when EBGP is configured, the IP header TTL for all neighbor session packets is set to 1. This setting was originally assumed to be useful because it prevents the establishment of an EBGP session beyond a single hop. However, an attacker could be located up to 255 hops away and still send spoofed packets to the BGP-speaking router successfully by setting the TTL to the maximum value of 255. For example, an attacker could send many TCP SYN packets to the BGP peer in hopes of overwhelming the BGP process. The BGP MD5 neighbor authentication technique described earlier does not protect against this kind of attack, and can actually exacerbate its effects by causing the router CPU to expend resources while it attempts to compute MD5 hashes on large numbers of attack packets.

The BGP TTL security check leverages the fact that the vast majority of EBGP peering sessions are established between routers that are next to each other (for example, either between directly connected interfaces or possibly between loopbacks). Because successful TTL spoofing is considered nearly impossible, a mechanism was developed that is based on an expected TTL value, in order to provide a simple and robust defense from infrastructure attacks that are based on forged BGP packets.

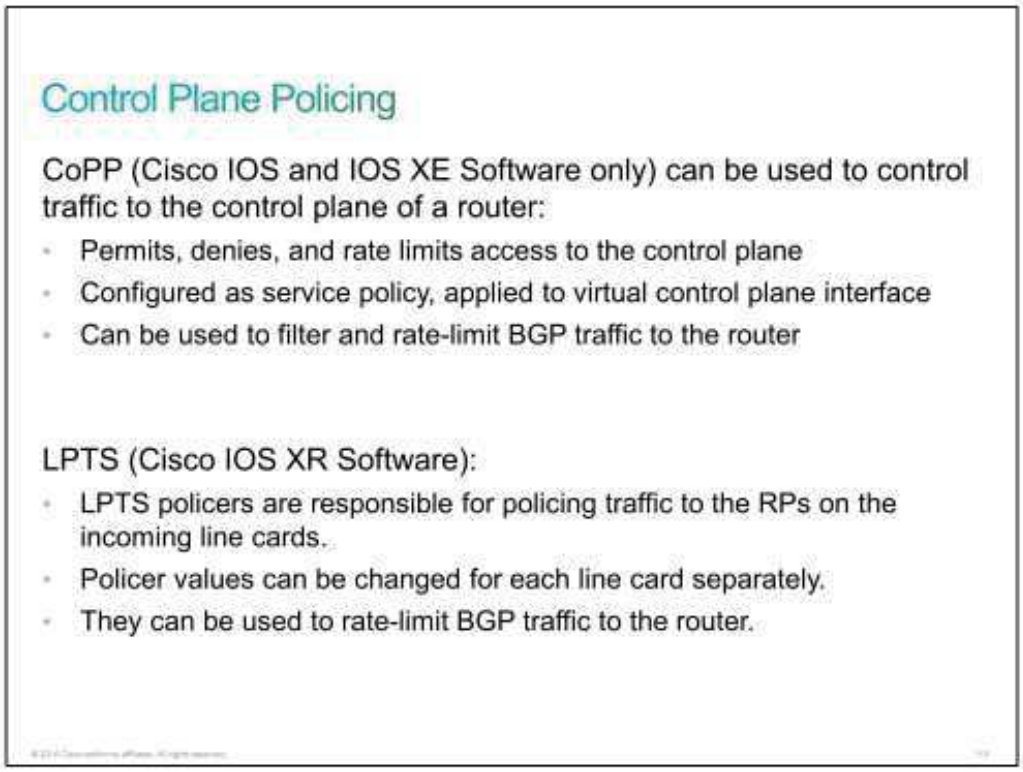
When the BGP TTL security check is enabled, the initial TTL value for an EBGP packet is set to 255 rather than 1, and a minimum TTL value is enforced on all BGP packets that are associated with that EBGP session. For example, if the TTL security check is configured to the value of 1, the receiving BGP router expects all BGP packets to have a TTL of 254 (the initial value of 255 minus one hop). Because the IP header TTL value is decremented by each router hop along its path to its final destination, the diameter from which an attacker could possibly source packets is restricted to those routers that are directly connected.

This feature needs to be configured on each participating router. It protects the EBGP peering session in the incoming direction only, and has no effect on outgoing IP packets on the remote router.

https://t.me/learningnets

Control Plane Policing

CoPP is available on the Cisco IOS and IOS XE Software platforms. LPTS is available on the Cisco IOS XR Software platforms.



CoPP is available on the Cisco IOS and IOS XE Software platforms. CoPP uses the concept of early rate-limiting and dropping of traffic that is sent to the central processor of the network device by applying QoS policies to a virtual aggregate CPU-bound queue called the control plane interface. This queue receives all aggregated traffic that is bound for the control plane (which includes the routing protocols), the management plane (management processes), and the slow data plane path traffic of the network device.

CoPP can granularly permit, drop, or rate-limit traffic to the CPU using an MQC interface. Because CoPP aggregates all traffic that is forwarded to the CPU of the network device, it is independent of interfaces, allowing a central configuration mechanism to protect the CPU resources of the network device.

A network administrator can use CoPP to specify which hosts can establish BGP peering with BGP routers in the service provider network. The network administrator also can rate-limit traffic from allowed BGP peers.

With LPTS, available on the Cisco IOS XR Software platforms, the control packets—which are bound for the RP—are policed using a set of ingress policers in the incoming line cards. These policers are programmed statically during bootup by LPTS components. The policers are applied based on the flow type of the incoming control traffic. The flow type is determined from the packet headers. The policer rates for these static ingress policers are defined in a configuration file, which is programmed on the line card during bootup.

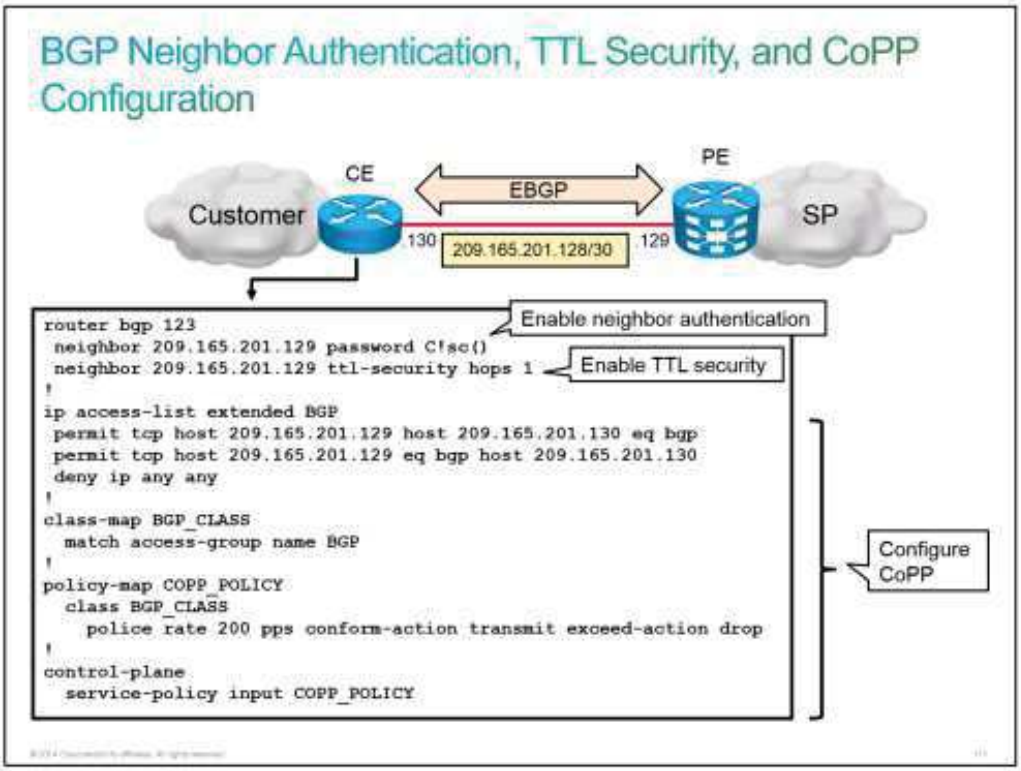
You can change the policer values based on the flow types of traffic. You can configure the rate per policer per node (locally) and globally using the CLI, thereby overwriting the static policer values.

The network administrator can use LPTS to rate-limit BGP traffic to the service provider routers.

https://t.me/learningnets

BGP Neighbor Authentication, TTL Security, and CoPP Configuration

The BGP neighbor authentication, TTL security, and control plane policing configuration is shown below:



The figure shows a configuration of the BGP security options on the Cisco IOS and IOS XE Software platforms. To enable BGP MD5 authentication, first enter BGP router configuration mode and use the **neighbor password** command. To enable TTL security, use the **neighbor ttl-security hops** command that is followed by the maximum distance away (number of hops) that a BGP neighbor can be.

To configure CoPP to rate-limit BGP traffic from allowed BGP hosts, first define an ACL of trusted hosts, using specific protocols to access the router using extended IP access lists. In the example, the ACL specifies TCP traffic from the PE router to the CE router by using the BGP port number. Two entries exist because which router initiates the BGP session is not known.

Next, enter class map global configuration command mode by creating an MQC class map using the **class-map** command. Inside the class map, specify criteria to match using the **match** command. In this example, the **match access-group name** command matches the previously configured ACL.

Next, enter policy map configuration mode to define a named policy using the **policy-map** command. Enter the class map for which you want to create or change policy, using the **class** command. This command associates a service policy with a class. Then configure traffic policing or, alternatively, drop this class using the **drop** command. There are three distinct ways to configure policing:

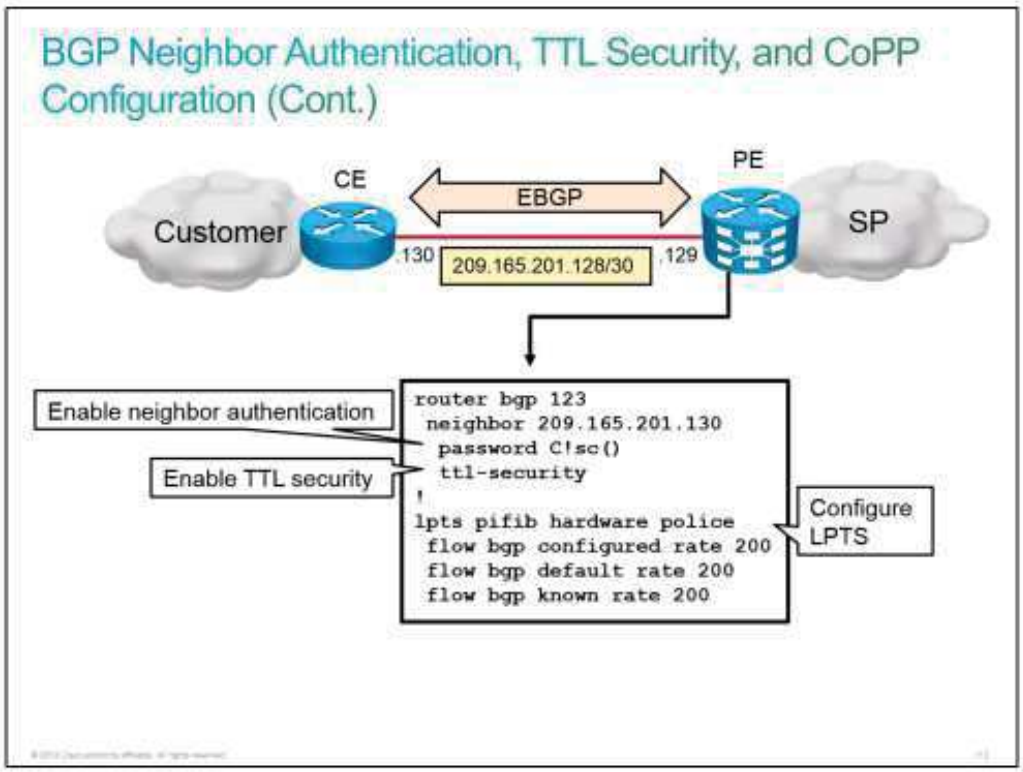
- Packets per second (p/s) using the **police rate** *units* **pps** command
- Bits per second (b/s) using the **police rate** *units* **bps** command
- Percentage (percentage of bandwidth) using the **police rate** *percent* **percentage** command

https://t.me/learningnets

Finally, enter control plane configuration mode using the **control-plane** command and attach the policy map to a control plane interface using the **service-policy input** command.

The example in the figure will rate-limit BGP traffic to 200 packets per second. Excess packets will be discarded. All other traffic to the control plane will be processed without any restrictions. However, you should also rate-limit other traffic to the control plane to a reasonable value.

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



The figure shows a configuration of the BGP security options on the Cisco IOS XR Software platform. To enable BGP MD5 authentication, first enter BGP router configuration mode and then enter neighbor configuration mode. Then, use the **password** command. To enable TTL security, use the **ttl-security** command. The software will automatically determine the number of hops.

On the Cisco IOS XR Software platform, BGP authentication also can be configured using key chains. Key chains are useful when smooth transition from one key to another is desired, because each key in the key chain can have a specified send and accept validity. The following is an example of BGP authentication using key chains on a Cisco IOS XR Software router:

```
key chain BGP_KEY_CHAIN
 key 1
  cryptographic-algorithm HMAC-SHA1-12
  key-string C!sc()
!
router bgp 123
 neighbor 209.156.201.130
  keychain BGP_KEY_CHAIN
```

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

To use LPTS to change the allowed rate of incoming BGP packets, first configure the ingress policers and enter pifib policer global configuration mode using the **lpts pifib hardware police** command. You also can enable policers on a line card separately by providing a **location** keyword, followed by the node ID. Then specify the allowed rate of packets using the **flow** command, followed by the flow type, the **rate** keyword, and the number of packets per second. In the example, incoming BGP packets are rate-limited to 200 packets per second.

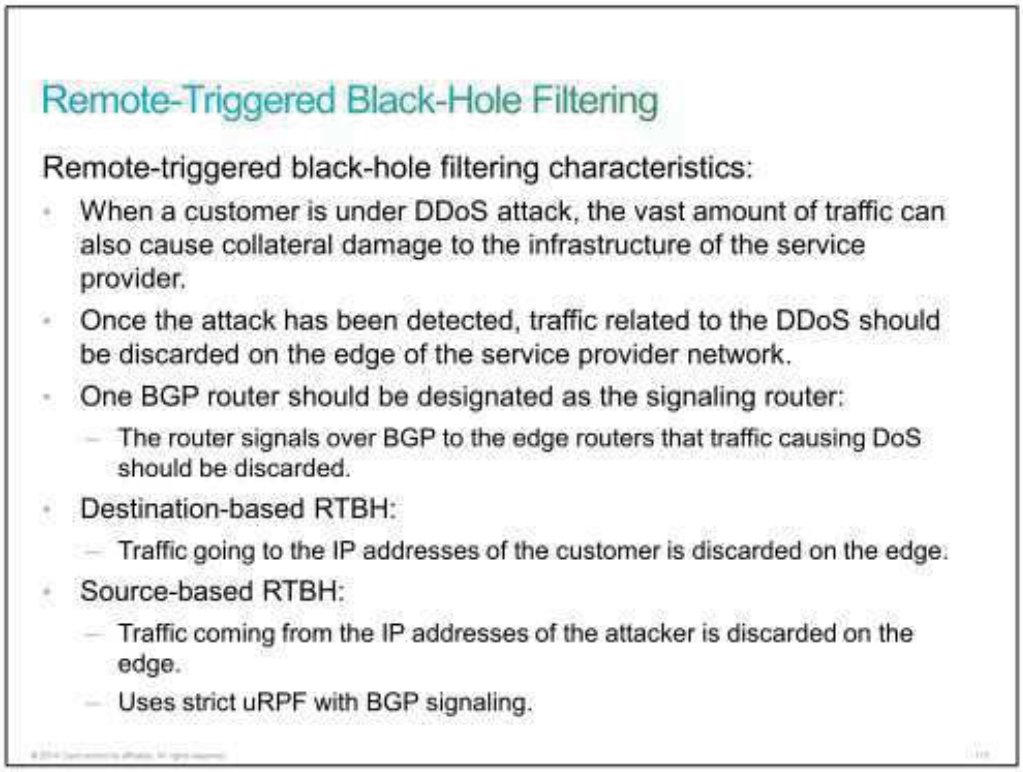
The following table describes flows that are associated with BGP.

Flow Type	Description	Default Packet Rate
bgp configured	Packets from a configured BGP peer	2000
bgp default	Packets from unconfigured, newly configured, or wildcard BGP peers	2500
bgp known	Packets from established BGP peering sessions	1500

https://t.me/learningnets

Remote-Triggered Black-Hole Filtering

RTBH filtering is another security option that can be implemented in service provider networks. RTBH is not a countermeasure to protect BGP, but it can be used when a customer is under DDoS attack and the vast amount of traffic could cause collateral damage to the infrastructure of the service provider. The first countermeasure against such attacks is to identify the source or destination of the attack. Then one router, called the signaling router, notifies the edge routers that traffic coming from the source (attacker) of the attack or going to the destination (victim) must be dropped. The BGP routing protocol is used for signalization between the signaling router and other edge routers. This way the harmful traffic is dropped on the edge of the service provider network and does not cause damage to the service provider devices.



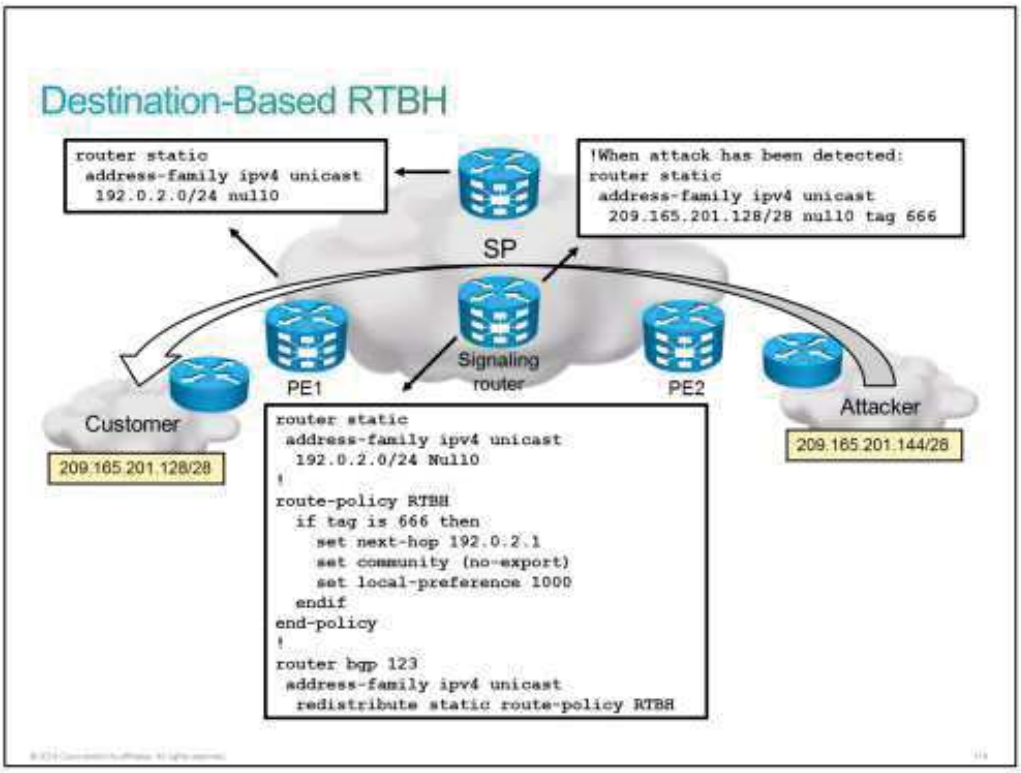
There are two options to implement RTBH:

- **Destination-based RTBH**: In this implementation, traffic going to the IP address of the victim is discarded on the edge of the service provider. It requires knowledge of the victim IP addresses, and willingly sacrifices communication between the victim and the rest of the Internet. This practice helps the attacker because the point of a DDoS attack is to deny all service to the victim. However, it conserves the service provider resources. Once the DDoS traffic is contained, the service provider can further investigate the attack and prevent the victim from further DoS attacks. Once the service provider activity is done, the service provider can stop discarding traffic to the customer. Black-hole filtering is implemented by using an unreachable (null interface) next hop for packets going to the IP addresses of the victim.
- **Source-based RTBH**: In this implementation, traffic coming from the IP addresses of the attacker is discarded on the service provider edge. It requires knowledge of the attacker IP addresses. However, in many DDoS attacks, the attacker IP addresses are difficult to determine because the number and variety of devices performing the attack can be very large due to worm infections. Source-based RTBH uses BGP signaling together with uRPF to drop packets coming from the IP addresses of the attacker.

https://t.me/learningnets

Destination-Based RTBH

The destination-based RTBH implementation is shown below:



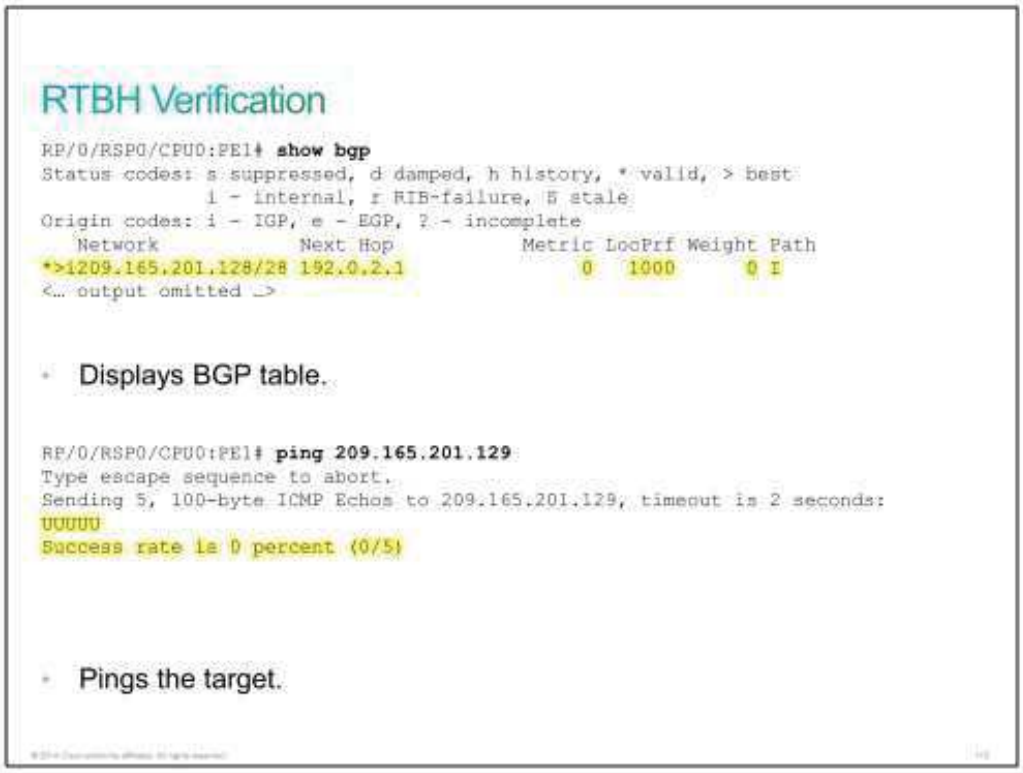
The figure shows the implementation of destination-based RTBH on Cisco IOS XR Software platforms. RTBH is configured in three steps:

1. Before the attack, all routers should be configured with a static route for a prefix that is not used in the network. In the example, prefix 192.0.2.0/24 is used, which is assigned as TEST-NET for use in documentation and example code. This static route should point to the null interface on all the routers. The IP address from this network will be used as the next hop for traffic that needs to be black-holed.
2. One router should be configured as the signaling router. This router will be used to trigger a traffic black hole once the attack has been detected and the victim IP addresses have been determined. Signaling is implemented as redistribution of static routes into BGP and the setting the next hop for the redistributed static route to a black-holed next hop (192.0.2.1 in the example). Redistribution is implemented with a route policy, which matches static routes that are tagged with 666 and sets local preference to a value that is high enough to prefer the black-holed path over the valid one. In the example, local preference is set to 1000. The route policy also adds a no-export community to the redistributed routes to prevent advertisement of the routes outside the autonomous system of the service provider.
3. Once the attack has been detected and the IP addresses of the victim have been determined, you should configure a static route for those IP addresses on the signaling router. This static route should be tagged with a tag that is used to match static routes for redistribution (666 in the example), and should point to the null interface.

https://t.me/learningnets

When the static route for the victim IP addresses is redistributed into BGP, it will be advertised to all BGP peers with a next hop set to 192.0.2.1. When an edge router receives traffic that is bound for the IP addresses of the victim, the router will perform a routing table lookup. The table entry will point to 192.0.2.1. The router will perform a recursive lookup for the 192.0.2.1 next hop, which will point to the null interface and cause traffic to be discarded.

RTBH Verification



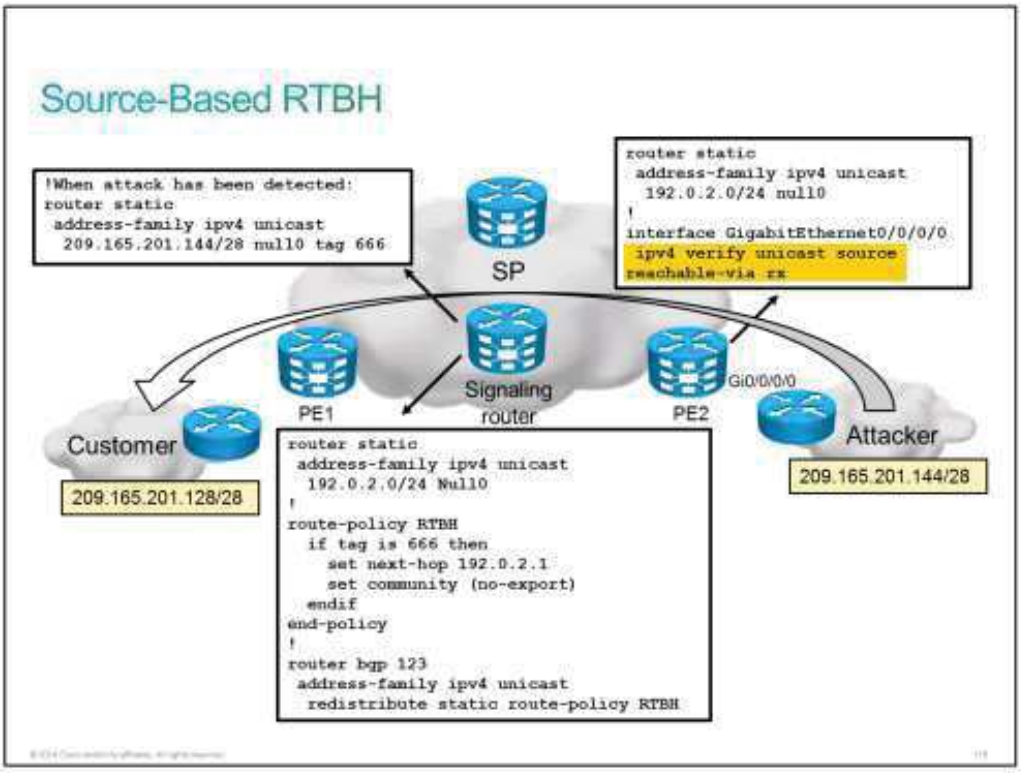
To verify RTBH, examine the BGP table. The next hop for the victim IP addresses (209.165.201.128/28) is set to 192.0.2.1, which is locally configured to point to the null interface.

If you ping the IP address of the victim, you should see that the destination is unreachable.

https://t.me/learningnets

Source-Based RTBH

Source-based RTBH is implemented similarly to destination-based RTBH. The difference is that when the attack has been detected, the attacker IP address must be determined. On the signaling router, the static route for the attacker IP addresses must be configured. This route will then be redistributed over BGP to all BGP peers. uRPF must be configured on all the edge routers on the interfaces where the attack is likely to be seen. These interfaces are customer-facing or upstream service provider-facing interfaces.



uRPF has two modes of operation:

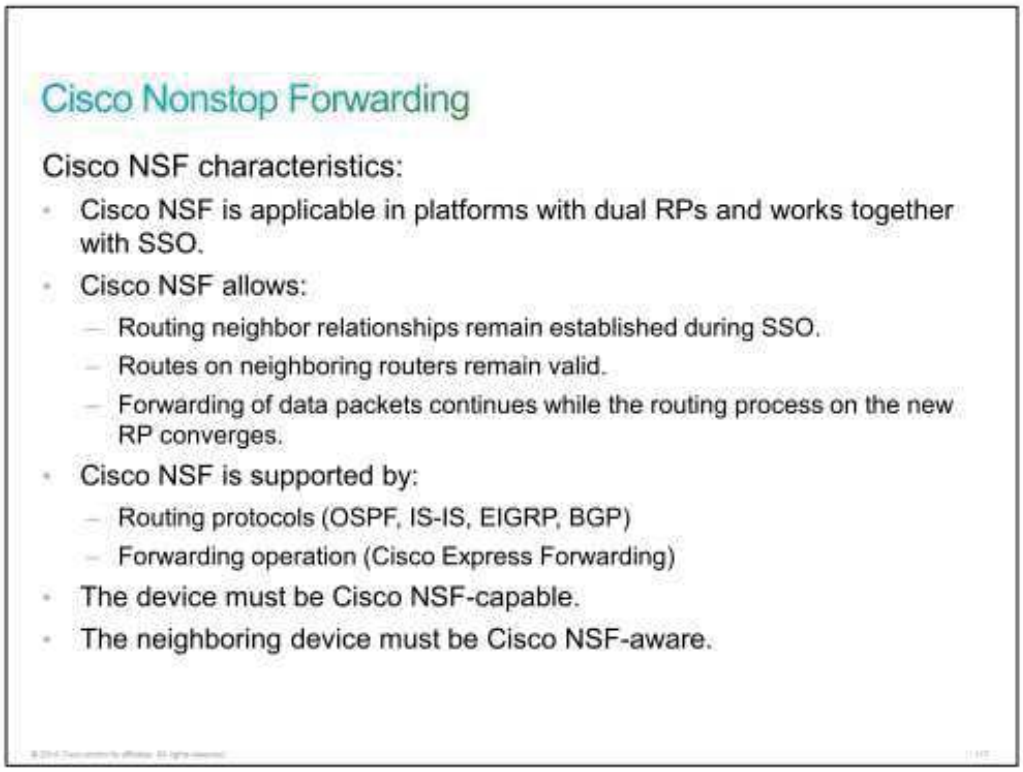
- **uRPF strict mode**: A router that receives a packet also performs the routing table lookup for the source IP address. The packet must be received on the interface that the router will use to forward the return packet, otherwise the packet is silently dropped.
- **uRPF loose mode**: The source address of the received packet must appear in the routing table. Additionally, a packet that contains a source address for which the return route points to the Null 0 interface will be silently dropped.

In the example, strict uRPF is enabled on the PE2 router on the GigabitEthernet0/0/0/0 interface using the **ipv4 verify unicast source reachable-via rx** command. A router that receives a packet will perform the routing table lookup for the source IP address. If the packet came from the attacker, the router will see that the routing table for the attacker IP addresses points to the 192.0.2.1 next-hop IP address, which in turn points to the null interface. The packet will be dropped because the packet was received through the other interface that would be used to forward the return traffic.

https://t.me/learningnets

Cisco Nonstop Forwarding

Cisco Nonstop Forwarding, also known as graceful restart, works with the SSO to minimize the amount of time a network is unavailable to its users following a switchover. The main objective of Cisco NSF is to continue forwarding IP packets following a route processor switchover in platforms with dual RPs.



When a networking device restarts, usually all routing peers of that device detect that the device went down and then came back up. This transition results in what is called a routing flap, which could spread across multiple routing domains. Routing flaps that are caused by routing restarts create routing instabilities, which are detrimental to the overall network performance. Cisco NSF helps to suppress routing flaps in SSO-enabled devices, thus reducing network instability.

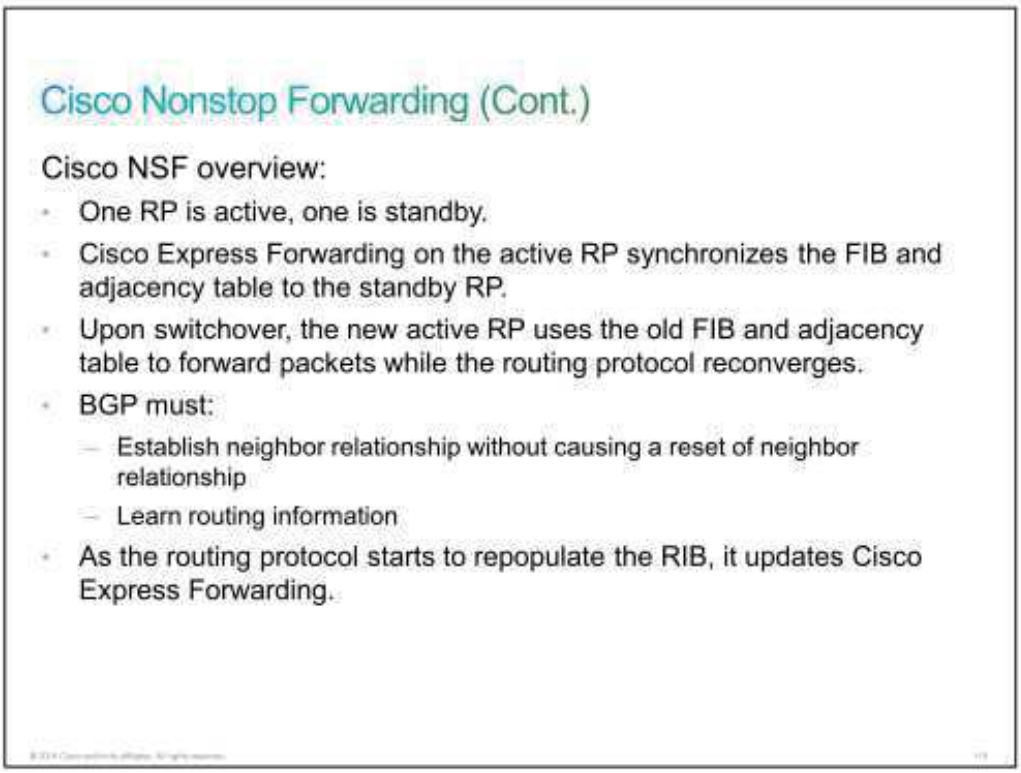
Cisco NSF allows for the forwarding of data packets to continue along known routes while the routing protocol information is being restored following a switchover. With NSF, peer networking devices do not experience routing flaps. Data traffic is forwarded through intelligent line cards or dual forwarding processors, while the standby RP assumes control from the failed active RP during a switchover.

Cisco NSF is supported by four protocols for routing (BGP, EIGRP, OSPF, and IS-IS), and by Cisco Express Forwarding for forwarding. The routing protocols have been enhanced with Cisco NSF capability and awareness, meaning that routers running these protocols can detect a switchover and take the necessary actions to continue forwarding network traffic and to recover route information from the peer devices.

A networking device is said to be Cisco NSF-aware if it is running NSF-compatible software. A device is said to be Cisco NSF-capable if it has been configured to support NSF and can rebuild routing information from NSF-aware or NSF-capable neighbors.

Cisco NSF depends on Cisco Express Forwarding to continue forwarding packets during switchover while the routing protocols rebuild the RIB tables. Once the routing protocols have converged, Cisco Express Forwarding updates the FIB table and removes old route entries. Cisco Express Forwarding, in turn, updates the line cards with the new FIB information.

https://t.me/learningnets



Cisco NSF always operates together with SSO. In specific Cisco networking devices that support dual RPs, SSO establishes one of the RPs as the active processor while the other RP is designated as the standby processor, and then synchronizes the FIB and adjacency table between them. A switchover from the active to the standby processor occurs when the active RP fails, is removed from the networking device, or is manually taken down for maintenance. During the switchover, the new active RP uses the old FIB and adjacency table to forward packets while the routing protocol reconverges. Once the routing protocols have updated the RIB, Cisco Express Forwarding updates the FIB table and removes old route entries. Cisco Express Forwarding, in turn, updates the line cards with the new FIB information.

The routing protocols (for example, BGP) run only on the active RP, and they receive routing updates from their neighbor routers. Routing protocols do not run on the standby RP. If Cisco NSF is not configured, a neighboring device will tear down the adjacency during a switchover. The result would be a routing flap that would spread across the network.

Using Cisco NSF, the routing protocols on an NSF-capable device requests that the NSF-aware neighbor devices send routing information to help rebuild the routing tables. The Cisco NSF-aware neighbor will not tear down the neighbor relationship. It will re-establish the neighbor relationship with the Cisco NSF-capable device, and will send routing information to synchronize the routing table on an NSF-capable device.

A Cisco NSF-aware router can peer with Cisco NSF-capable routers and facilitate the resynchronization of routing information with such routers.

https://t.me/learningnets

Cisco Nonstop Forwarding (Cont.)



Cisco NSF-aware neighbors:

- Do not reconverge
- Help the NSF-capable router restart
- Continue forwarding traffic to the restarting router

Cisco NSF-capable router:

- Data is forwarded in the hardware base.
- Rebuilds the Layer 3 routing protocol database from the neighbor.

When a Cisco NSF-capable router begins a BGP session with a BGP peer, it sends an OPEN message to the peer. Included in the message is a declaration that the Cisco NSF-capable or NSF-aware router has "graceful restart capability." If the BGP peer has received this capability, it is aware that the device sending the message is Cisco NSF-capable. Both the Cisco NSF-capable router and its BGP peers (NSF-aware peers) need to exchange the graceful restart capability in their OPEN messages at the time of session establishment. If both peers do not exchange the graceful restart capability, the session will not be capable of graceful restart.

If the BGP session is lost during the RP switchover, the Cisco NSF-aware BGP peer marks all the routes that are associated with the NSF-capable router as stale. However, it continues to use these routes to make forwarding decisions for a time. This functionality means that no packets are lost while the newly active RP is waiting for convergence of the routing information with the BGP peers.

After an RP switchover occurs, the Cisco NSF-capable router re-establishes the session with the BGP peer. In establishing the new session, it sends a new graceful restart message that identifies the Cisco NSF-capable router as having restarted.

At this point, the routing information is exchanged between the two BGP peers. Once this exchange is complete, the Cisco NSF-capable device uses the routing information to update the RIB and the FIB with the new forwarding information. The Cisco NSF-aware device uses the network information to remove stale routes from its BGP table. Once the stale routes are removed, the BGP protocol is fully converged.

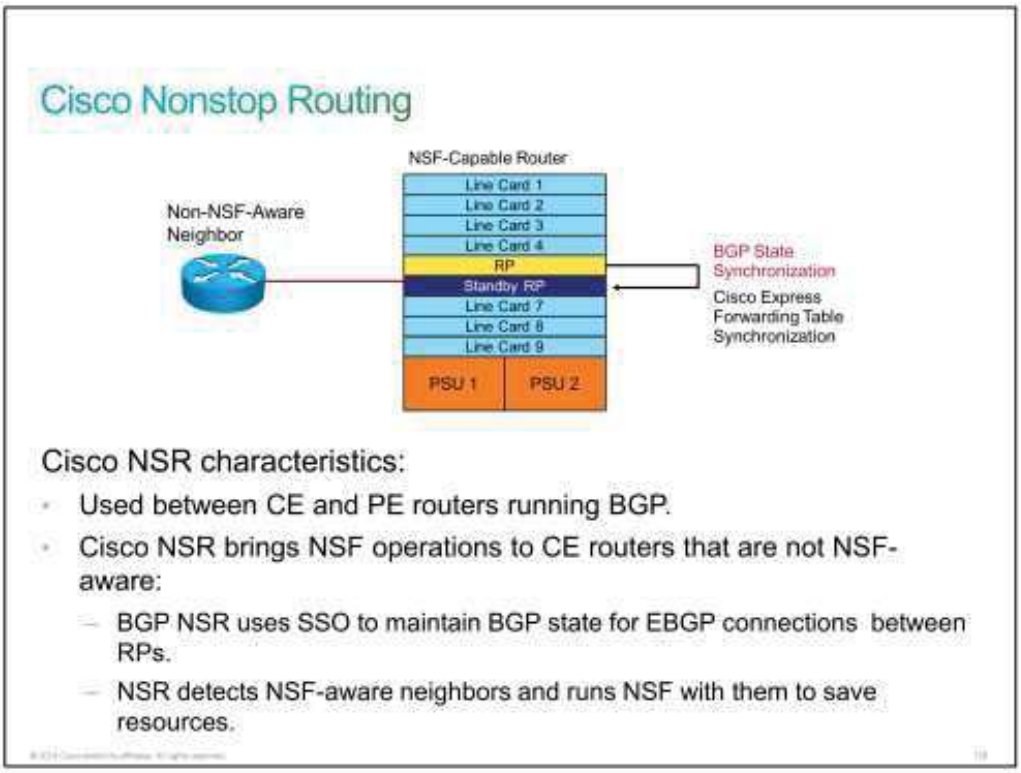
If a BGP peer does not support the graceful restart capability, it will ignore the graceful restart capability in an OPEN message but will establish a BGP session with the Cisco NSF-capable device. This functionality will allow interoperability with non-NSF-aware BGP peers (and without NSF functionality), but the BGP session with non-NSF-aware BGP peers will not be capable of graceful restart.

BGP support for Cisco NSF requires that neighbor routers are NSF-aware or NSF-capable. Cisco NSF awareness in BGP is also enabled by the graceful restart mechanism. A router that is NSF-aware functions like a router that is NSF-capable, with one exception: An NSF-aware router is cannot perform an SSO operation. However, a router that is NSF-aware can maintain a peering relationship with a NSF-capable neighbor during an NSF SSO operation and can hold routes for this neighbor during the SSO operation.

https://t.me/learningnets

Cisco Nonstop Routing

The BGP support for NSR with SSO feature enables PE routers to maintain a BGP state with CE routers and ensure continuous packet forwarding through an RP. CE routers need not be Cisco Nonstop Forwarding-capable or NSF-aware to benefit from BGP NSR capabilities on PE routers. Only PE routers need to be upgraded to support BGP NSR—no CE router upgrades are required, which means that BGP NSR with SSO enables service providers to provide the benefits of Cisco NSF with the additional benefits of NSR, without requiring CE routers to be upgraded to support BGP graceful restart.



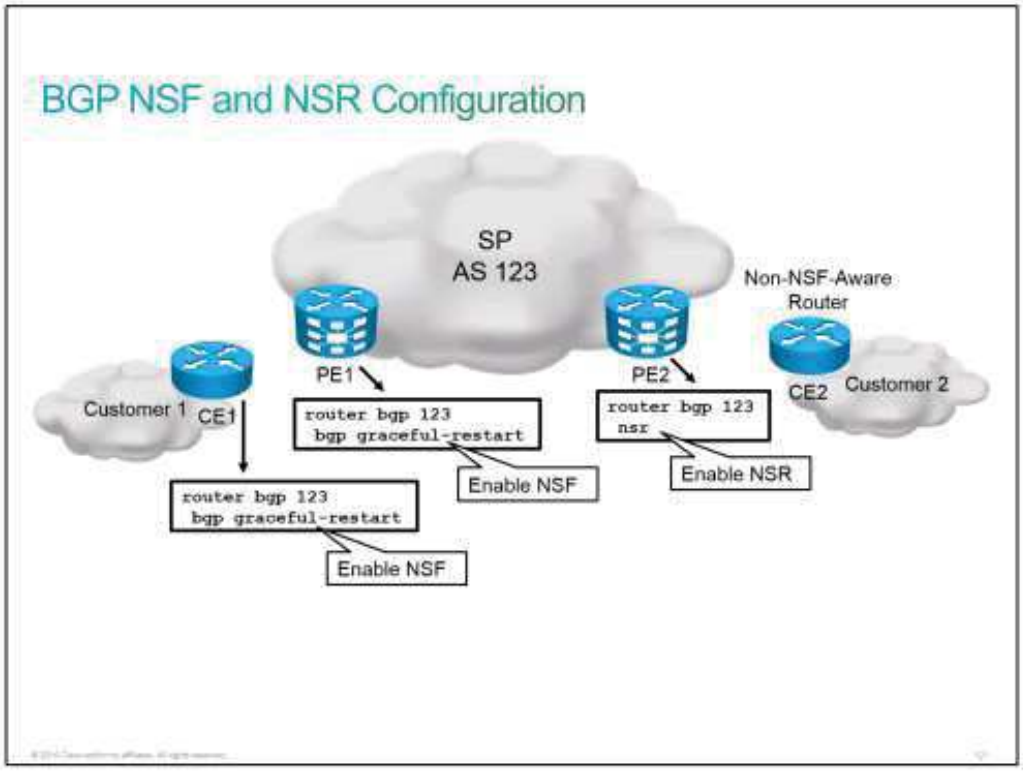
Before the introduction of BGP NSR with SSO, BGP required that all neighboring devices participating in BGP NSF be configured to be either Cisco NSF-capable or NSF-aware (through configuration of the devices to support the BGP graceful restart mechanism). That is, BGP NSF required that all neighboring devices were upgraded to a version of software that supported BGP graceful restart. However, in many deployments, there are situations where PE routers engage in exterior EBGP peering sessions with CE routers that do not support BGP graceful restart, and cannot be upgraded to a software version that supports BGP graceful restart in the same time frame as the provider routers.

BGP NSR with SSO provides a high-availability solution to service providers whose PE routers engage in EBGP peering relationships with CE routers that do not support BGP graceful restart. BGP NSR works with SSO to synchronize BGP state information between the active and standby RPs. SSO minimizes the amount of time a network is unavailable to its users following a switchover. When the BGP NSR with SSO feature is configured, in the event of an RP switchover, the PE router uses BGP NSR with SSO to maintain the BGP state for EBGP peering sessions with CEs that are not Cisco NSF-aware. The BGP NSR with SSO feature also dynamically detects NSF-aware peers and runs graceful restart with those CE routers. For EBGP peering sessions with NSF-aware peers and for internal IBGP sessions with BGP route reflectors in the service provider core, the PE uses Cisco NSF to maintain the BGP state. Therefore, BGP NSR with SSO enables service providers to provide the benefits of Cisco NSF with the additional benefits of NSR, without requiring CE routers to be upgraded to support BGP graceful restart.

https://t.me/learningnets

BGP NSF and NSR Configuration

The BGP NSF and NSR configuration is shown below:



The figure shows the configuration of Cisco Nonstop Forwarding on Cisco IOS, IOS XE, and IOS XR Software platforms. To enable NSF support for BGP, enter BGP configuration mode and use the **bgp graceful-restart** command.

To configure NSR support on the Cisco IOS XR Software platform, use the **nsr** command in BGP configuration mode. Enter this command on only PE routers that are NSR-capable.

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

Configuration of NSR on Cisco IOS and Cisco IOS XE Software is different and is displayed here:

```
router bgp 123
 bgp graceful-restart
 neighbor 209.165.201.130 ha-mode sso
```

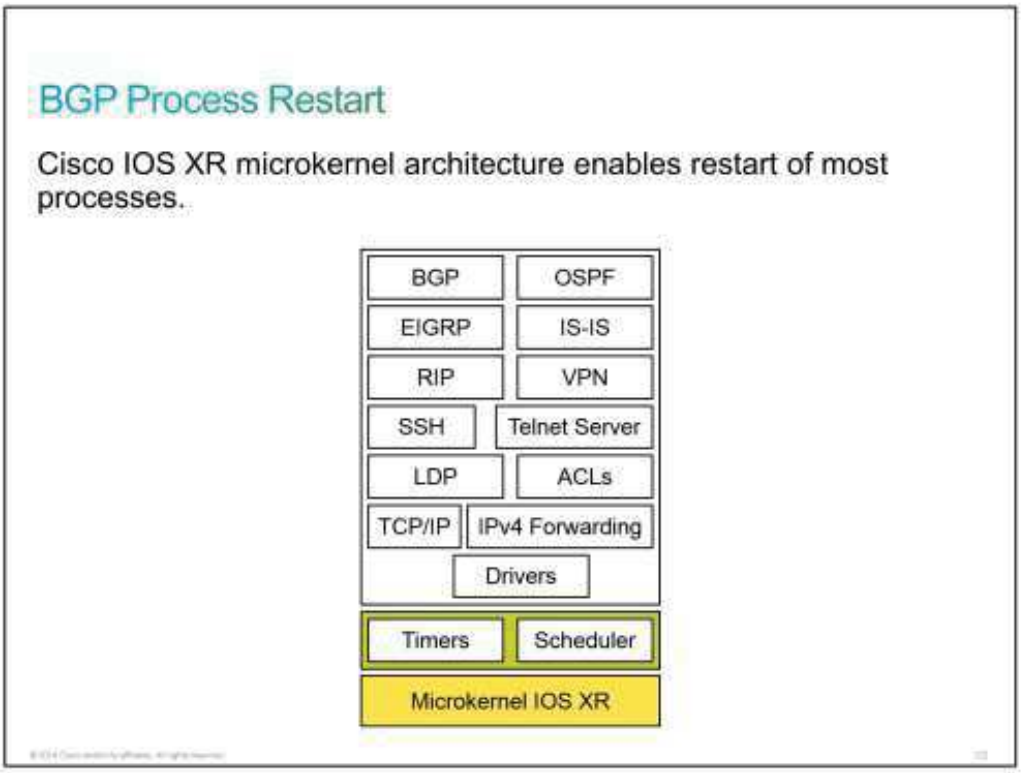
Note	For a complete command reference, refer to the Cisco IOS and IOS XE Software Command Reference guides on www.cisco.com.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

BGP Process Restart

Process restartability is another BGP optimization option that is available on Cisco IOS XR Software platforms. Critical control-plane processes (such as the BGP routing process) can be both manually and automatically restarted in response to a process failure, as opposed to restarting the entire operating system. This feature supports the Cisco IOS XR goal of continuous system availability, and allows for quick recovery from process or protocol failures with minimal disruption to customers or traffic.



You can manually restart the BGP process by entering the admin mode and using the **process restart** command, followed by the job ID of the BGP process. You can determine the BGP process ID using the **show processes bgp** command.

The following is a sample output from the **show processes bgp** command:

https://t.me/learningnets

```
RP/0/RSP0/CPU0:PE1# show processes bgp
                  Job Id: 1047
                     PID: 25186559
         Executable path: /disk0/iosxr-routing-4.1.0/bin/bgp
              Instance #: 1
              Version ID: 00.00.0000
                 Respawn: ON
           Respawn count: 8
Max. spawns per minute: 12
            Last started: Wed Nov 16 21:26:29 2011
           Process state: Run
           Package state: Normal
       Started on config: ipc/gl/ip-bgp/meta/speaker/default
                    core: MAINMEM
                Max. core: 0
               Placement: Placeable
            startup path: /pkg/startup/bgp.startup
                   Ready: 0.614s
               Available: 11.366s
        Process cpu time: 40.183 user, 7.763 kernel, 47.946 total
<… output omitted …>
```

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Summary

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

Summary

- BGP security and optimization options are used in the infrastructure layer of the Cisco IP NGN.
- BGP as any service can be a target of malicious attacks.
- Protection mechanisms can reduce the risk of an attack if implemented.
- Maximum prefix limitation will shut down peering session if too many prefixes are received.
- BGP passwords add security mechanism that prevents arbitrary connections even if other parameters match.
- TTL security check can be used to verify hop distance of the peer.
- Control plane policing can limit the amount of traffic sent to the control plane of the router and thus reduce the possibility of an attacker overwhelming the CPU.
- All these features are configured on per-neighbor basis.

Summary (Cont.)

- RTBH makes it possible for the network to signal border routers which traffic to discard to reduce the effects of DDoS attacks based on destination or source addresses.
- With the destination-based RTBH implementation, traffic going to the IP address of the victim is discarded on the edge of the service provider.
- With the source-based RTBH implementation, traffic coming from the IP addresses of the attacker is discarded on the service provider edge.
- NSF allows forwarding to continue in the event of supervisor failover.
- NSR allows NSF capabilities with routers that are not NSF-aware.
- BGP supports both NSF and NSR.
- In case of a software failure IOS XR supports automatic recovery by restarting BGP processes.

https://t.me/learningnets

https://t.me/learningnets

Lesson 2: Improving BGP Convergence

Overview

As the number of routes in the Internet increases, demands on CPU and memory resources on the routers in a service provider will increase. BGP processing affects both router resources and network convergence time. Network convergence must be as fast as possible to ensure accurate routing information between domains. Router resources also must be optimized whenever possible.

The lesson first describes BGP features that are available to improve BGP convergence. The lesson then describes BGP timers and intervals that can also improve convergence.

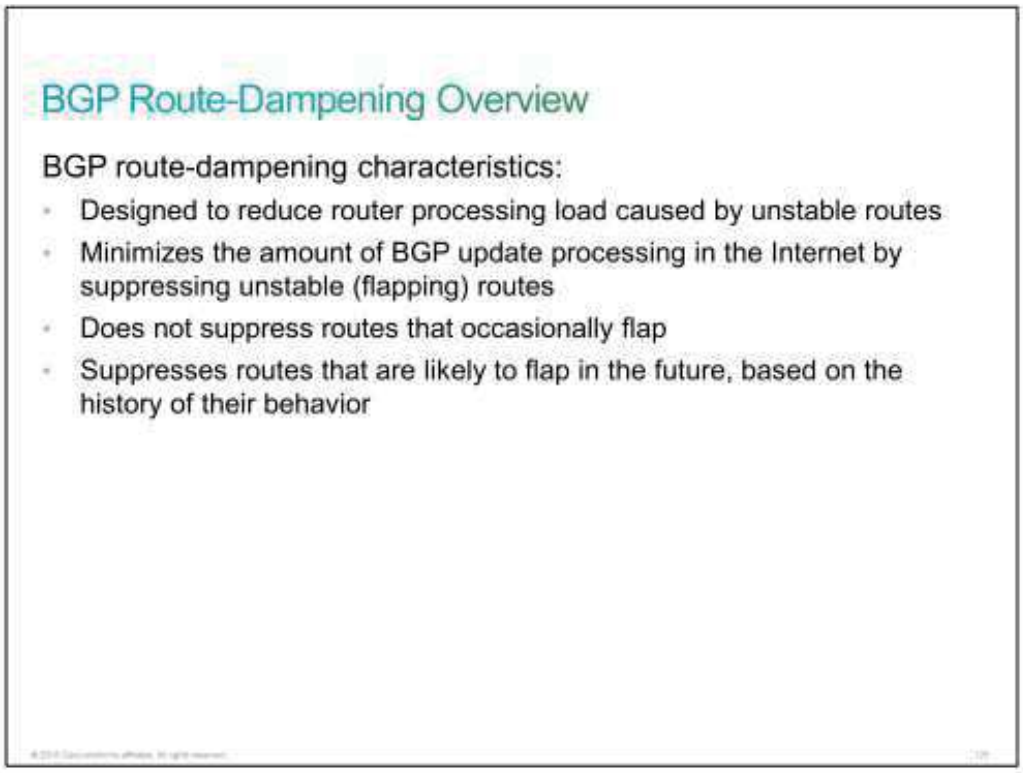
Upon completing this lesson, you will be able to:

- Describe the purpose and the operation of BGP route dampening
- Describe the configuration of BGP route dampening
- Describe the verification of BGP route dampening
- Describe BGP convergence
- Describe BGP processes
- Describe the features that can be used to improve BGP convergence
- Describe using distributed BGP
- Describe using PMTU discovery
- Describe increasing the input queue depth
- Explain PMTU discovery, hold queue, and distributed BGP configuration
- Describe the BGP prefix independent convergence feature and configuration
- Describe the BGP BFD feature and configuration
- Explain BGP timer and interval configuration

https://t.me/learningnets

BGP Route-Dampening Overview

A "flap" refers to a route that is repeatedly available, then unavailable, then available, then unavailable, and so on. BGP is the only routing protocol that is designed for large internetworks with the specific intention of carrying a large number of prefixes. Several mechanisms are built into BGP that ensure maximum router stability.



For example, a BGP router does not forward BGP routing updates immediately after receiving them. Every time a BGP router sends an update, it starts a 30-second timer (advertisement interval) for external neighbors. No new updates can be sent until that timer expires. The result is that if a router contains a link that is flapping at a rate of once per second, external routers see the flap at a much slower rate. Routers that are external to the source of the flap are not forced to recalculate the best path every second but, at most, every 30 seconds.

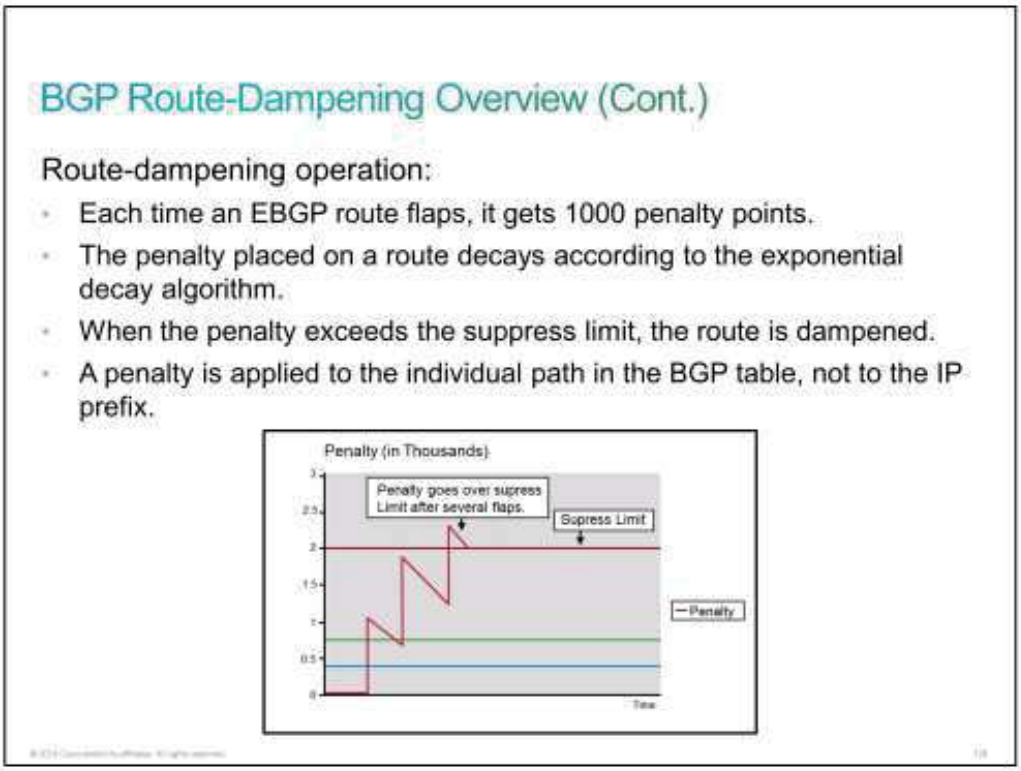
Reducing the rate at which neighboring routers process flapping routes helps to reduce the requirements to process the BGP update. However, routers that process routing updates for unstable routes are still wasting resources in determining the best route to the destination. Because the unstable route is oscillating between up and down, each route update that a router receives causes it to process the unstable route all over again. A better approach is to remove the update about the route until it can be guaranteed that the destination is more stable.

With this goal in mind, an additional BGP scalability mechanism called route dampening was created to reduce route update processing requirements by suppressing unstable routes.

Most service providers hold routing information for the entire Internet. Therefore, a flapping link somewhere in the Internet can cause all routers in the Internet to keep processing changes because of one single link. If, however, one of the autonomous systems in the Internet implements route dampening, the flapping network is suppressed. The route is not propagated further to other autonomous systems until the configured rules of route dampening allow it.

https://t.me/learningnets

A route that flaps once or twice is typically not considered a flap from an administrative perspective. If the flapping happens more often, however, something is probably wrong with the destination and the route should be suppressed. The BGP router stores a suppressed route in the BGP table but does not consider it in the BGP path-selection process, and does not therefore propagate it to other BGP neighbors or use it for data forwarding.



A BGP router with route dampening enabled tracks all routes (even those routes that are unreachable) so that it can recall the penalties that are assigned to each route. Every time a route flap occurs, the flapping route receives 1000 penalty points. The penalty is gradually decreased through the use of a decaying algorithm. A route that flaps several times will be penalized (gain enough penalty points) and then reach and exceed the suppress limit.

Any route that reaches the suppress limit is no longer forwarded to other neighbors until the assigned penalty is once again below the reuse limit. An exponential decay algorithm, which is automatically calculated based on the half-life time, reduces penalty points that are applied to a flapping route.

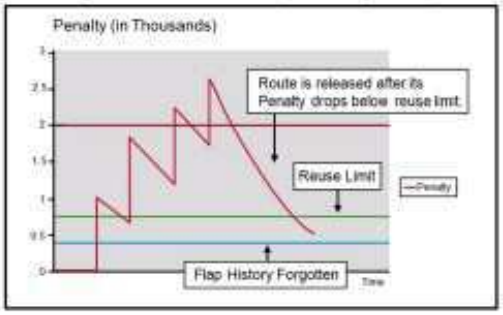
A penalty is always applied to a path and not a prefix. One path flapping does not mean that the destination is flapping because the destination might be reachable over other paths.

https://t.me/learningnets

BGP Route-Dampening Overview (Cont.)

Route-dampening operation continuation:

- The route is never dampened for longer than the maximum suppression time limit.
- An unreachable route with a flap history is put in the history state.
- A dampened route is propagated when the penalty drops below the reuse limit.
- The flap history is forgotten when the penalty drops below half of the reuse limit.



The maximum suppress limit defines the maximum duration that a route can be suppressed.

After route dampening is enabled, the router never removes a route from the BGP table. A route that has been withdrawn by a BGP neighbor can still be seen in the BGP table and is marked with an "h" (history state).

After the number of penalty points that are assigned to a route falls below the reuse limit, the BGP router once again advertises the route.

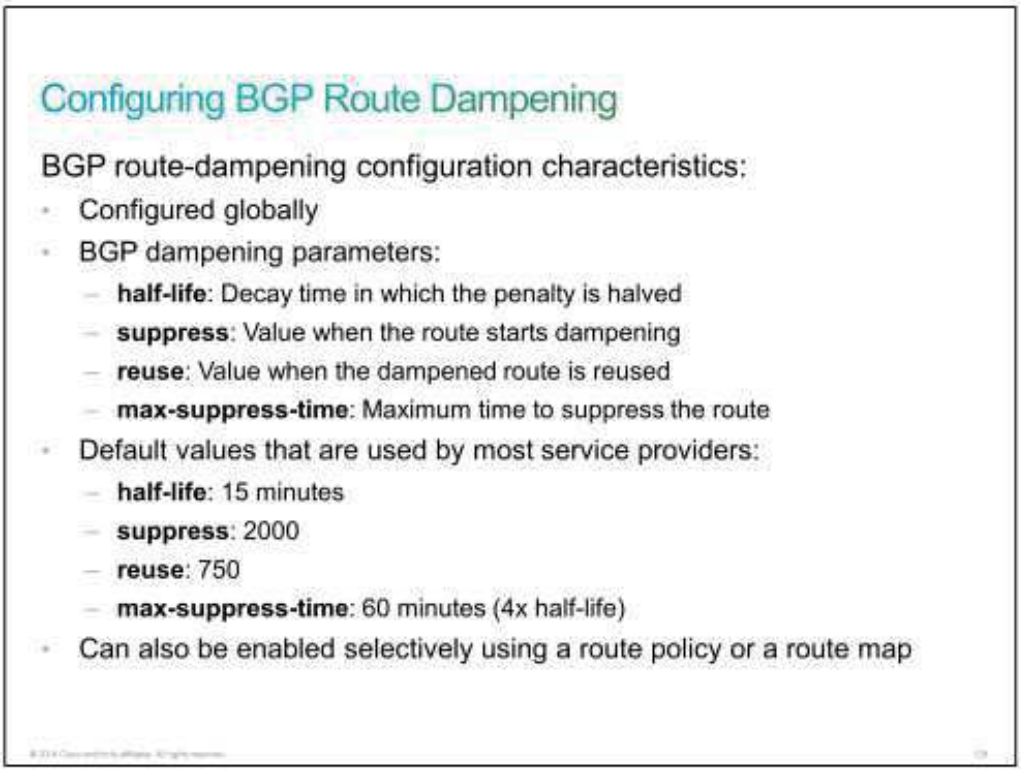
A router stops tracking penalty points when they are below half of the reuse limit.

 © 2014 Cisco Systems, Inc.

https://t.me/learningnets

Configuring BGP Route Dampening

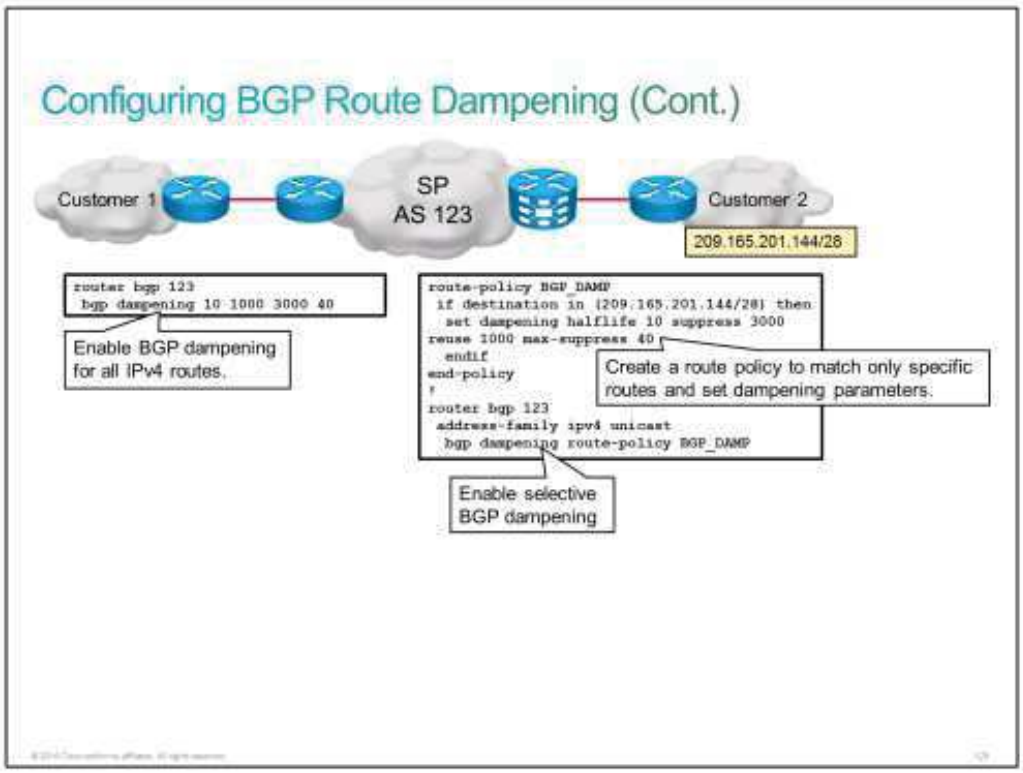
Route dampening can be enabled globally for all routes, or it can be enabled selectively, with different parameters for different sets of routes using route maps or route policies.



BGP route dampening is enabled globally. Route flap dampening configuration supports the following parameters:

- **half-life**: The half-life is the time that is needed for the penalty to halve (default is 15 minutes).
- **suppress**: When a route has more penalty points than the suppress limit, the route is suppressed (default is 2000).
- **reuse**: After the flapping has stopped and the penalty for a route has fallen below the reuse limit, the route is unsuppressed (default is 750). Every 10 secs, the router sees which routes are unsuppressed and advertises them again.
- **max-suppress-time**: No route can be suppressed longer than the **max-suppress-time** parameter (default is 60 minutes, maximum is 255 minutes).

https://t.me/learningnets



To configure BGP route dampening on Cisco IOS, IOS XE, and IOS XR Software using default parameters, use the **bgp dampening** command under BGP router configuration mode, under the desired address family.

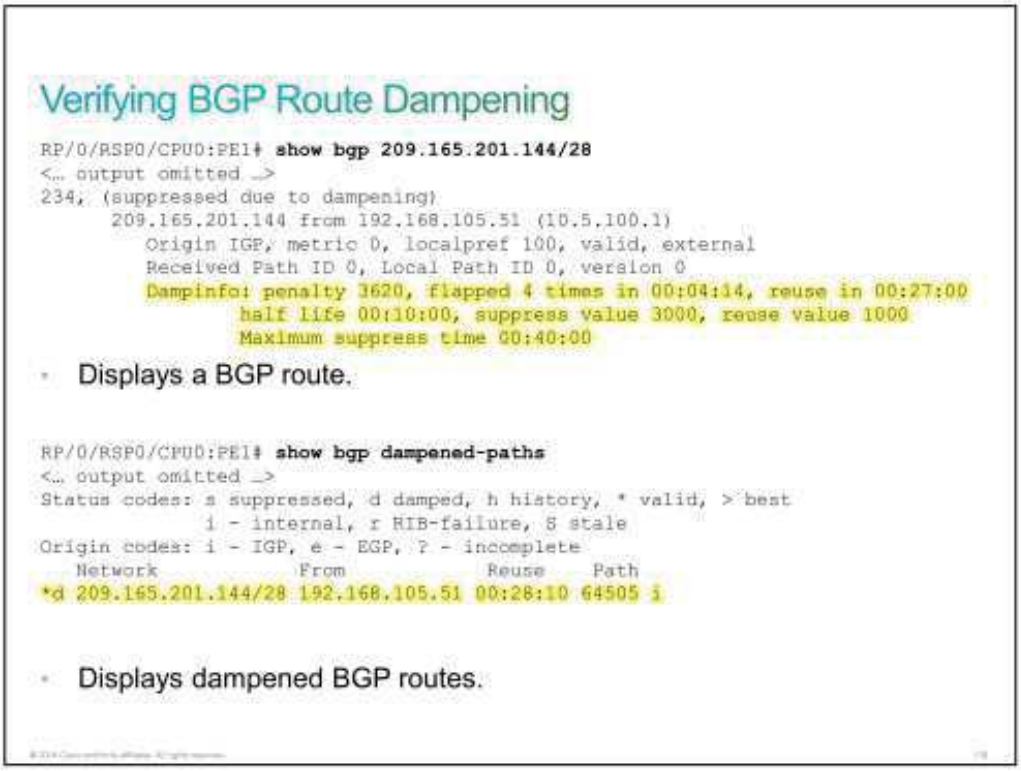
The example on the Cisco IOS Software router shows how to enable route dampening for all IPv4 routes and how to change dampening parameters. In the example, the half-life is set to 10 minutes, the reuse limit is set to 1000, the suppress limit is set to 3000, and the maximum suppress time is set to 40 minutes. To enable route dampening for IPv6, first enter IPv6 address family configuration mode.

The example on the Cisco IOS XR Software router shows how to enable route dampening selectively for a particular set of routes. In the example, a route policy is configured that matches the prefix of the customer. Then dampening parameters are set for matched routes. To enable selective dampening, enter BGP router configuration mode and enter the desired address family configuration mode. Then enable BGP dampening using the **bgp dampening** command, followed by the route policy name.

https://t.me/learningnets

Verifying BGP Route Dampening

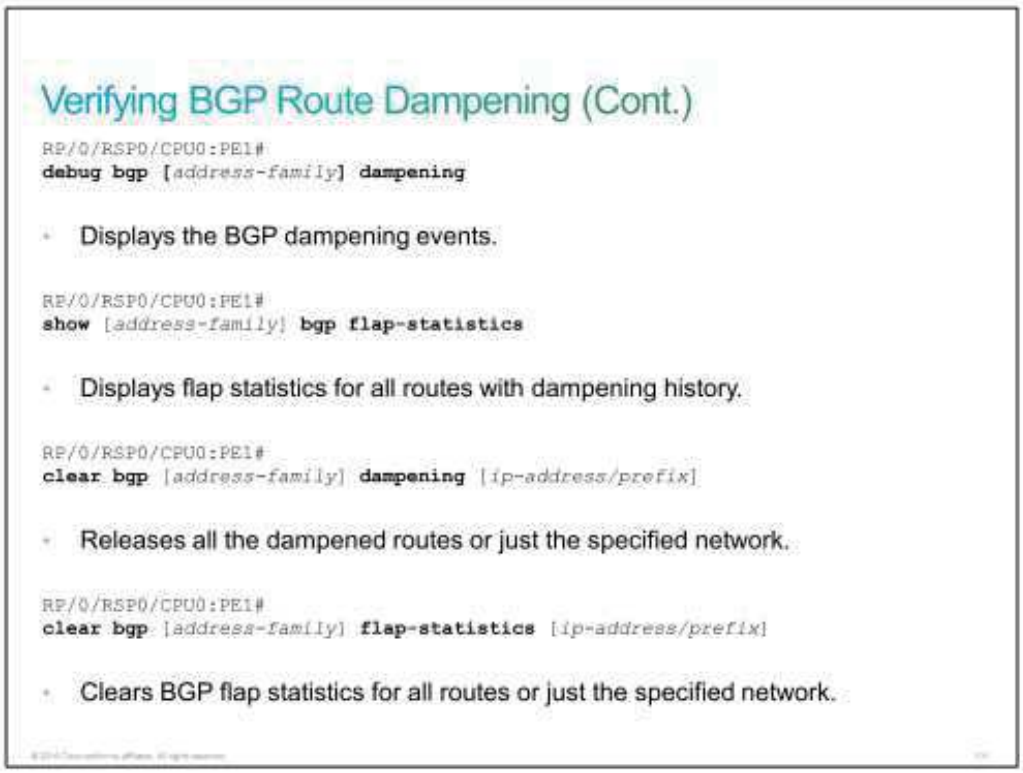
This topic describes verification of BGP route dampening.



To verify if a BGP route has any BGP route dampening penalties, use the **show bgp** command, followed by a prefix. In the example, the route flapped four times and received a penalty of 3620. Because the suppress limit is set to 3000, the route will be suppressed for a maximum of 40 minutes.

To verify suppressed routes, use the **show bgp dampened-paths** command. You should see all routes that are suppressed. If you display the BGP table, you should see that the dampened routes are marked with "d" in the first column.

https://t.me/learningnets



You can also use the following commands to monitor BGP route dampening:

- Use the **debug bgp dampening** command to display BGP dampening events as they occur in real time.
- Use the **show ip bgp flap-statistics** command to list all routes that have a penalty that is still above the time-to-forget limit. You can also use this command in combination with regular expressions and filter lists.
- To clear BGP route-dampening information and unsuppress the suppressed routes, use the **clear bgp dampening** command.
- To clear BGP flap statistics, use the **clear bgp flap-statistics** command.

The following messages are examples of BGP dampening events, displayed as a result of enabling BGP dampening debugging.

BGP dampening is enabled with default parameters. When a route flaps for the second time, the penalty is set to 1950:

```
RP/0/RSP0/CPU0:PE1#RP/0/RSP0/CPU0:Oct 14 07:23:13.551 : bgp[1048]: [1-rtr] (ip4u):
Charge penalty for 209.165.201.144/28 path 64505 with halflife-time 15 min
reuse/suppress 750/2000 Flapped 2 times in 00:01:12. New penalty is 1950
```

When a route flaps for the third time, the penalty is set to 2875 and exceeds the suppress limit:

```
RP/0/RSP0/CPU0:PE1#RP/0/RSP0/CPU0:Oct 14 07:24:14.014 : bgp[1048]: [1-rtr] (ip4u):
Charge penalty for 209.165.201.144/28path 64505 with halflife-time 15 min
reuse/suppress 750/2000 Flapped 3 times in 00:02:12. New penalty is 2875
```

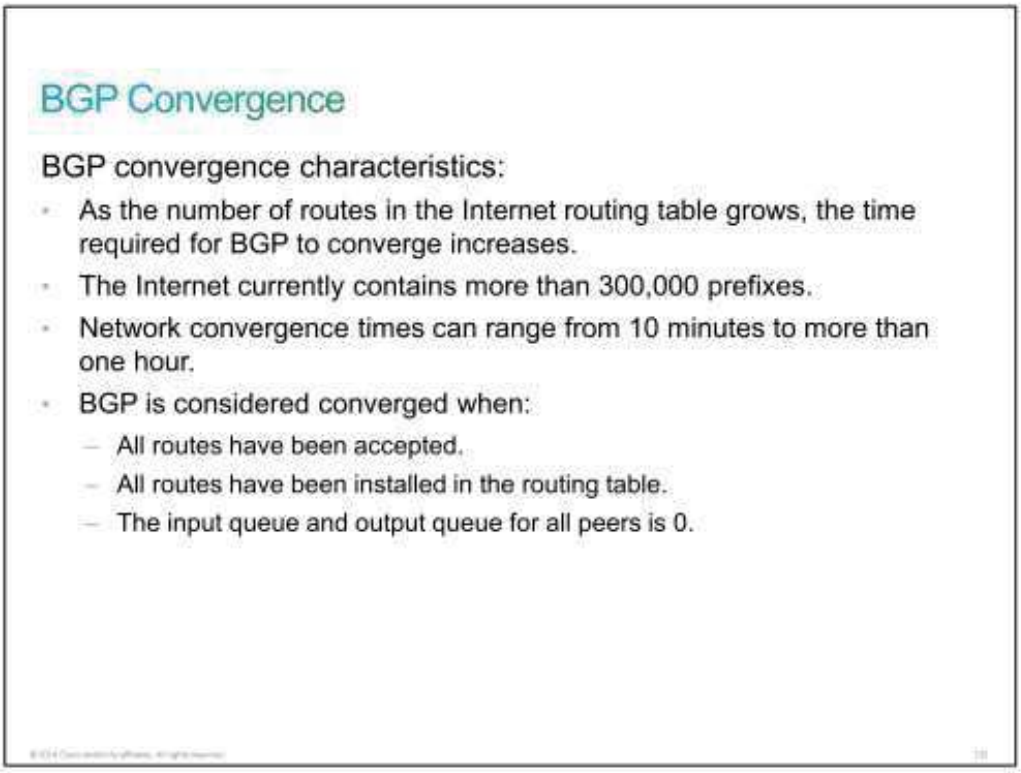
The route is suppressed:

```
RP/0/RSP0/CPU0:PE1#RP/0/RSP0/CPU0:Oct 14 07:24:44.734 : bgp[1048]: [1-rtr] (ip4u):
Suppress 209.165.201.144/28 path 64505 for 00:28:30 (penalty 2809) halflife-time 15,
reuse/suppress 750/2000
```


https://t.me/learningnets

BGP Convergence

As the number of routes in the Internet routing table grows, service providers and large enterprise customers are experiencing a dramatic increase in the amount of time that BGP takes to converge. Networks that once converged in 10 or 15 minutes may now take up to an hour in some cases and even longer in extreme situations. In general, convergence is defined as the process of bringing all routing tables to a state of consistency.



The BGP routing protocol is considered converged when the following conditions are true:

- All routes have been accepted.
- All routes have been installed in the routing table.
- The input queue and output queue for all peers is 0.

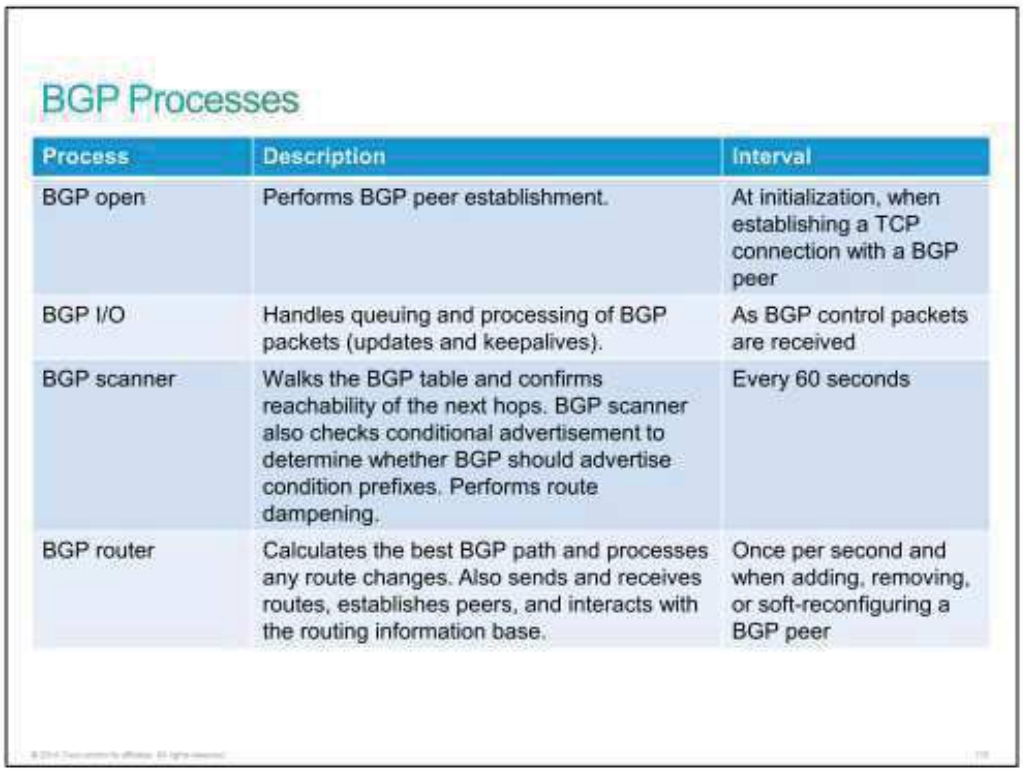
Convergence time is an important consideration in a network because unconverged networks can cause routing loops, packet delays, and even packet loss as a result of black holes.

https://t.me/learningnets

BGP Processes

In general, a process on a router consists of the individual threads and associated data that perform router tasks, such as system maintenance, packet switching, and implementing routing protocols.

BGP consists of several processes:

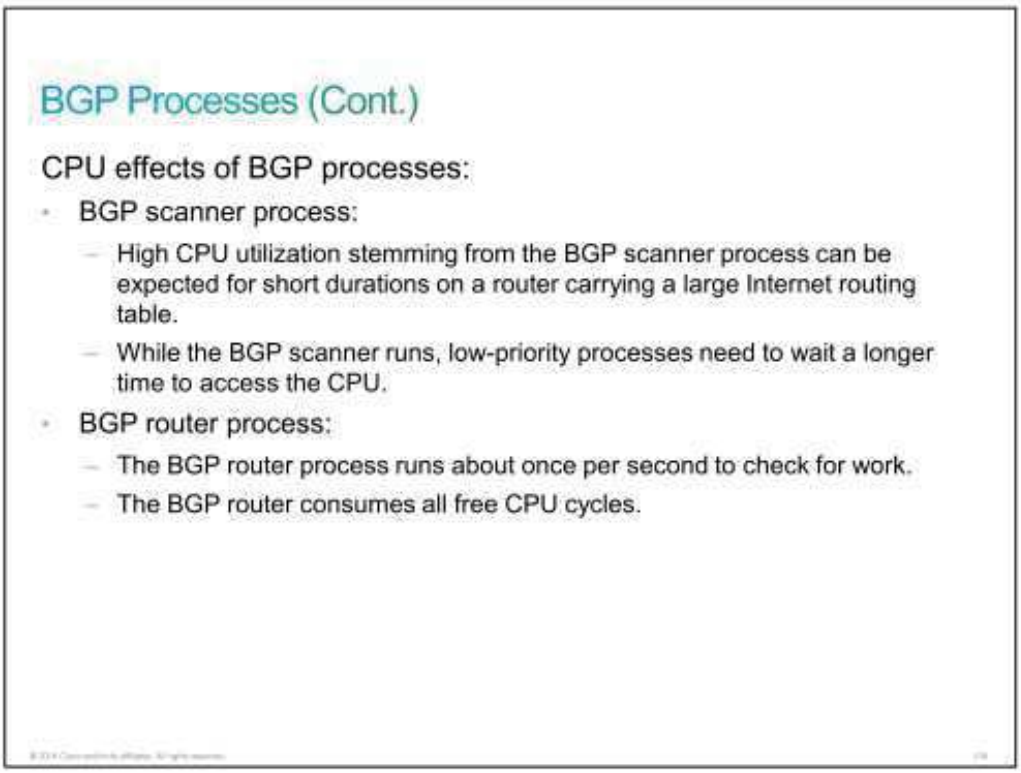


Note	A thread is an information placeholder that allows a single process to be halted (interrupted) on the router so that the CPU can service another process. The information that is contained within the thread allows the interrupted process to restart exactly where it left off when the CPU is ready to continue to service that process thread.

Several processes that are executed on the router enable BGP to run. You can use the **show process cpu | include BGP** command on Cisco IOS and IOS-XE or the **show processes bgp** on Cisco IOS-XR to display the volume of CPU resources that are consumed (utilization) because of running BGP processes.

The figure lists the function of each BGP router process and how often each process is executed on the router. It shows that each process runs at different times, depending on the tasks that are managed by the specific process. Because the BGP scanner and BGP router processes are responsible for a large number of calculations, you may notice high CPU utilization during the running of either of these processes.

https://t.me/learningnets



On routers that carry a large Internet routing table, high CPU utilization stemming from the BGP scanner process can be expected for short periods of time. By default, the BGP scanner "walks" (scans) the BGP routing table once per minute and performs important maintenance tasks. These tasks include checking the next hop that is referenced in the BGP table of the router and verifying that the next-hop devices can be reached. Thus, a large BGP table takes an equally large amount of time to be walked and validated.

The BGP scanner walks the BGP routing table to update any data structures and for route redistribution purposes. In this context, the routing table is also known as the RIB, which the router outputs when the **show ip route** command is executed. Both tables are stored separately in the router memory and can be very large, thus consuming CPU and memory resources.

Because the BGP scanner runs through the entire BGP table, the duration of the high CPU utilization condition that is caused by the BGP scanner process varies with the number of neighbors and the number of routes that are learned per neighbor.

While the BGP scanner runs, low-priority processes need to wait a longer time to access the CPU. One low-priority process controls ICMP packets such as pings. Packets that are destined to or have originated from the router may experience higher than expected latency because the ICMP process must wait behind the BGP scanner. The BGP scanner process runs for some time and is suspended, then ICMP runs and is suspended, then the BGP scanner runs, and so on. In contrast, pings sent through a router should be switched via Cisco Express Forwarding and should not experience any additional latency. When you are troubleshooting periodic spikes in latency, compare forwarding times for packets that are forwarded through a router versus packets that are processed directly by the CPU on the router.

The BGP router process runs about once per second to check for work. BGP convergence defines the duration between the time when the first BGP peer is established and the point at which BGP is converged. To ensure the shortest possible convergence times, the BGP router consumes all free CPU cycles. However, after it starts, it relinquishes (or suspends) the CPU intermittently.

https://t.me/learningnets

Example: CPU Effects of BGP Processes

Convergence time is a direct measurement of how long the BGP router process runs on the CPU, not the total time that the process is actually running. This example investigates the high CPU utilization condition on a Cisco IOS Software router during BGP convergence as BGP exchanges prefixes with two EBGP peers.

- Capture a baseline for normal CPU utilization before starting the test.

```
router# show process cpu
CPU utilization for five seconds: 0%/0%; one minute: 4%; five minutes: 5%
```

- After the test starts, the CPU reaches 100 percent utilization. The **show process cpu** command shows that the high CPU condition is caused by the BGP router, denoted by 139 (the Cisco IOS process ID for the BGP router) in the following output:

```
router# show process cpu
CPU utilization for five seconds: 100%/0%; one minute: 99%; five minutes:
81%[output omitted] 139      6795740    1020252    6660 88.34% 91.63% 74.01%    0 BGP
Router
```

- Monitor the router by capturing multiple outputs of the **show ip bgp summary** and **show process cpu** commands during the event. The **show ip bgp summary** command captures the state of the BGP neighbors.

```
router# show ip bgp summary
Neighbor        V    AS  MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
10.1.1.1        4 64512   309453  157389    19981    0  253 22:06:44 111633
172.16.1.1      4 65101   188934    1047    40081   41    0 00:07:51 58430
```

- When the router completes prefix exchange with its BGP peers, the CPU utilization rates should return to normal levels. The computed 1-minute and 5-minute averages also will settle back down, but may show higher than normal levels for a longer period than the 5-second rate.

```
router# show process cpu
CPU utilization for five seconds: 3%/0%; one minute: 82%; five minutes: 91%
```

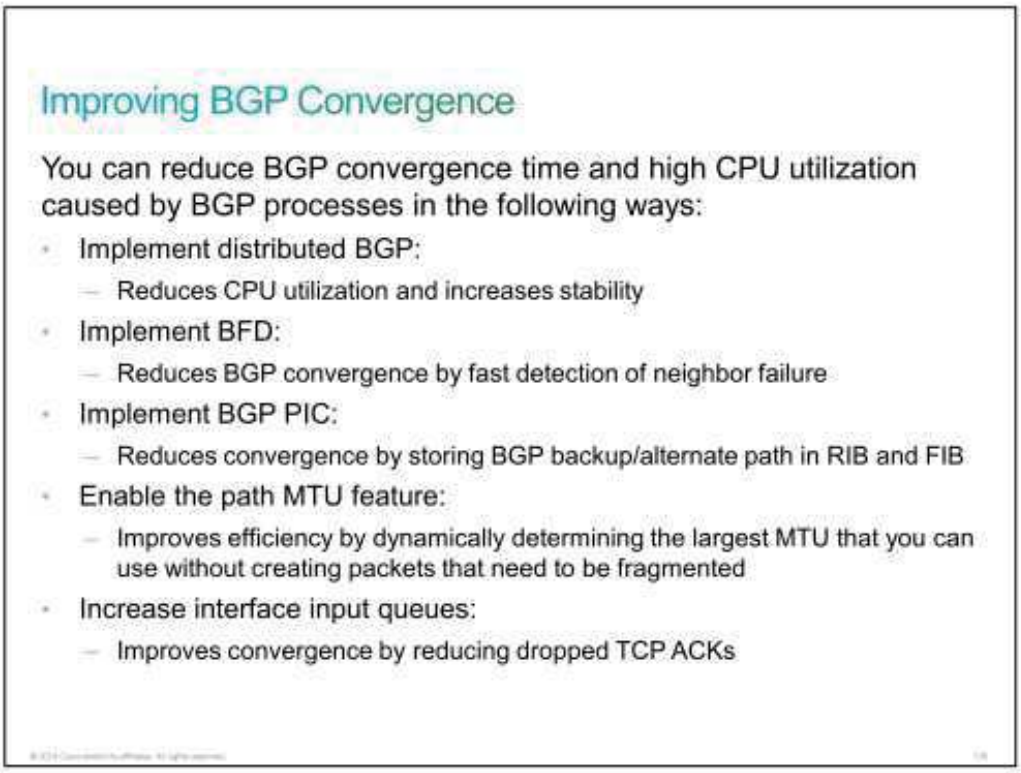
- Using the output from the **show** commands will allow you to compute the BGP convergence time. In particular, the Up/Down column of the **show ip bgp summary** command is compared to the start and stop times of the high CPU utilization condition. Typically, BGP convergence can take several minutes when routers exchange a large Internet routing table.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Improving BGP Convergence

BGP convergence can often be an issue in networks that require quick propagation of routing information.



To reduce BGP convergence time and the high CPU utilization that is caused by a running BGP process, the following performance improvement features are available:

- **Distributed BGP**: Available on Cisco IOS XR Software routers. Distributed BGP splits BGP functionality into three process types and is used to reduce the impact that a fault in one address family has on another address family. It also reduces CPU utilization.
- **Bidirectional Forwarding Detection:** BFD is a detection protocol designed to provide fast forwarding path failure detection times for all media types, encapsulations, topologies, and routing protocols.
- **BGP prefix independent convergence:** The BGP PIC feature improves BGP convergence after a network failure. This convergence is applicable to both core and edge failures and can be used in both IP and MPLS networks. The BGP PIC feature creates and stores a backup/alternate path in the RIB, FIB, and Cisco Express Forwarding. When a failure is detected, the backup/alternate path can immediately take over, thus enabling fast failover.
- **Path MTU feature:** All TCP sessions are bounded by a limit on the number of bytes that a single packet can transport. This limit, which is known as the MSS, is 536 bytes by default. That is, TCP breaks up packets in a transmit queue into 536-byte chunks before passing packets down to the IP layer. The advantage of a 536-byte MSS is that packets are not likely to be fragmented at an IP device along the path to the destination, because most links use an MTU of at least 1500 bytes. The disadvantage is that smaller packets increase the amount of bandwidth that is used for transport overhead.

https://t.me/learningnets

Because BGP builds a TCP connection to all peers, a 536-byte MSS affects BGP convergence times. The solution is to enable the PMTU feature. You can use this feature to dynamically determine how large the MSS value can be without creating packets that need to be fragmented. PMTU allows TCP to determine the smallest MTU size among all links in a TCP session. TCP then uses this MTU value, minus room for the IP and TCP headers, as the MSS for the session. The PMTU feature may be enabled by default, depending on the software version running on a router.

- **Increase interface input queues:** If BGP is advertising thousands of routes to many neighbors, TCP must transmit thousands of packets. BGP peers receive these packets and send TCP ACKs to the advertising BGP speaker, causing the BGP speaker to receive a flood of TCP ACKs in a short period of time. If the ACKs arrive at a rate that is too high for the router CPU, packets back up in inbound interface queues.

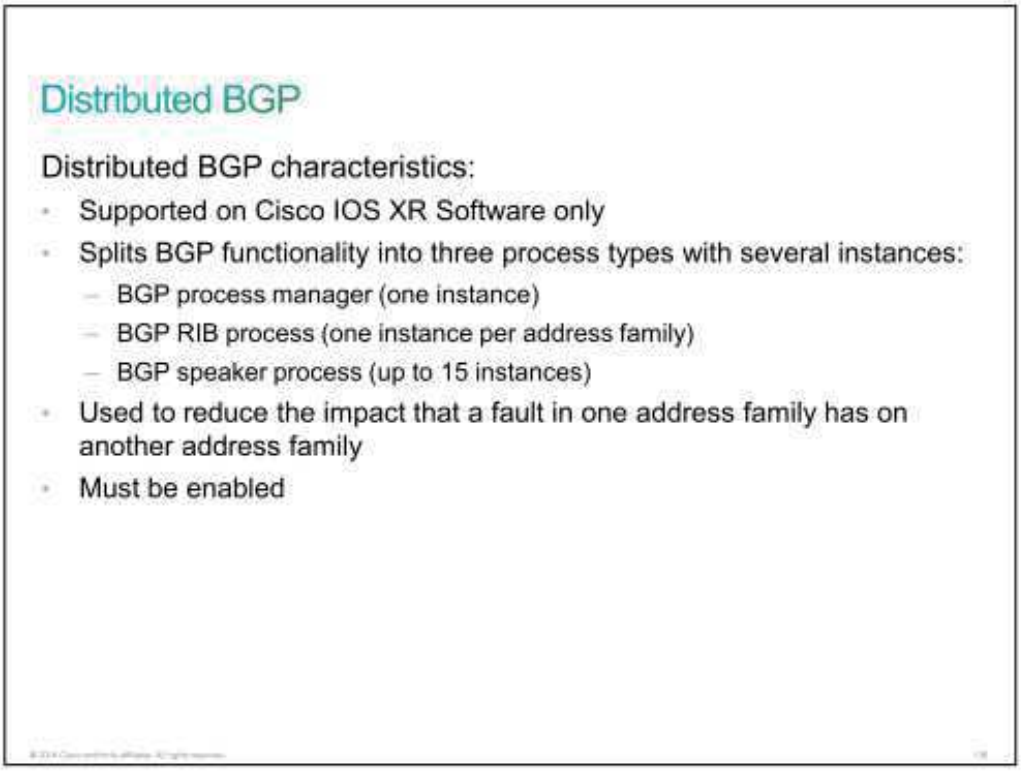
For example, Cisco IOS Software router interfaces use an input queue size of 75 packets. In addition, special control packets such as BGP updates use a special queue with SPD. This special queue holds 100 packets. During BGP convergence, TCP ACKs can quickly fill the 175 spots of input buffering, causing newly arriving packets to be dropped. On routers with 15 or more BGP peers that also exchange the complete Internet routing table, more than 10,000 drops per interface per minute may be seen.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Distributed BGP

Distributed BGP is available on Cisco IOS XR Software platforms.



Distributed BGP splits BGP functionality into three process types:

- **BGP process manager**: Responsible for verifying configuration changes and calculating and publishing the distribution of neighbors among BGP speaker processes. There is a single instance of this process.
- **bRIB process**: Responsible for performing the best-path calculation of routes (receives partial best paths from the speaker). The best route is installed into the bRIB and is advertised back to all speakers. The bRIB process is also responsible for installing routes in the routing table, and for managing routes that are redistributed from the routing table. To accommodate route leaking from one routing table to another, bRIB may register for redistribution from multiple routing table routes into a single route in the bRIB process. There is a single instance of this process for each address family.
- **BGP speaker process**: Responsible for processing all BGP connections to peers. The speaker stores received paths in the RIB and performs a partial best-path calculation, advertising the partial best paths to the bRIB (limited best-path calculation). Speakers perform a limited best-path calculation because, in order to compare MEDs, paths need to be compared from the same AS but may not be received on the same speaker. Because BGP speakers do not have access to the entire BGP local routing table, they can only perform a limited best-path calculation. Only the best paths are advertised to the bRIB to reduce speaker/bRIB IPC, and to reduce the number of paths to be processed in the bRIB. BGP speakers can mark a path as active only after learning the result of the full best-path calculation from the bRIB. Neighbor import and export policies are imposed by the speaker.

  There are multiple instances of this process in which each instance is responsible for a subset of BGP peer connections. Up to a total 15 speakers for all address families and one bRIB for each address family (IPv4, IPv6, and VPNv4) are supported.

https://t.me/learningnets

Distributed BGP is used to reduce the impact that a fault in one address family has on another address family. For example, you can have the following scenario:

- One speaker with only IPv6 neighbors (peering to IPv6 addresses)
- A separate speaker with only IPv4 neighbors (peering to IPv4 addresses)
- Another speaker with only VPNv4 PE or CE neighbors (peering to IPv4 addresses distinct from the non-VPN neighbors)

In this scenario, there is no overlap in processes (BGP, bRIB, and RIB) between IPv4, IPv6, and VPNv4. Therefore, a BGP, bRIB, or RIB process reload affects only one address family.

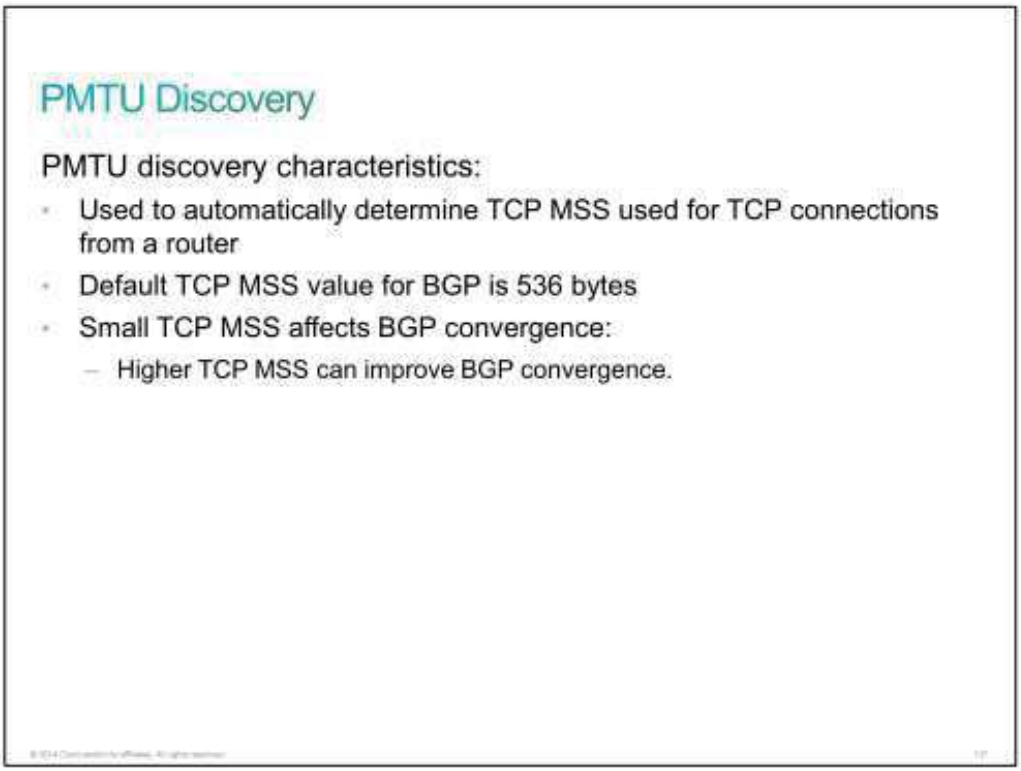
Distributed BGP also allows more CPU capacity for receiving, computing, and sending BGP routing updates. When the router is in distributed BGP mode, you can control the number of distributed speakers that are enabled and which neighbors are assigned to each speaker. If no distributed speakers are enabled, BGP operates in standalone mode. If at least one distributed speaker is enabled, BGP operates in distributed mode.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

PMTU Discovery

Another feature that can be used to reduce BGP convergence is PMTU discovery. PMTU discovery is a method for maximizing the use of available bandwidth in the network between the endpoints of a TCP connection.



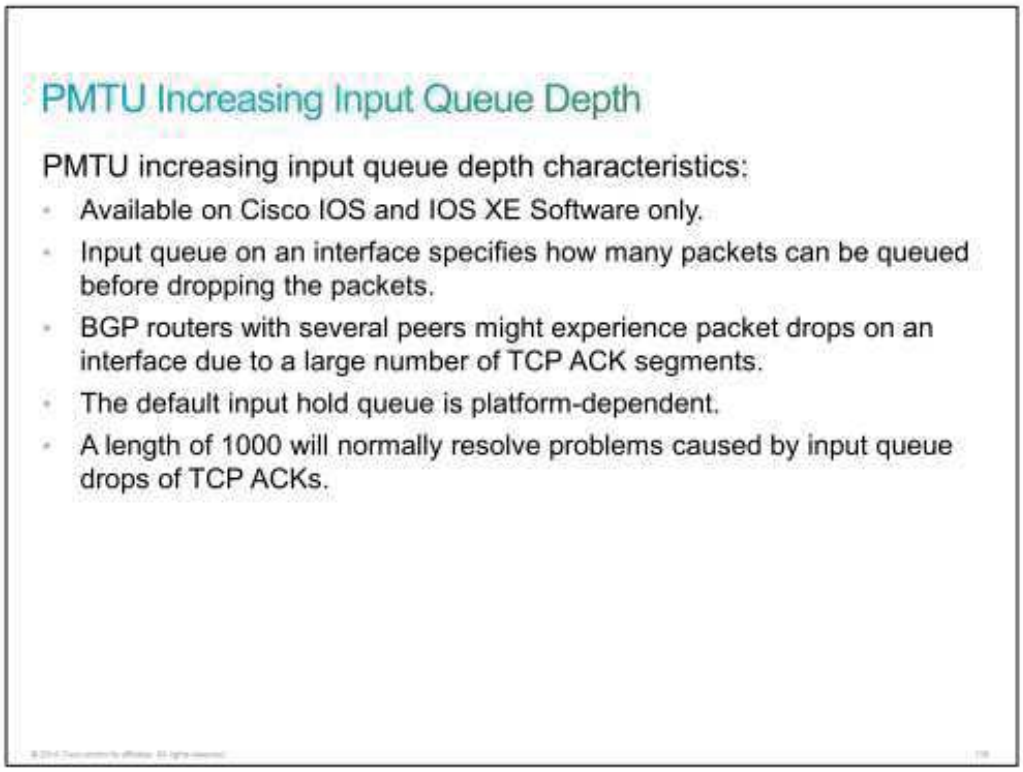
PMTU discovery works by setting the "don't fragment" option bit in the IP headers of outgoing packets. Then, any device along the path whose MTU is smaller than the packet will drop it and send back an ICMPFragmentation Needed (Type 3, Code 4) message containing its MTU, allowing the source host to reduce its PMTU appropriately. The process repeats until the MTU is small enough to traverse the entire path without fragmentation.

The default TCP MSS for BGP traffic is 536 bytes. Enabling PMTU discovery, and thus using a higher MSS for BGP traffic, can significantly improve BGP convergence because fewer packets are required to send BGP updates.

https://t.me/learningnets

PMTU Increasing Input Queue Depth

Another feature that can be used to improve BGP convergence on Cisco IOS and IOS XE Software platforms is to increase input queue depth. BGP routers might experience packet drops on an interface due to large number of TCP ACK segments, which are used to acknowledge receipt of BGP updates.



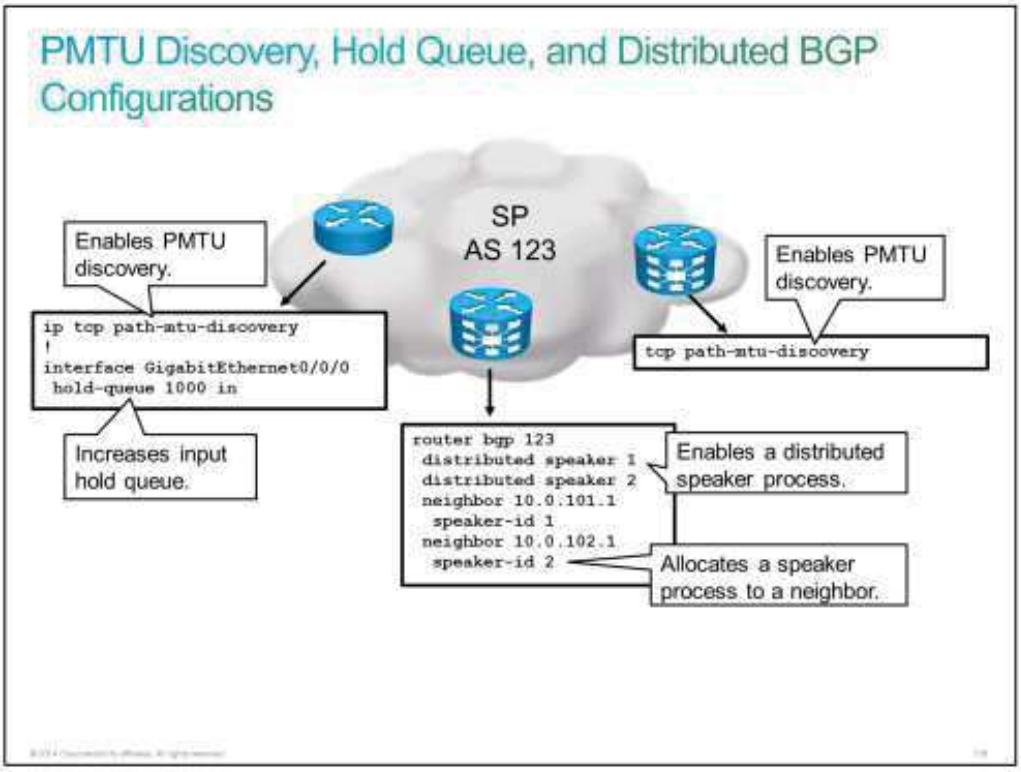
The default input hold queue is platform-dependent, and is 75 on Cisco IOS Software platforms and 375 on Cisco IOS XE Software platforms. Setting the length of an interface to 1000 will normally resolve problems that are caused by input queue drops of TCP ACKs.

Note	Consider several factors when increasing interface input queues. Increasing the interface input queues will increase memory utilization.

https://t.me/learningnets

PMTU Discovery, Hold Queue, and Distributed BGP Configurations

The figure shows a configuration to improve BGP convergence.



To enable PMTU discovery on Cisco IOS and IOS XE Software platforms, use the **ip tcp path-mtu-discovery** command in global configuration mode. On Cisco IOX XR Software platforms, use the **tcp path-mtu-discovery** command in global configuration mode.

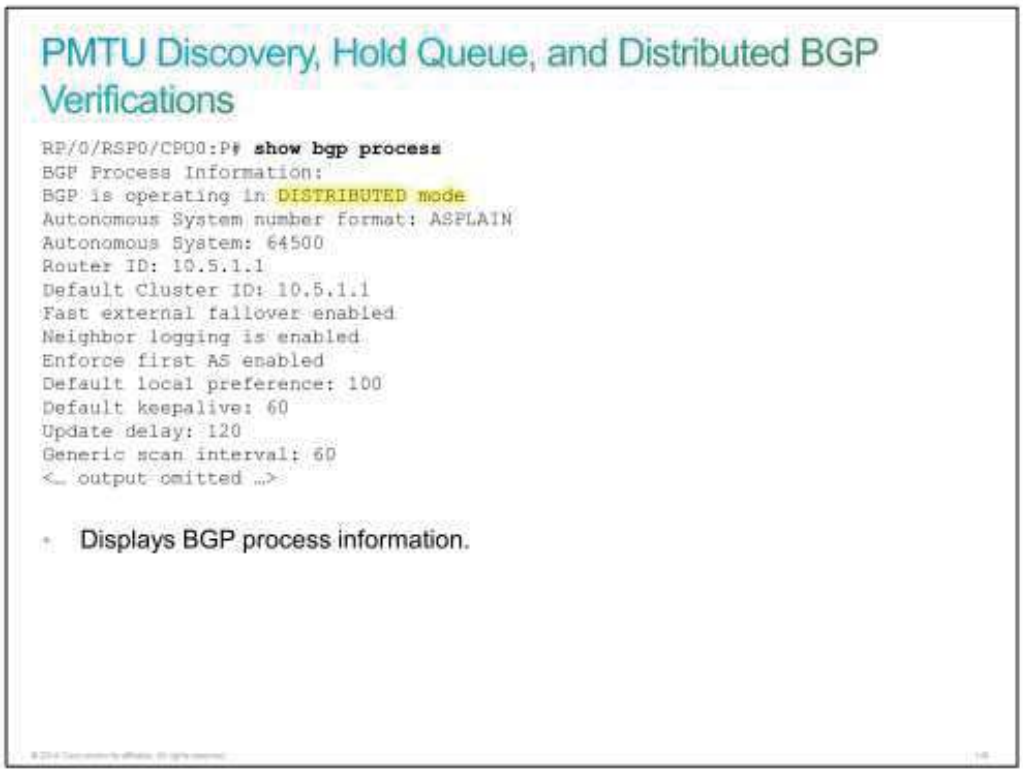
To change the interface input hold queue on Cisco IOS and IOS XE Software platforms, enter the interface configuration mode and use the **hold-queue** command, followed by packet length and directionality. To change the queue length for incoming packets, use the **in** keyword.

To enable distributed BGP on a Cisco IOS XR Software router, first enter BGP configuration mode. Then, create a BGP speaker process using the **distributed speaker** command, followed by the ID of the speaker process. The range is 1 to 15. Then, assign a BGP neighbor to the distributed speaker process by entering neighbor configuration mode and using the **speaker-id** command. In the example, two speaker processes are created. Neighbor 10.0.101.1 is assigned to process ID 1 and neighbor 10.0.102.1 is assigned to process ID 2. Finally, you must instruct the router to clear the current BGP mode and switch to distributed mode by using the **clear bgp current-mode** 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

PMTU Discovery, Hold Queue, and Distributed BGP Verifications



To verify whether BGP is running in distributed mode, use the **show bgp process** command. The output will show you the BGP mode, as shown in the example.

If BGP is not running in distributed mode, the output will read as follows:

```
BGP is operating in STANDALONE mode
```


https://t.me/learningnets

PMTU Discovery, Hold Queue, and Distributed BGP Verifications (Cont.)

```
PE1# show ip bgp neighbors | include Datagrams
Datagrams (max data segment is 1460 bytes):
<… output omitted …>
```

- Displays BGP neighbor information, including TCP MSS

```
PE1# show interface GigabitEthernet0/0/0
GigabitEthernet0/0/0 is up, line protocol is up
  Hardware is ASR1001, address is e8b7.48fb.7100 (bia e8b7.48fb.7100)
  Internet address is 192.168.106.60/24
  MTU 1500 bytes, BW 100000 Kbit/sec, DLY 100 usec,
     reliability 255/255, txload 1/255, rxload 1/255
  Encapsulation ARPA, loopback not set
  Keepalive not supported
  Full Duplex, 100Mbps, link type is auto, media type is T
  output flow-control is on, input flow-control is on
  ARP type: ARPA, ARP Timeout 04:00:00
  Last input 00:00:00, output 00:00:13, output hang never
  Last clearing of "show interface" counters never
  Input queue: 0/1000/0/0 (size/max/drops/flushes); Total output drops: 0
<… output omitted …>
```

- Displays interface information, including input queue depth.

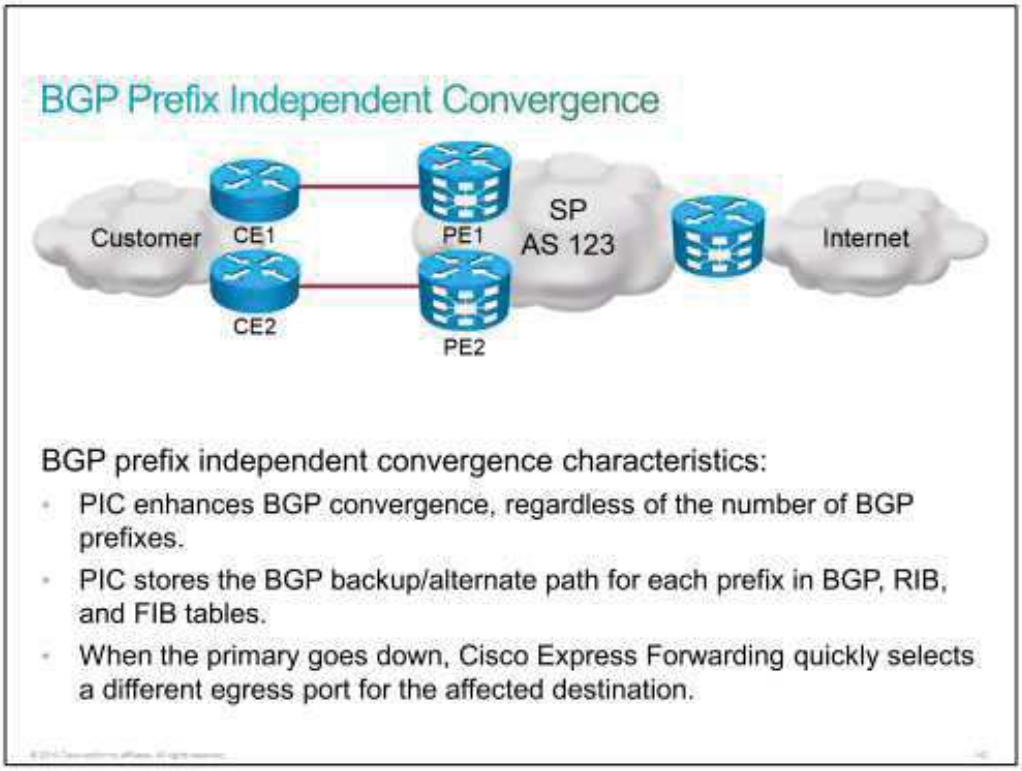
To verify the TCP MSS that is used between BGP neighbors, use the **show ip bgp neighbors** command (Cisco IOS and IOS XE Software only). You can use the **include** filter to display relevant output only, as shown in the example. In the example, the MSS for BGP sessions with three neighbors is 1460 bytes.

To verify interface queue depth, use the **show interface** command and search for the input queue section. In the example, input queue depth has been increased to 1000 packets.

https://t.me/learningnets

BGP Prefix Independent Convergence

This topic explains the BGP PIC feature.



BGP Process

Under normal circumstances, BGP can take several seconds to a few minutes to converge after a network change. At a high level, BGP goes through the following process:

1. BGP learns of failures through either IGP or BFD events or interface events.
2. BGP withdraws the routes from the RIB, and the RIB withdraws the routes from the FIB and distributed FIB. This process clears the data path for the affected prefixes.
3. BGP sends withdraw messages to its neighbors.
4. BGP calculates the next best path to the affected prefixes.
5. BGP inserts the next best path for affected prefixes into the RIB, and the RIB installs them in the FIB and distributed FIB.

This process takes a few seconds or a few minutes to complete, depending on the latency of the network, the convergence time across the network, and the local load on the devices. The data plane converges only after the control plane converges.

The BGP PIC functionality is achieved by additional functionality in the BGP, RIB, and Cisco Express Forwarding.

https://t.me/learningnets

BGP PIC affects prefixes under IPv4 and VPNv4 address families. For those prefixes, BGP calculates a second best path, along with the primary best path. (The second best path is called the backup/alternate path.) BGP installs the best path and the backup/alternate paths for the affected prefixes into the BGP RIB. The backup/alternate path provides a fast reroute mechanism to counter a single network failure.

For BGP PIC, RIB installs an alternate path per route if one is available. With the BGP PIC functionality, a RIB that selects a BGP route containing a backup/alternate path installs the backup/alternate path with the best path.

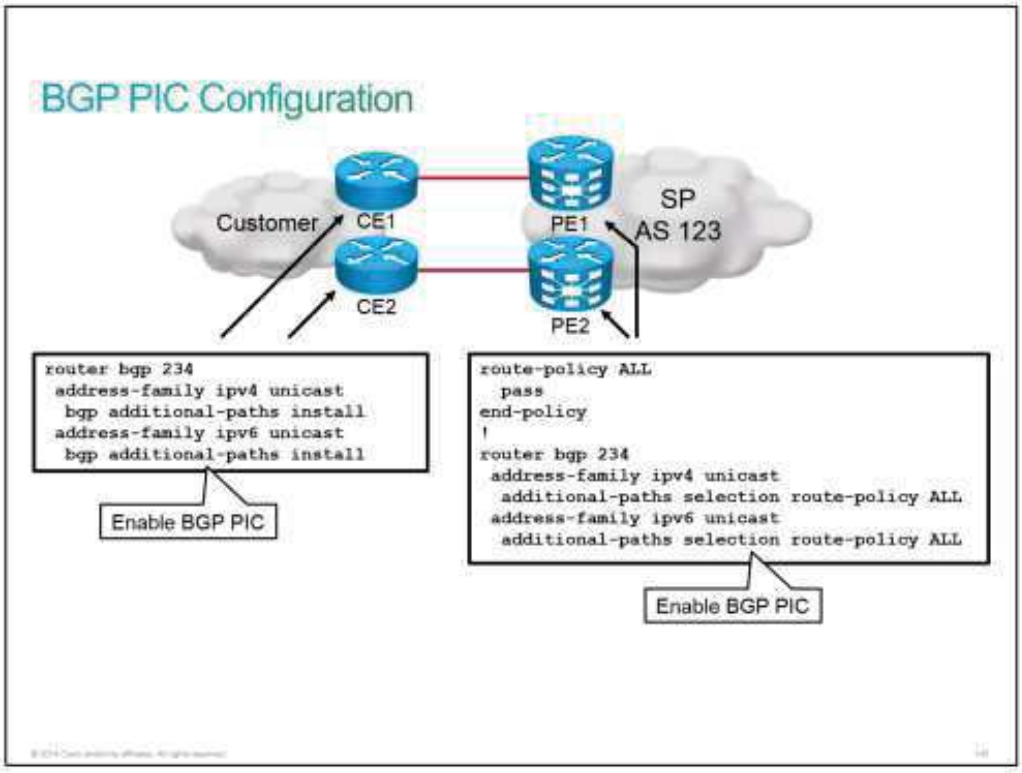
With BGP PIC, Cisco Express Forwarding stores an alternate path per prefix. When the primary path goes down, Cisco Express Forwarding searches for the backup/alternate path in a prefix-independent manner. Cisco Express Forwarding also listens to BFD events to rapidly detect local failures. Upon detection of a failure, Cisco Express Forwarding detects the alternate next hop for all prefixes affected by the failure.

In the example, CE1 and CE2 are configured with the BGP PIC feature. BGP computes PE1 as the best path and PE2 as the backup/alternate path, and installs both the routes into the RIB and Cisco Express Forwarding plane.

No policies should be set on CE1 and CE2 for the EBGP peers PE1 and PE2. Both CE routers must point to the EBGP route as the next hop. On CE1, the next hop to reach the Internet is through PE1. On CE2, the best path to reach the Internet is PE2. CE2 advertises itself as the next hop to CE1, and CE1 does the same to CE2. As a result, CE1 has two paths for the specific prefix and it usually selects the directly connected EBGP path over the IBGP path, according to the best path selection rules. Similarly, CE2 has two paths, an EBGP path through PE2 and an IBGP path through CE1-PE1.

If the CE1-PE1 link or PE1 goes down, BGP recomputes the best path, removing the next hop PE1 from RIB and reinstalling CE2 as the next hop into the RIB and Cisco Express Forwarding. CE1 automatically gets a backup/alternate repair path into Cisco Express Forwarding, and the traffic loss during forwarding is now in subseconds, thereby achieving fast convergence.

BGP PIC Configuration



https://t.me/learningnets

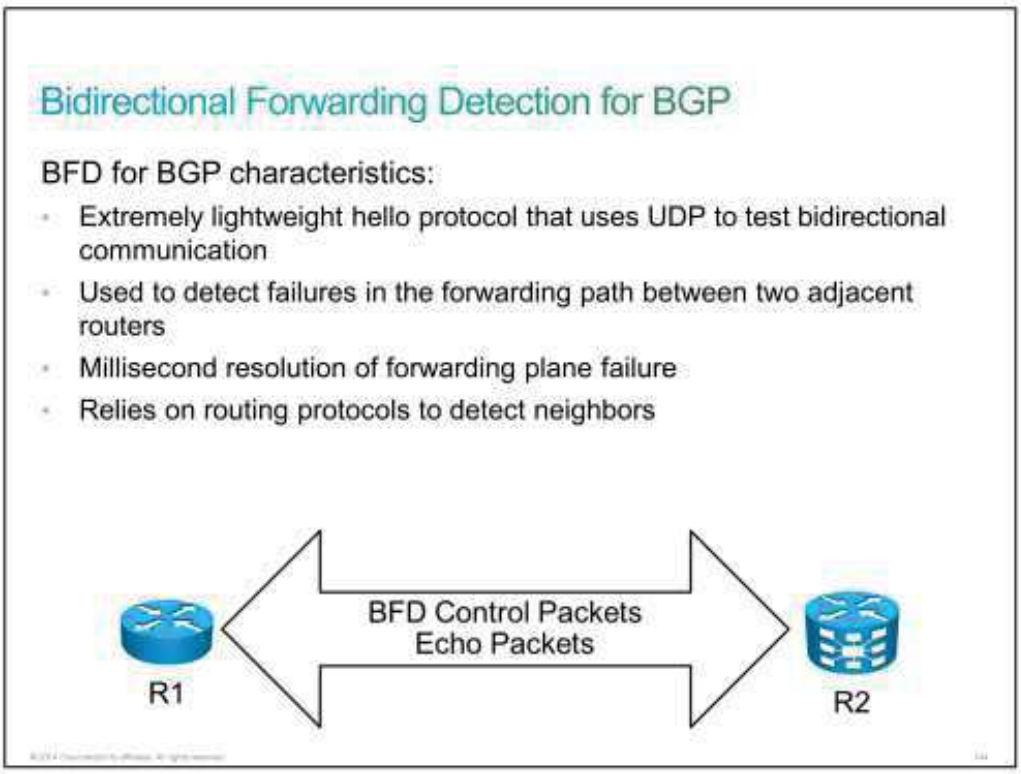
To enable BGP PIC on Cisco IOS and IOS XE Software routers, first enter BGP configuration mode using the **router bgp** command. Then enter the appropriate address family configuration mode and use the **bgp additional-paths install** command to enable BGP PIC. To enable BGP PIC on Cisco IOS XR Software routers, use the **additional-paths selection route-policy** command, which enables you to selectively enable PIC for all prefixes or for prefixes defined by a route policy. In the example, BGP PIC is enabled for all IPv4 and IPv6 prefixes.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Bidirectional Forwarding Detection for BGP

BFD provides a low-overhead, short-duration method of detecting failures in the forwarding path between two adjacent routers, including the interfaces, data links, and forwarding planes. BFD is a detection protocol that you enable at the interface and routing protocol levels. Cisco supports the BFD asynchronous mode, which depends on the sending of BFD control packets between two systems to activate and maintain BFD neighbor sessions between routers. Therefore, in order for a BFD session to be created, you must configure BFD on both BFD peers. Once BFD has been enabled on the interfaces and at the router level for the appropriate routing protocols, a BFD session is created, BFD timers are negotiated, and the BFD peers will begin to send BFD control packets to each other at the negotiated interval.



BFD provides fast BFD peer failure detection times independently of all media types, encapsulations, topologies, and routing protocols BGP, EIGRP, IS-IS, and OSPF. By sending rapid failure detection notices to the routing protocols in the local router to initiate the routing table recalculation process, BFD contributes to greatly reduced overall network convergence time.

The BFD protocol has no discovery mechanisms to detect neighbors; it is designed solely as an agent for other applications requiring fast failure detection. Whenever a routing protocol that is configured to use BFD detects a new neighbor, it requests availability tracking from BFD.

BFD can rely on control packets or on echo packets. Echo packets are IP packets addressed to the router itself but sent to the Layer 2 address of the next-hop node. The echo packets thoroughly test the complete bidirectional forwarding path between adjacent routers because they have to be transmitted by the originating router, propagated to the adjacent router, received by its interface logic, switched by its forwarding engine, and sent back to the originator (because the IP packet is addressed to the router itself).

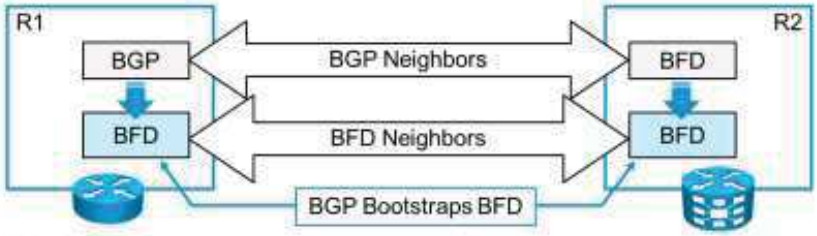
For example, when R1 sends a BFD echo packet, it sets the destination IP address in the packet to its own interface IP address and the MAC address in the Layer 2 frame header to the MAC address of the neighbor (R2). When R2 receives the packet, it performs a Layer 3 lookup and sends the packet toward its final destination (back to R1).

https://t.me/learningnets

Bidirectional Forwarding Detection for BGP (Cont.)

BFD operation:

- Routing protocol (BFD client) bootstraps BFD to create a BFD session to a neighbor:
  - BFD client receives link status change notification.
  - Receive and transmit intervals are negotiated and configurable.
- The two systems agree on a method to detect failure.
- In case of failure, BFD notifies the BFD client:
  - The BFD client independently decides on the action.



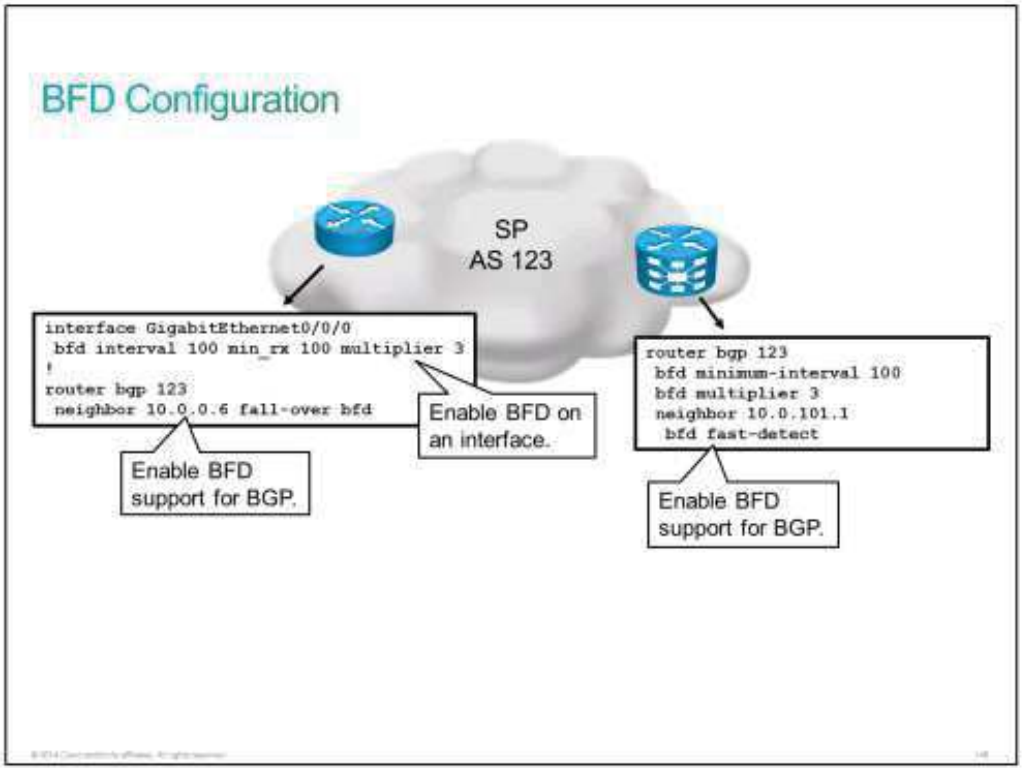
When a routing protocol (BGP, for example) discovers a neighbor, it sends a request to the local BFD process to initiate a BFD neighbor session with the BGP neighbor router. Then the BFD neighbor session with the BGP neighbor router is established. If there is a failure on the link between neighbors, the BFD neighbor session with the BGP neighbor router is torn down. BFD notifies the local BGP process that the BFD neighbor is no longer reachable. The local BGP process tears down the BGP neighbor relationship. If an alternative path is available, the routers will immediately start converging on it.

Note	If multiple routing protocols want to establish BFD sessions with the same remote system for the same routed protocol (IPv4 or IPv6), all must share a single BFD session.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

BFD Configuration



The figure shows a configuration of BGP BFD. To enable BFD on the Cisco IOS and IOS XE Software router, first enter interface configuration mode and enable BFD using the **bfd interval** *send-timer* **min_rx** *receive-timer* **multiplier** *multiplier* command. The *send-timer* argument specifies the frequency of BFD packets originated by the router, and the *receive-timer* argument specifies the minimum interval between packets accepted from BFD peers. (BFD adjacency will not form if the send timer on one peer is lower than the receive timer on another peer.) The *multiplier* number is the number of BFD packets that can be lost before the BFD peer is declared "down."

To enable BFD support for BGP, enter BGP routing mode and enable BFD for an individual neighbor using the **neighbor fall-over bfd** command.

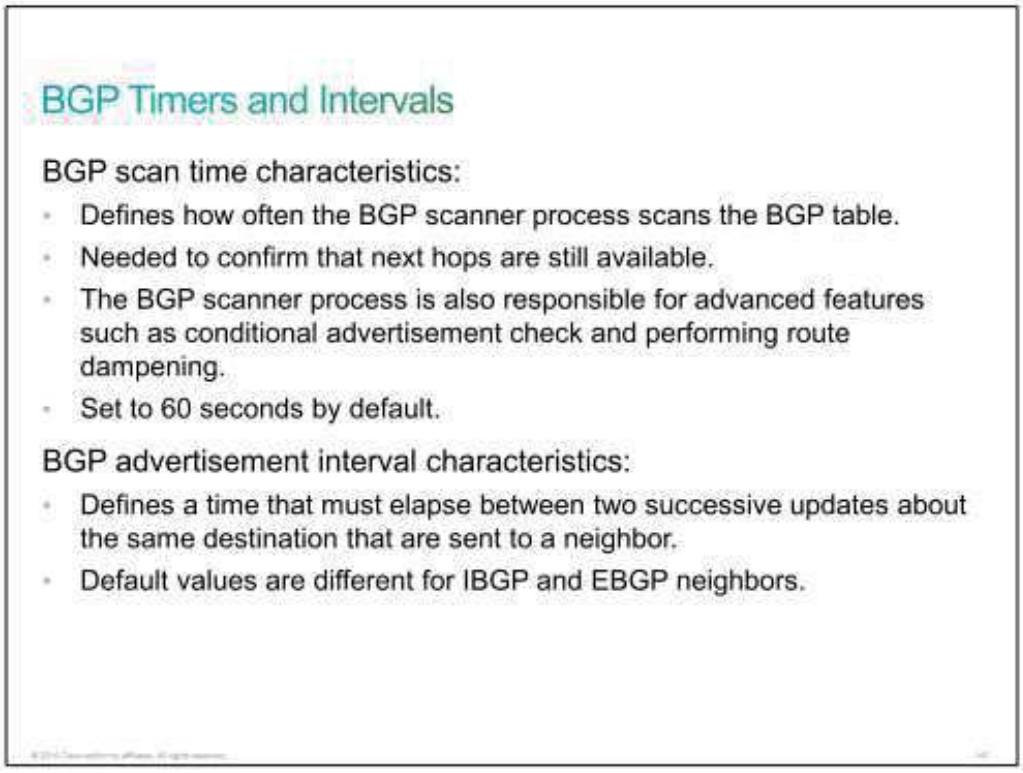
To enable distributed BFD support for BGP on the Cisco IOS XR Software router, first enter BGP configuration mode. Then specify the interval between BFD packets using **the bfd minimum-interval** *milliseconds* command. Then specify the multiplier using the **bfd multiplier** command, followed by a number. The multiplier specifies the number of times a BFD packet can be missed before BFD declares the neighbor down. Finally, enable BFD support for each neighbor individually by entering neighbor configuration mode and using the **bfd fast-detect** 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

BGP Timers and Intervals

This topic explains BGP timer and interval configuration.



BGP convergence can also be reduced by using certain BGP timers. Two BGP timers can be used to influence BGP convergence speed:

- **Scan time:** Defines how often the BGP scanner process scans the BGP table. The scanner process is responsible for verifying information in the BGP table. By default, the BGP scan time is set to 60 seconds.
- **Advertisement interval:** Controls the rate at which successive advertisements are sent to a BGP neighbor. The default values are different for IBGP and EBGP peers. For IBGP peers, the interval is set to 0 seconds. For EBGP peers, the interval is set to 30 seconds. For EBGP peers that are configured in a VRF instance, the interval is also set to 0 seconds.

https://t.me/learningnets

BGP Timers and Intervals (Cont.)

BGP keepalive timer characteristics:

- Defines a time between successive keepalive messages.
- Set to 60 seconds by default.

BGP hold-down timer characteristics:

- Defines how long a router will wait from the last received keepalive or update message before declaring the session dead.
- Set to 3x keepalive = 180 seconds by default.

BGP keepalive and hold-down timers are two BGP timers that also influence BGP convergence. The keepalive timer defines how often a BGP router will probe a BGP neighbor by sending keepalive messages. The hold-down timer defines how long a router will wait from the last received keepalive or update message before declaring the session dead.

By default, the keepalive timer is set to 60 seconds, and the hold-down timer is set to 180 seconds. Therefore even if the IP routing table indicates that the neighbor is no longer reachable, BGP will not trigger the convergence process until the BGP session hold-down timer expires.

BGP Timers and Intervals (Cont.)

Improving BGP Convergence

- BGP convergence can also be improved to some extent by:
  - Lowering the scan time interval for the BGP scanner process.
  - Lowering the advertisement interval between BGP neighbors.
  - Lowering the keepalive and hold-down timers.
- Limitations:
  - Not recommended in routers dealing with large BGP tables.
  - Could lead to CPU or memory exhaustion.
  - Lower hold-down timers could lead to undesired session terminations.

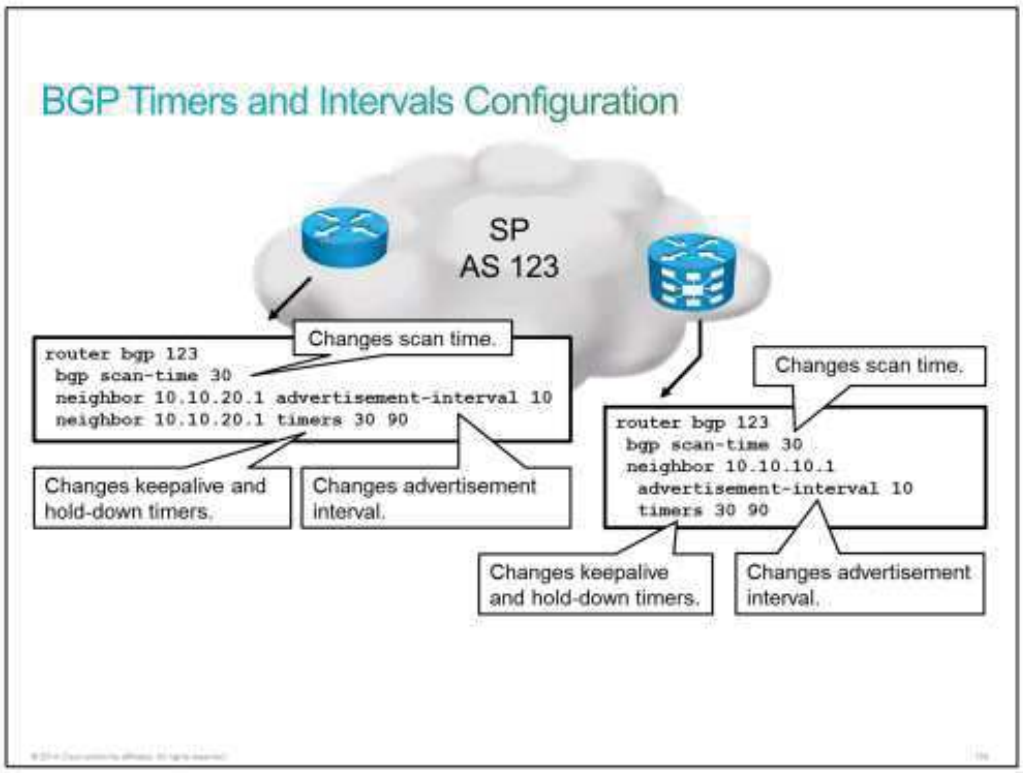
https://t.me/learningnets

BGP convergence can also be improved to some extent by:

- Configuring a smaller scan time interval for the BGP scanner process.
- Configuring a smaller advertisement interval between BGP neighbors.
- Configuring smaller keepalive and hold-down timers.

Network administrators must be careful when configuring these parameters. Setting the values too low for a specific network environment could lead to a significant consumption of router resources. The larger the BGP tables and the more unstable the BGP network, the greater the danger of exhausting the resources of a router. Lowering the keepalive and hold-down timers could lead to undesired session terminations when keepalives are not received due to network congestion.

BGP Timers and Intervals Configuration



The figure shows an example of how to change the BGP scan time and advertisement interval.

To change the BGP scan time on Cisco IOS and IOS XE Software platforms, first enter BGP configuration mode and use the **bgp scan-time** command, followed by the scan time in seconds. In the example, the scan time is set to 30 seconds. To change the neighbor advertisement interval, use the **neighbor advertisement-interval** command, followed by the advertisement interval in seconds. In the example, the advertisement interval is set to 10 seconds for neighbor 10.10.20.1. To change the keepalive and hold-down timers, use the **neighbor timers** command, followed by the keepalive and hold-down time in seconds. In the example, the keepalive and hold-down timers are set to 30 seconds and 90 seconds, respectively, for neighbor 10.10.20.1.

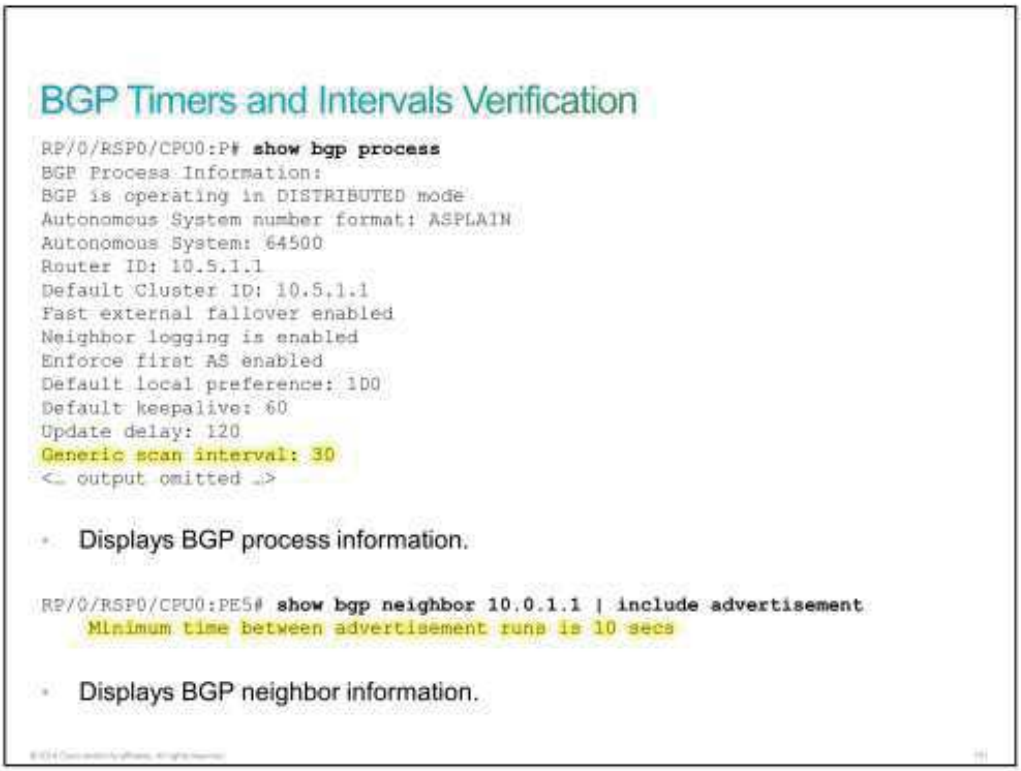
 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

To change the BGP scan time on Cisco IOS XR Software platforms, first enter BGP configuration mode and use the **bgp scan-time** command, followed by the scan time in seconds. In the example, the scan time is set to 30 seconds. To change the advertisement interval, first enter neighbor configuration mode and then use the **advertisement-interval** command, followed by the advertisement interval in seconds. In the example, the advertisement interval is set to 10 seconds for neighbor 10.10.10.1. To change the keepalive and hold-down timers, use the **timers** command, followed by the keepalive and hold-down time in seconds. In the example, the keepalive and hold-down timers are set to 30 seconds and 90 seconds, respectively, for neighbor 10.10.10.1.

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

BGP Timers and Intervals Verification



To verify the scan interval, use the **show bgp process** command.

To verify the neighbor advertisement interval, use the **show bgp neighbor** command. The command will show you the minimum time between advertisement runs, which is 10 seconds in the example.

https://t.me/learningnets

Summary

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

Summary

- BGP route dampening reduces the effects of route flaps by removing offending prefixes from the routing table for a period of time.
- To enable BGP route dampening, use the **bgp dampening** command under BGP router configuration mode.
- To verify BGP route dampening, use the **show bgp** and **show bgp dampened-paths** commands.
- BGP is considered converged when no outstanding updates are to be sent to neighbors.
- BGP comprises several processes of equal importance for the functioning of BGP as a complete protocol.
- BGP provides several features to improve convergence times.
- Distributed BGP splits processing of different address families to reduce impact in case of a failure.
- PMTU discovery benefits BGP because it benefits any transmission over IP.

Summary (Cont.)

- Input queue depth can be increased to improve performance.
- PMTUD, hold queue, and distributed BGP need to be configured for the BGP process.
- PIC enhances BGP convergence regardless of the number of BGP prefixes.
- BFD reduces conversion time by providing rapid detection of failed peers.
- BFD needs to be configured for directly connected eBGP peers or else its capabilities are available through the use of an IGP in conjunction with BFD.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Lesson 3: Improving BGP Configuration Scalability

Overview

Scaling routers to meet the demands of full Internet routing and associated administrative policies requires protocols such as BGP with embedded scalability mechanisms. In environments where network administrators must configure a large number of BGP peers, peer groups and configuration templates are scalability tools that reduce both administrative overhead and router resource requirements.

The lesson describes peer groups, peer templates, and configuration templates that are used to optimize BGP configuration.

Upon completing this lesson, you will be able to:

- Describe BGP peer groups and provide examples of BGP peer groups use
- Describe the role of BGP peer groups as a performance tool
- Describe the configuration of BGP peer groups
- Describe the limitations of BGP peer groups
- Describe BGP dynamic update groups
- Describe BGP peer templates
- Describe the three types of BGP configuration templates on Cisco IOS XR
- Describe BGP configuration templates inheritance
- Describe how to configure the BGP configuration templates
- Describe BGP configuration templates verification
- Describe the two types of BGP peer templates on Cisco IOS and IOS XE
- Describe BGP peer template inheritance
- Explain BGP peer template configuration

https://t.me/learningnets

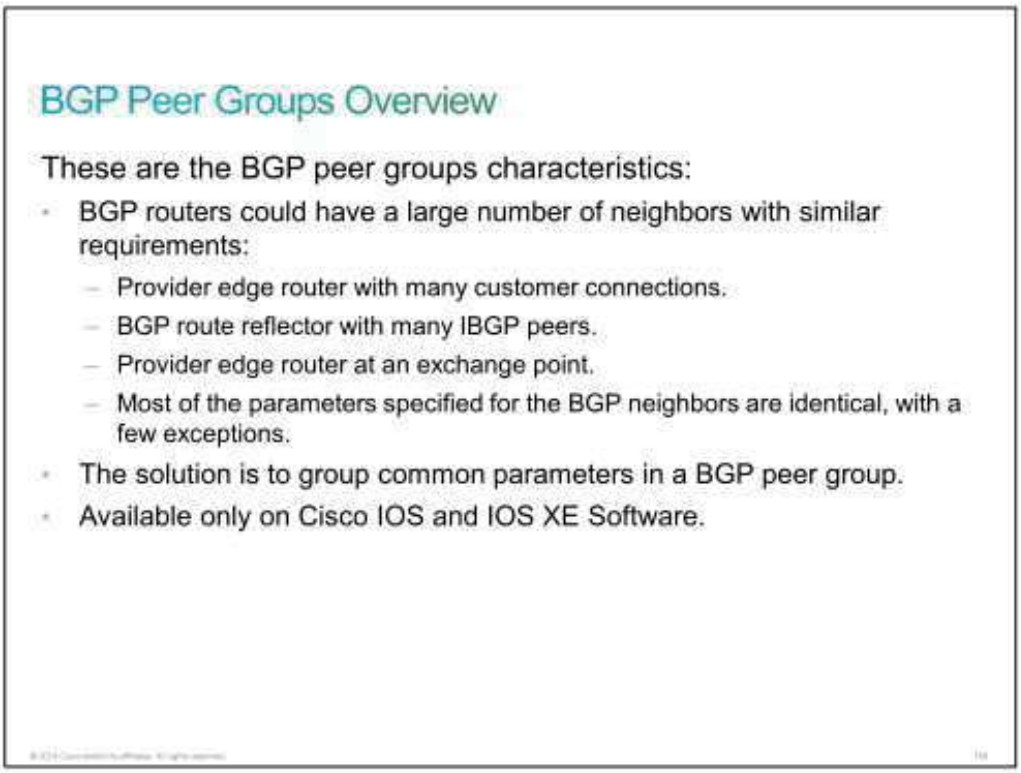
- Explain BGP peer template verification

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

BGP Peer Groups Overview

A network administrator often must configure a single router with a large number of neighbors, each neighbor having parameters similar to the others. This situation may cause time-consuming configuration work because the network administrator must configure almost the same parameters for all of the neighbors. If you have a service provider network that has an edge router with a large number of customers attached to it, where each customer requires its own BGP session, you may find that all of the BGP sessions to its customer routers have almost identical configurations.



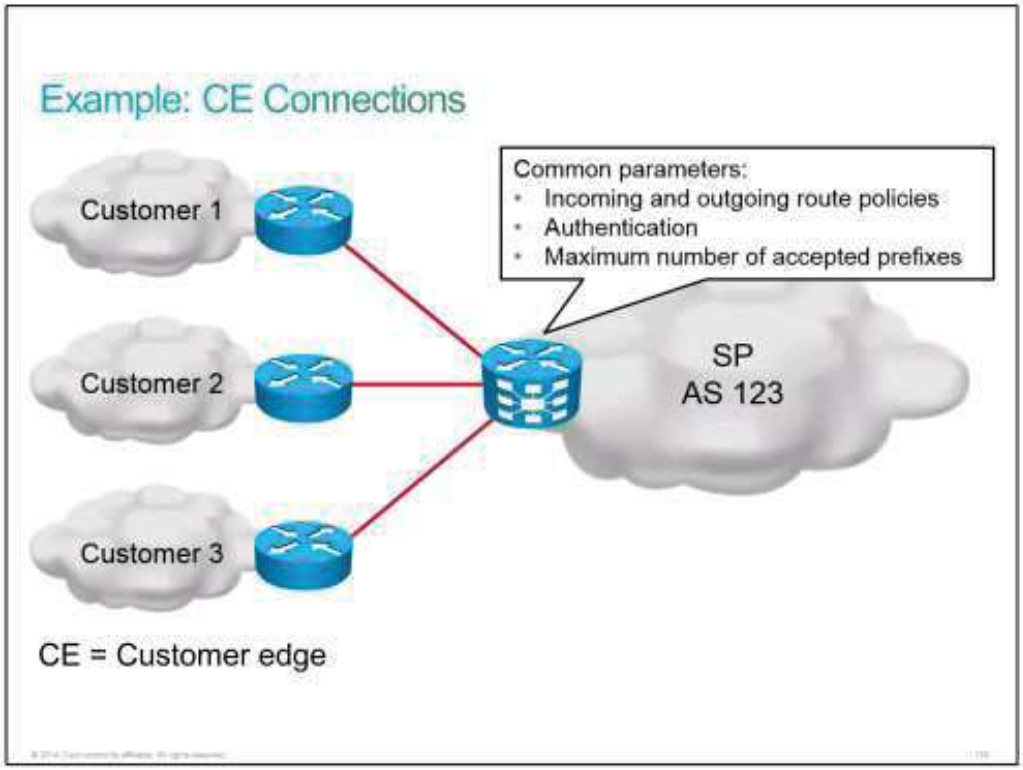
Likewise, IBGP sessions are almost always identically configured. If a full mesh is deployed within an AS, a large number of peer configurations might exist. Recall that an AS containing only 15 routers will require ([15 * 14] / 2) = 105 neighbor sessions to meet the full-mesh requirement of BGP. Configuring 105 neighbors with duplicate parameters leads to a large amount of redundant configuration.

Use the concept of peer groups to ease the burden of configuring a large number of neighbors with identical or similar parameters. The administrator configures the peer group with all the BGP parameters that are to be applied to many BGP peers. Actual BGP neighbors are bound to the peer group, and the network administrator applies the peer group configuration on each of the BGP sessions. BGP neighbors of a single router can be divided into several groups, each group having its own BGP parameters. Actual neighbors are then bound to the appropriate group, resulting in an optimal BGP configuration.

BGP peer groups are available only on the Cisco IOS and IOS XE Software platforms.

https://t.me/learningnets

Example: CE Connections



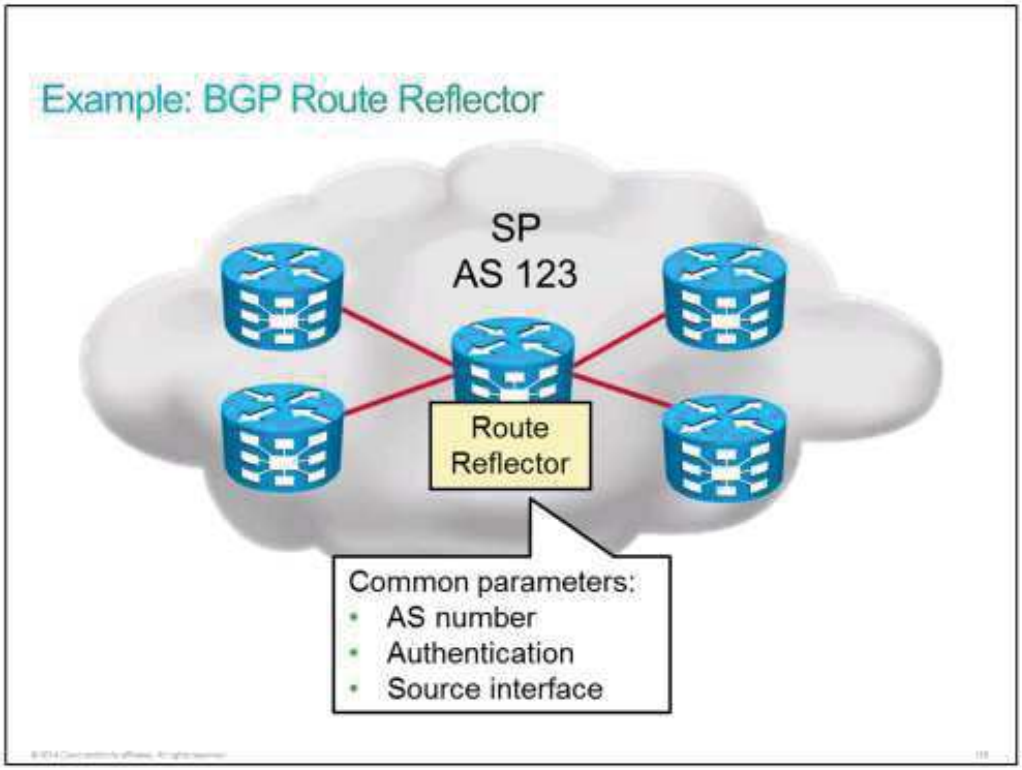
The figure shows an example where peer groups are useful. The customer autonomous systems are all assumed to announce local networks only. All customer autonomous systems should receive BGP updates with the same set of Internet routes, and the customer autonomous systems are all assumed to generate only a small number of prefixes.

This situation makes the neighbor configuration almost identical for each of the customers, with only a few changes that are specific to each neighbor.

In this scenario, the use of the peer group function is highly desirable. The network administrator can configure BGP neighbors in the customer autonomous systems using a single peer group. The administrator configures the peer group template with references to route policies, authentication settings, and the maximum number of received prefixes. Then the IP addresses of the customer routers are bound to the peer group, and the peer group configuration is applied to all neighbors.

https://t.me/learningnets

Example: BGP Route Reflector



Another example of using peer groups is within the entire local AS, on the route reflectors, where every IBGP session is configured identically. If a router in the AS is supplied with some information, then all the routers should be supplied with the same information. Otherwise, an inconsistent routing policy within the AS might cause inconsistent routing or application of BGP policies.

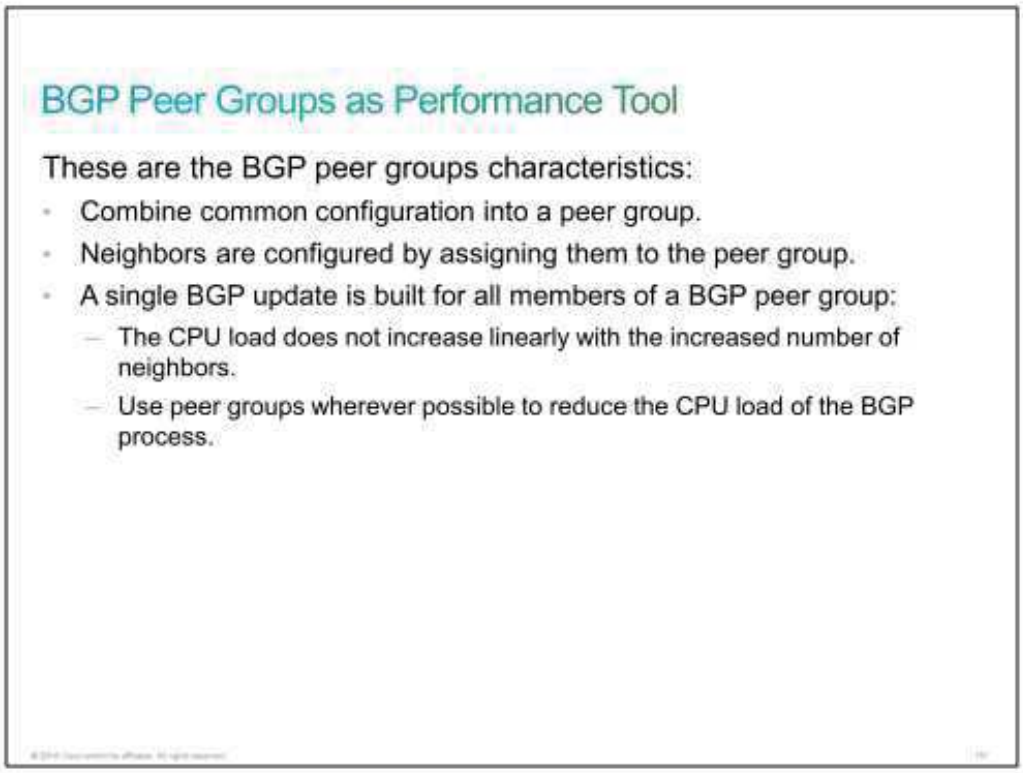
The peer group function is a good tool to make sure that all IBGP peers receive the same configuration information. The network administrator configures a peer group with the required parameters, such as the neighbor AS number, setting of the update source to a loopback interface, and router authentication mechanisms. Then all the internal neighbor IP addresses are bound to the peer group, and the peer group configuration is applied to all of them. This approach ensures a consistent routing policy within the AS.

In a service provider network, the routers that are assigned as route reflectors are the routers with the largest number of IBGP sessions. These are the routers where the peer group function is most useful.

https://t.me/learningnets

BGP Peer Groups as Performance Tool

As described here, peer groups can be used to combine common BGP configuration into a peer group. Neighbors are then configured by assigning them to the peer group. By default, the router builds BGP updates for each neighbor individually. Building BGP updates involves a number of router-CPU-consuming tasks, including scanning the BGP table and applying a variety of outgoing filtering mechanisms (filter lists, route maps, and prefix lists). These tasks mean that when a router is configured with a large number of neighbors, the CPU load grows proportionally.



However, with the use of peer groups, some of the router CPU utilization that is imposed by BGP update generation is significantly reduced because the use of peer groups allows the router to run the BGP update (including all outgoing filter processing) only once for the entire peer group. After the router has finished building the BGP update, it sends the update to each member of the peer group. The actual TCP transmission still must be done on a per-neighbor basis because of the connection-oriented characteristics of BGP sessions.

Router CPU load does increase when there are more neighbors of a router, due to increased TCP workload, but the use of peer groups can substantially reduce the increase. Therefore, network administrators should use peer groups whenever possible to reduce the CPU load.

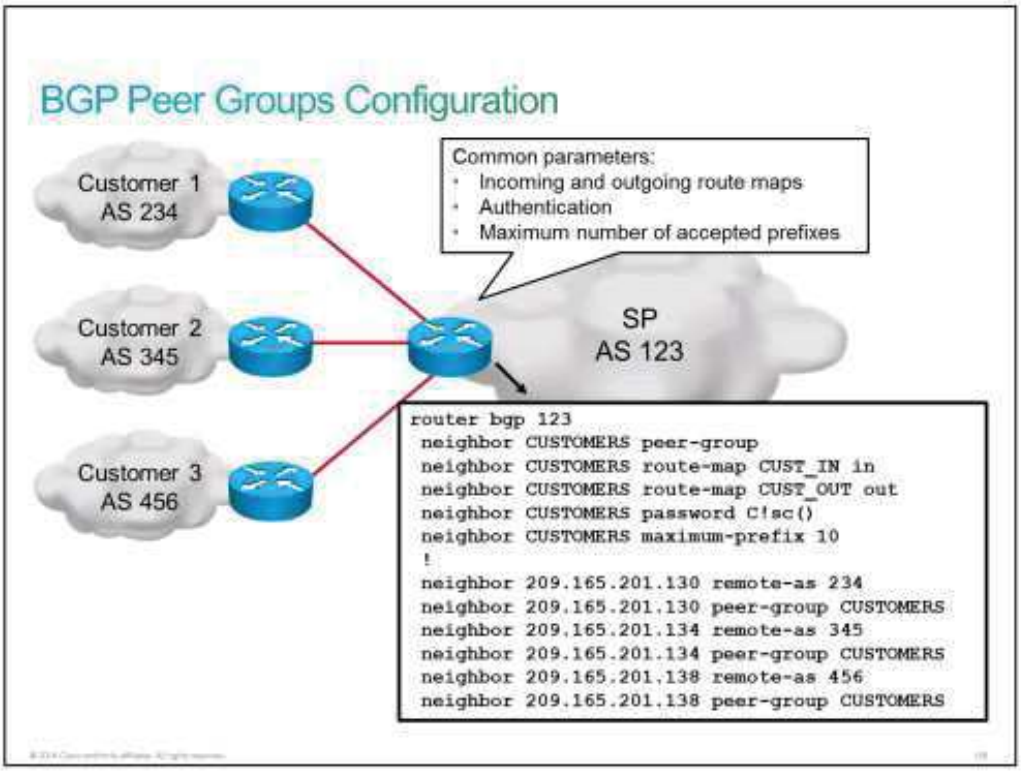
Note	Routers built BGP updates for each neighbor individually before the introduction of dynamic update groups, which were introduced in Cisco IOS Release 12.0(24)S and are also available on Cisco IOS XE and XR Software. Dynamic update groups, which are enabled by default, allow routers to dynamically calculate and optimize updates that are sent to neighbors that share the same outbound policies. Dynamic update groups will be discussed later. Although this new feature effectively made manually defined peer groups obsolete, many administrators still use peer groups because they offer convenience during configuration.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

BGP Peer Groups Configuration

This topic describes configuration of BGP peer groups.



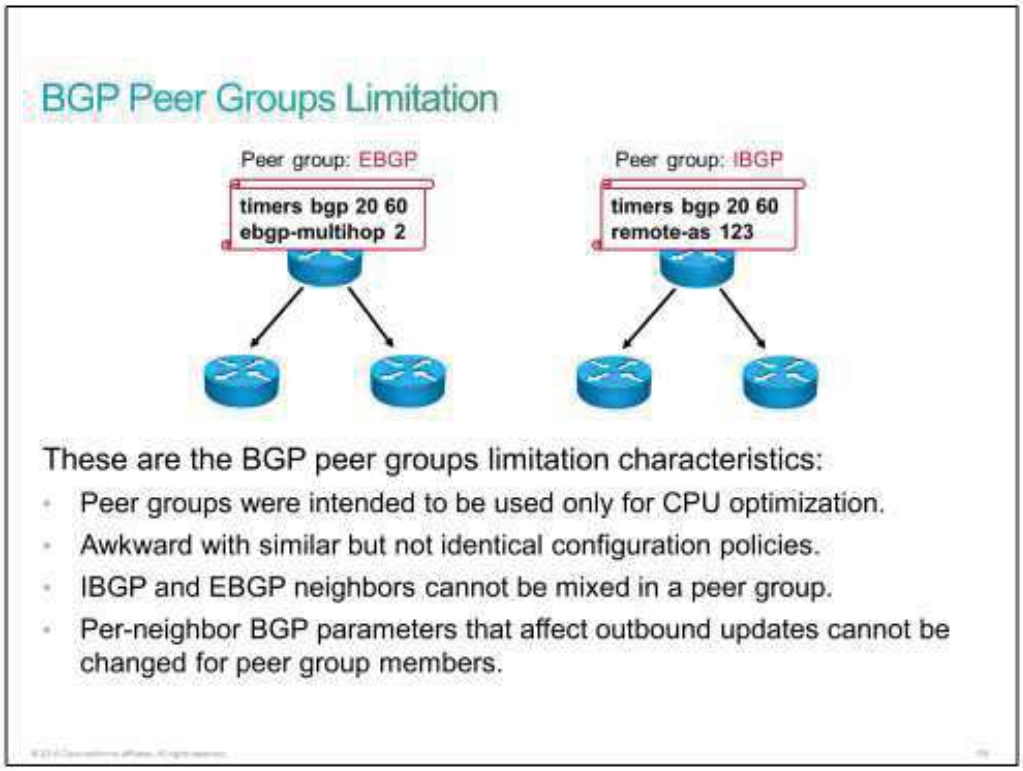
The figure shows a configuration of peer groups on the Cisco IOS and IOS-XE Software platforms. A peer group will be configured on the PE router and will combine incoming and outgoing route maps, authentication settings, and the maximum number of accepted prefixes. To create a peer group, first enter BPG router configuration mode and then use the **neighbor** *name* **peer-group** command. Then assign specific BGP parameters to the peer group using the same commands that are used to apply the parameters to neighbors. When all parameters are added to the peer group, configure BGP neighbors and assign them to the peer group, using the **neighbor peer-group** command, followed by the peer group name. In the example, three neighbors are configured, each in a different AS. Then all the neighbors are assigned to the peer group, which enables them to inherit common configuration from the peer group.

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

BGP Peer Groups Limitation

BGP peer groups were introduced primarily for CPU usage optimization, and therefore have some limitations.



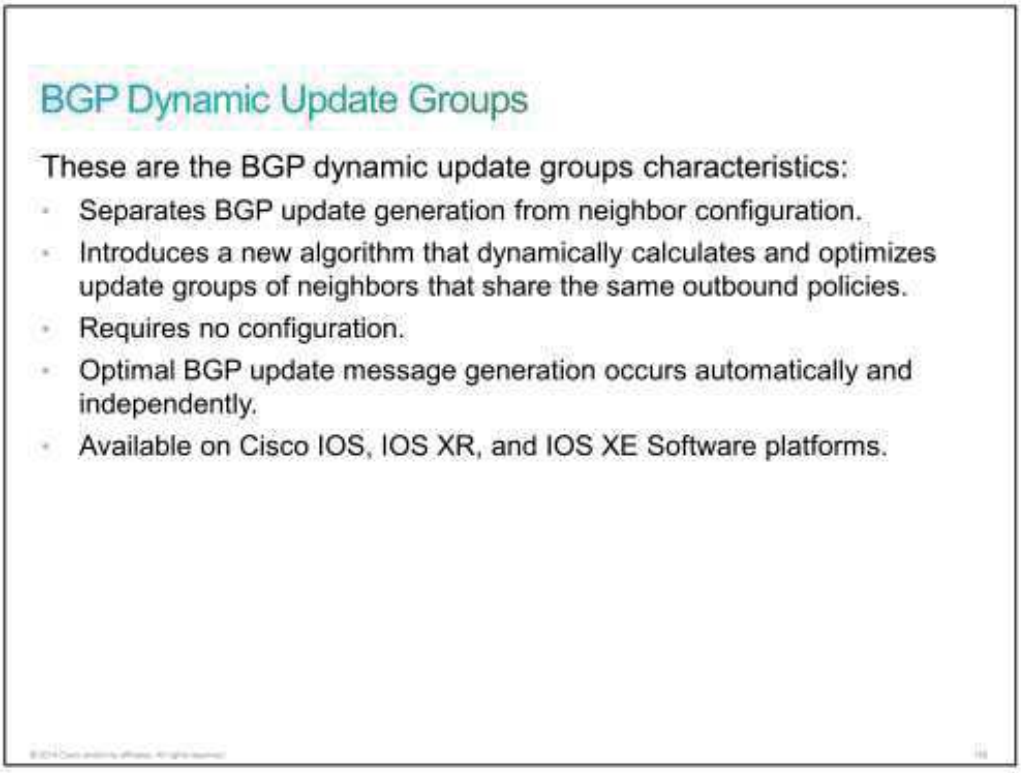
BGP parameters and outbound routing policies cannot differ between group members because doing so could cause different updates to be sent to two members of the same peer group. Therefore, peer groups are awkward when used on a router with several neighbors that have similar but not completely identical policies. For example, the same peer group cannot be used for EBGP and IBGP neighbors, even though a part of the configuration is the same for both types of neighbors. An example would be if you wanted to change the BGP keepalive and hold-down timers for both IBGP and EBGP sessions. Using peer groups, you must specify the timers in both peer groups (IBGP and EBGP),because they use different configurations due to the different nature of IBGP and EBGP sessions. Therefore common BGP parameters must be replicated across all peer groups, and there is no way to use only one peer group for common configuration.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

BGP Dynamic Update Groups

Cisco introduced dynamic update peer groups in the Cisco IOS 12.0(24)S Software release, a feature that enables the routers to automatically group BGP neighbors into groups and generate update messages accordingly. The BGP dynamic update groups feature contains an algorithm that dynamically calculates and optimizes update groups of neighbors that share outbound policies and can share the update messages. The BGP dynamic update groups feature separates update group replication from peer group configuration, improving convergence time and flexibility of neighbor configuration. The BGP update groups require no configuration, and a router optimizes BGP update message generation automatically.



When a change to the configuration occurs, the router automatically recalculates update group memberships and applies the changes by triggering an outbound soft reset after a 3-minute timer expires. This behavior is designed to provide the network operator with time to change the configuration if a mistake is made.

Note	A soft reset, which is performed on a per-neighbor basis, does not clear the BGP session and facilitates the application of new policies. There are two methods of performing a soft reset. Dynamic inbound soft reset is used to generate inbound updates from a neighbor. An outbound soft reset is used to send a new set of updates to a neighbor.

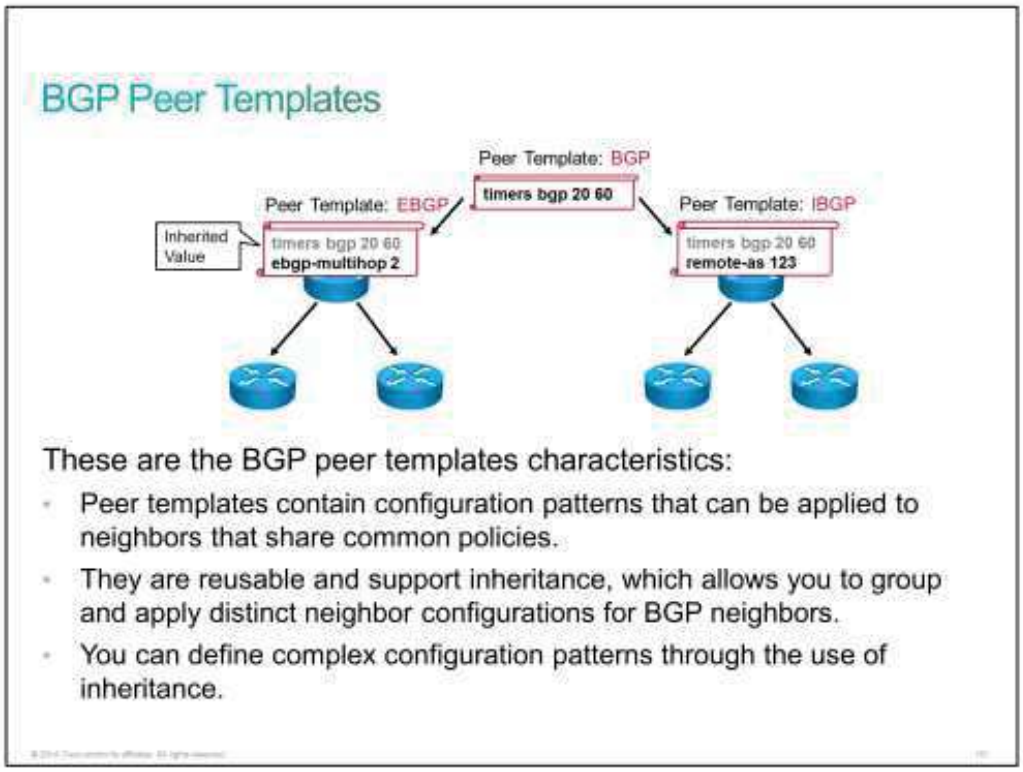
BGP dynamic update groups are supported on Cisco IOS, IOS XE, and IOS XR Software platforms.

With the burden of CPU optimization delegated to dynamic update peer groups instead of peer groups, Cisco introduced a mechanism called BGP peer templates that is used to optimize BGP neighbor configuration. This mechanism is discussed in the next topic.

https://t.me/learningnets

BGP Peer Templates Overview

The BGP dynamic update groups feature separates peer group configuration from update group generation. BGP neighbor configuration is no longer restricted by outbound routing policies, and update groups can belong to different address families.



Even though BGP update message generation has been separated from peer group configuration, peer group configuration still has the following limitations:

- A neighbor can belong only to one peer group.
- Neighbors that belong to different address families cannot belong to the same peer group.
- Different sets of policies cannot be grouped and applied to a neighbor.

To address the limitations of peer groups, the BGP peer templates feature was introduced, along with the BGP dynamic update groups feature. A peer template is a configuration pattern that can be applied to neighbors that share common policies. Peer templates are reusable and support inheritance, which allows the network operator to group and apply distinct neighbor configurations for BGP neighbors that share common policies. Peer templates also allow the network operator to define complex configuration patterns because a peer template can inherit a configuration from another peer template.

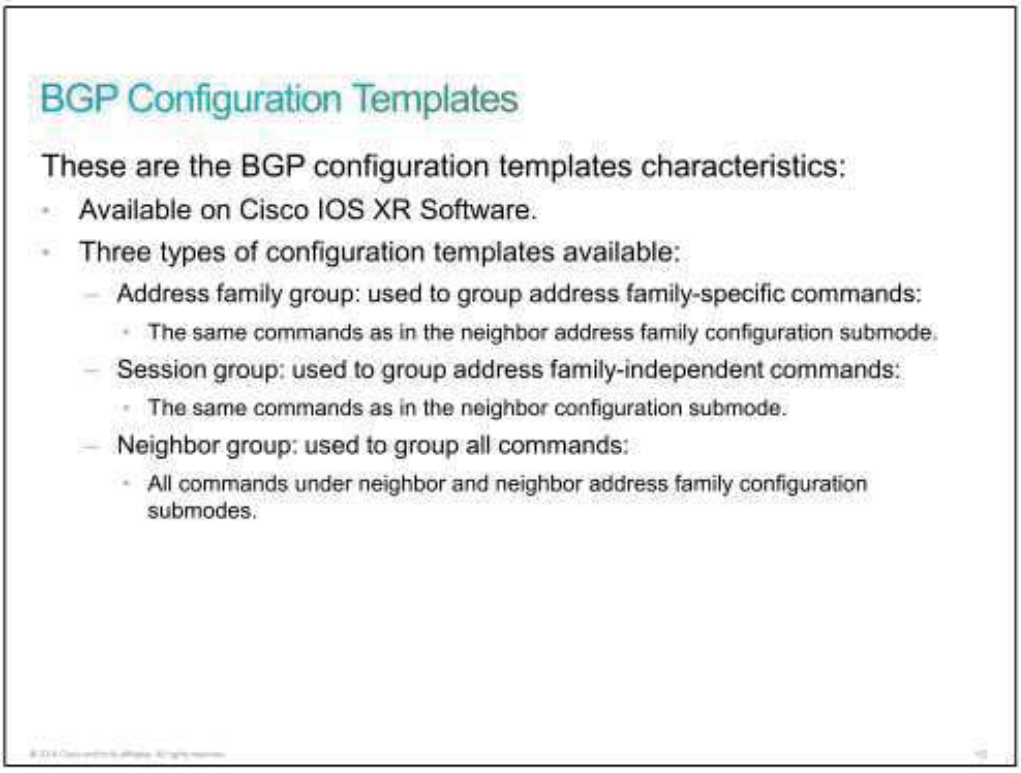
The example in the figure presents peer template inheritance. One peer template, called BGP, is configured and contains parameters that are common to all BGP neighbors. In the example, such parameters are the BGP keepalive and hold-down timers. Then one peer template is configured for IBGP and one for EBGP neighbors. Both templates inherit settings from the common peer template that is called BGP and add other parameters that are common only to IBGP or EBGP settings, respectively. Therefore templates can be created more effectively with fewer configurations when compared to BGP peer groups.

BGP peer templates are used on Cisco IOS and IOS XE Software platforms. On the Cisco IOS XR Software platforms, a similar feature called BGP configuration templates is available.

https://t.me/learningnets

BGP Configuration Templates

Three types of BGP configuration templates on Cisco IOS XR are available: address family group, session group, and neighbor group.



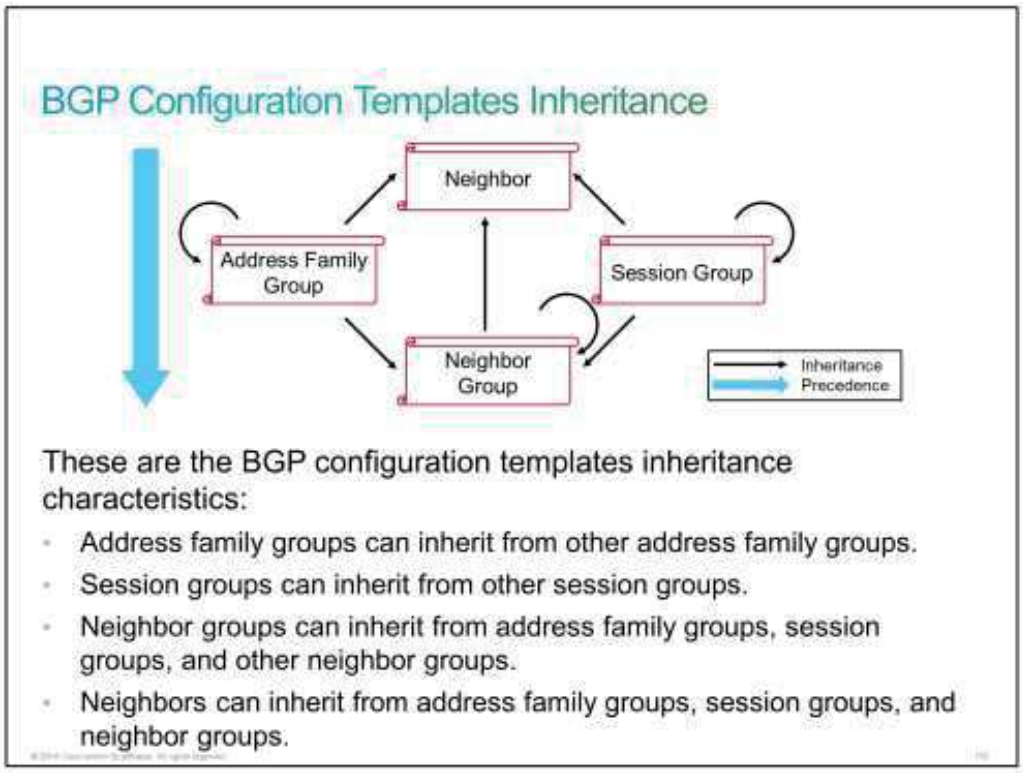
BGP configuration templates are supported on Cisco IOS XR Software platforms and support three different types of templates:

- **Address family group:** Allows you to group address family-specific neighbor commands within an IPv4, IPv6, or VPNv4 address family. Neighbors that have the same address family configuration are able to use the address family group for their address family-specific configuration. A neighbor inherits the configuration from an address family group. If a neighbor is configured to use an address family group, the neighbor (by default) inherits the entire configuration from the address family group. However, a neighbor does not inherit the entire configuration from the address family group if items are explicitly configured for the neighbor. The address family group configuration is entered under the BGP router configuration mode.
- **Session group:** Allows you to create a session group from which neighbors can inherit address family-independent configuration. A neighbor inherits the configuration from a session group. If a neighbor is configured to use a session group, the neighbor (by default) inherits the entire configuration of the session group. A neighbor does not inherit the entire configuration from a session group if a configuration is done directly on that neighbor.
- **Neighbor group:** Allows you to apply the same configuration to one or more neighbors. Neighbor groups can include session groups and address family groups, and can comprise the complete configuration for a neighbor. After a neighbor group is configured, a neighbor can inherit the configuration from the neighbor group. A neighbor that is configured to use a neighbor group inherits its entire BGP configuration.

https://t.me/learningnets

BGP Configuration Templates Inheritance

This topic explains BGP configuration templates inheritance.



The figure presents inheritance between different template groups, and precedence of groups if one feature is configured in several groups (and a neighbor inherits configuration from a group) and directly on a neighbor.

For address family-*independent* configuration, the following applies:

- Neighbors can inherit from session groups and neighbor groups.
- Neighbor groups can inherit from session groups and other neighbor groups.
- Session groups can inherit from other session groups.

For address family-*dependent* configurations, the following applies:

- Address family groups can inherit from other address family groups.
- Neighbor groups can inherit from address family groups and other neighbor groups.
- Neighbors can inherit from address family groups and neighbor groups.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

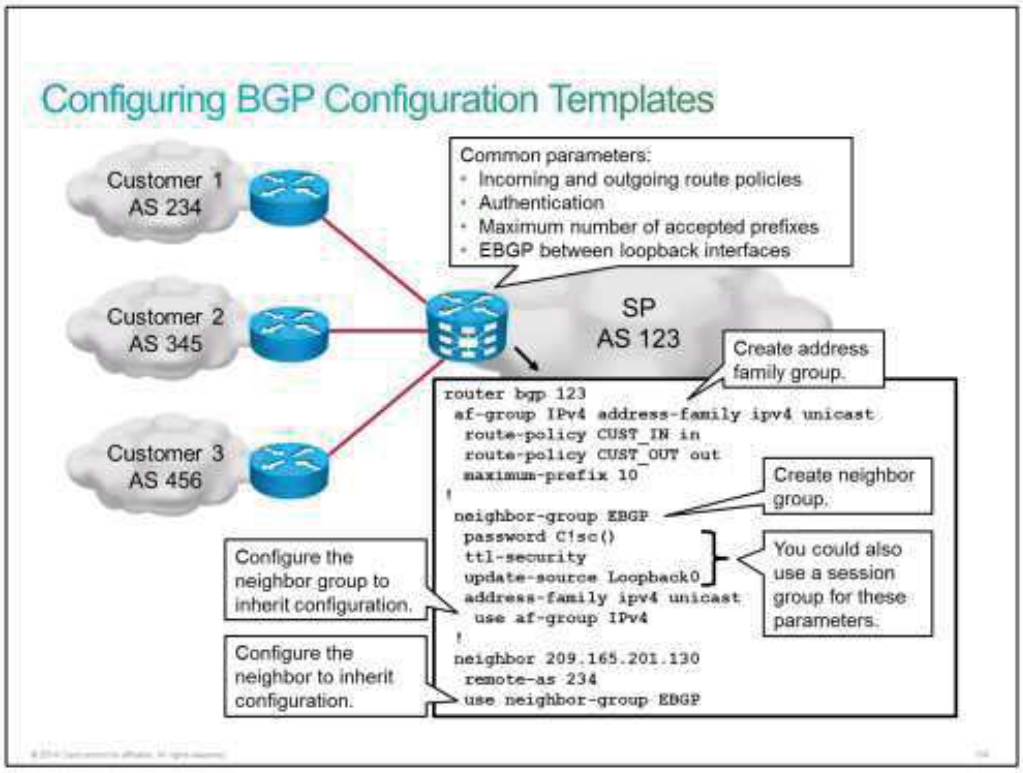
Configuration group inheritance rules are numbered in order of precedence, as follows:

1. If the item is configured directly on the neighbor, that value is used.
2. Otherwise, if an item is configured to be inherited from a session group or neighbor group and on the neighbor directly, then the configuration on the neighbor is used. If a neighbor is configured to be inherited from a session group or address family group, but there is no directly configured value, the value in the session group or address family group is used.
3. If a neighbor uses a session group and a neighbor group, the configurations in the session group are preferred over the configurations in the neighbor group. The same applies if a neighbor uses an address family group and a neighbor group.
4. Otherwise, if the neighbor uses a neighbor group and does not use a session group or an address family group, the configuration value can be obtained from the neighbor group either directly or through inheritance.
5. Otherwise, the default value is used.

https://t.me/learningnets

Configuring BGP Configuration Templates

The PE router will be configured with configuration templates for EBGP peers. Common parameters are route policies, authentication settings, maximum number of accepted prefixes, and EBGP session between loopback interfaces. Parameters that are applied to a specific address family will be grouped in the address family group. Other parameters will be grouped in the neighbor group.



The figure shows an example of configuration templates. To create an address family group, first enter BGP router configuration mode. Then create an address family group using the **af-group** command, followed by a group name and address family for which the group is being configured. Then specify BGP parameters using the same commands as are used for direct configuration on a neighbor. In the example, the group called IPv4 is created, which combines routing policies and the maximum number of accepted prefixes.

To create a neighbor group, first enter BGP router configuration mode. Then create a neighbor group using the **neighbor-group** command, followed by a group name. Then specify BGP parameters using the same commands as are used for direct configuration on a neighbor. To enable the neighbor group to inherit the settings from the address family group, enter the proper neighbor address family configuration mode, and then refer to the address family group using the **use af-group** command. In the example, the neighbor group that is called EBGP is created, which combines authentication settings and TTL security, and sets the source interface for BGP updates. The neighbor group also refers to the address family group and thus inherits all settings from the address family group. An alternative approach would be to create a session group and to combine address family-independent configuration inside the session group instead of the neighbor group. In such a case, the neighbor group would have to inherit settings from both the session group and the address family group.

To configure the neighbor to inherit configuration from the neighbor group, enter neighbor configuration mode and use the **use neighbor-group** command, followed by the neighbor group name. The neighbor could also inherit configuration directly from the address family or session group through use of the **use af-group** or **use session-group** commands, respectively.

 ©2014 Cisco Systems, Inc.

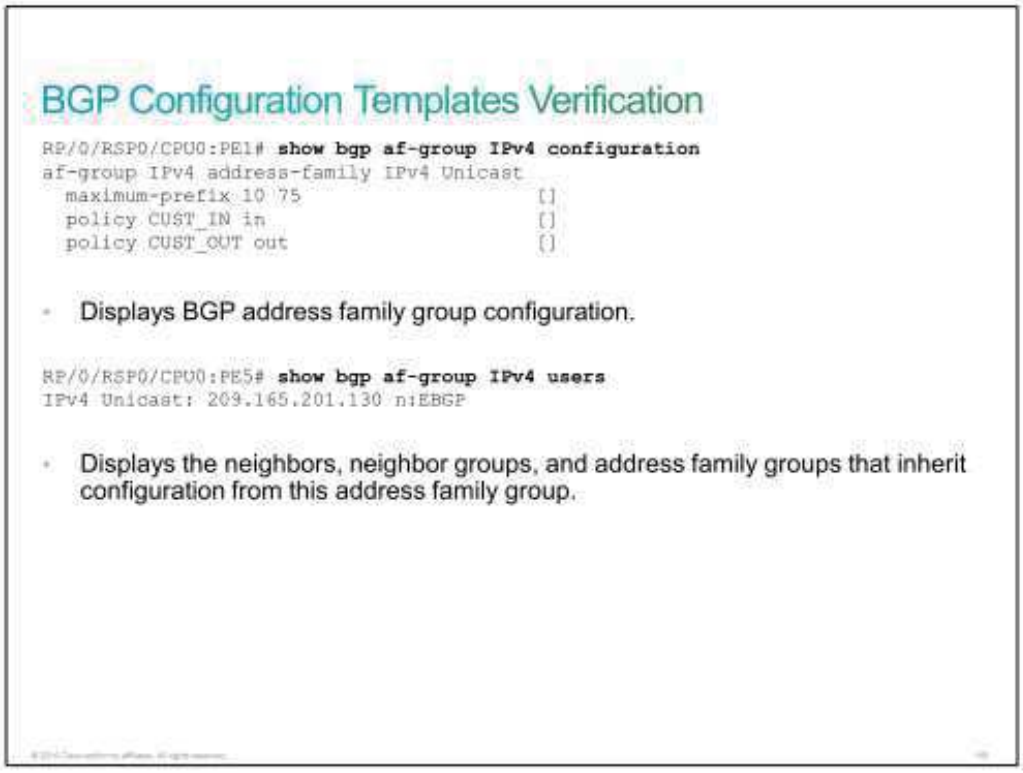
https://t.me/learningnets

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

https://t.me/learningnets

BGP Configuration Templates Verification

This topic explains how to verify the BGP configuration templates.



To display information about configuration for address family groups, use the **show bgp af-group** command. The example shows configuration for the address family group that is named IPv4. Parameters that are configured inside the group and not inherited are marked with empty brackets.

To display the neighbors, neighbor groups, and address family groups that inherit configuration from an address family group, use the **show bgp af-group users** command. In the example, the neighbor group that is called EBGP inherits configuration from the *IPv4* address family group.

https://t.me/learningnets

BGP Configuration Templates Verification (Cont.)

```
RP/0/RSP0/CPU0:PE1# show bgp neighbor-group EBGP configuration
neighbor-group EBGP
 password encrypted 143453180F4C63 []
 update-source Loopback0           []
 ttl-security                      []
 address-family IPv4 Unicast       []
  maximum-prefix 10 75             [a:IPv4]
  policy CUST_IN in                [a:IPv4]
  policy CUST_OUT out              [a:IPv4]
```

- Displays BGP neighbor group configuration.

```
RP/0/RSP0/CPU0:PE1# show bgp neighbor-group EBGP users
Session:      209.165.201.130
IPv4 Unicast: 209.165.201.130
```

- Displays the neighbors and neighbor groups that inherit configuration from this neighbor group.

To display information about configuration for neighbor groups, use the **show bgp neighbor-group** command. The example shows the configuration for the neighbor group that is called EBGP. You can also see that address family-specific parameters are inherited from the IPv4 address family group (as indicated inside the brackets).

To display the neighbors and neighbor groups that inherit configuration from a neighbor group, use the **show bgp neighbor-group users** command. In the example, BGP neighbor 209.165.201.130 inherits session and address family configuration from the neighbor group called EBGP.

BGP Configuration Templates Verification (Cont.)

```
RP/0/RSP0/CPU0:PE1# show bgp neighbors 209.165.201.130 configuration
neighbor 209.165.201.130
 remote-as 234                     []
 password encrypted 143453180F4C63 [n:EBGP]
 update-source Loopback0           [n:EBGP]
 ttl-security                      [n:EBGP]
 address-family IPv4 Unicast       [n:EBGP]
  maximum-prefix 10 75             [n:EBGP a:IPv4]
  policy CUST_IN in                [n:EBGP a:IPv4]
  policy CUST_OUT out              [n:EBGP a:IPv4]
```

- Displays the effective configuration for the neighbor, including any settings that have been inherited from groups.

https://t.me/learningnets

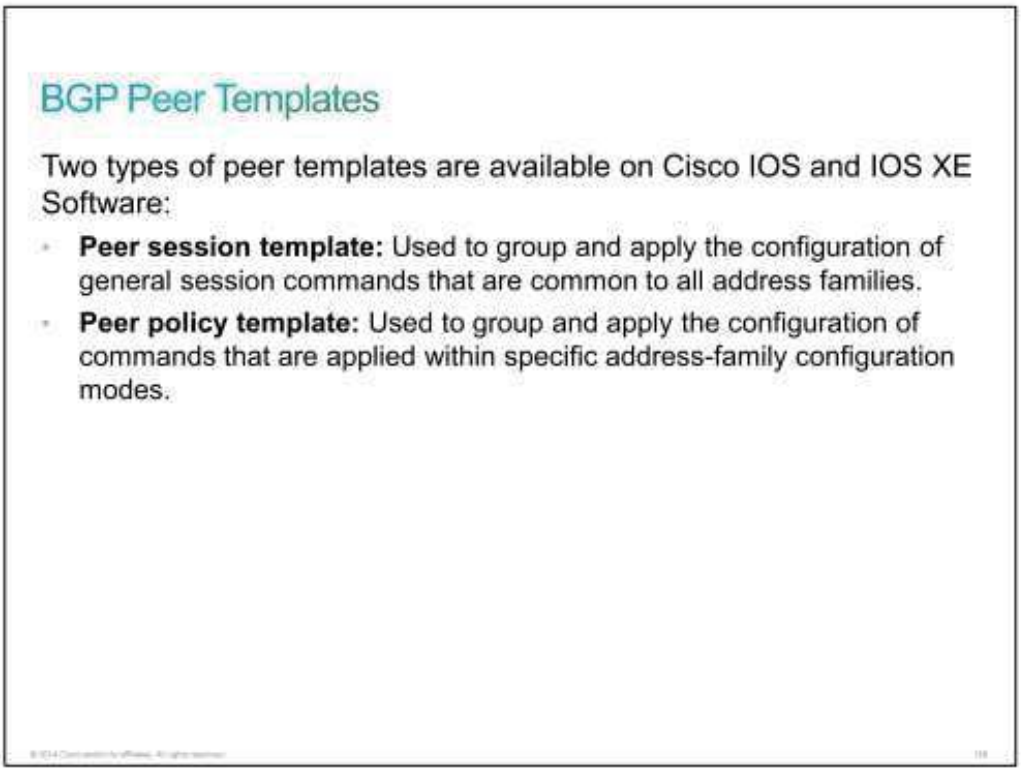
To display the effective configuration for a neighbor, including any settings that have been inherited from session groups, neighbor groups, or address family groups, use the **show bgp neighbors configuration** command. Parameters that are configured directly on the neighbor are marked with empty brackets. Parameters that are inherited from a group are marked with the group type and name inside the brackets. For example, the neighbor inherits session-specific parameters from the neighbor group that is called EBGP (marked with "n"). The neighbor also inherits address family-specific configuration from the same neighbor group, which in turn inherits configuration from the address-family group (marked with "a").

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

BGP Peer Templates

To address some of the limitations of peer groups, such as configuration management, BGP peer templates were introduced on Cisco IOS and IOS XE Software. A peer template is a configuration pattern that can be applied to neighbors that share policies.



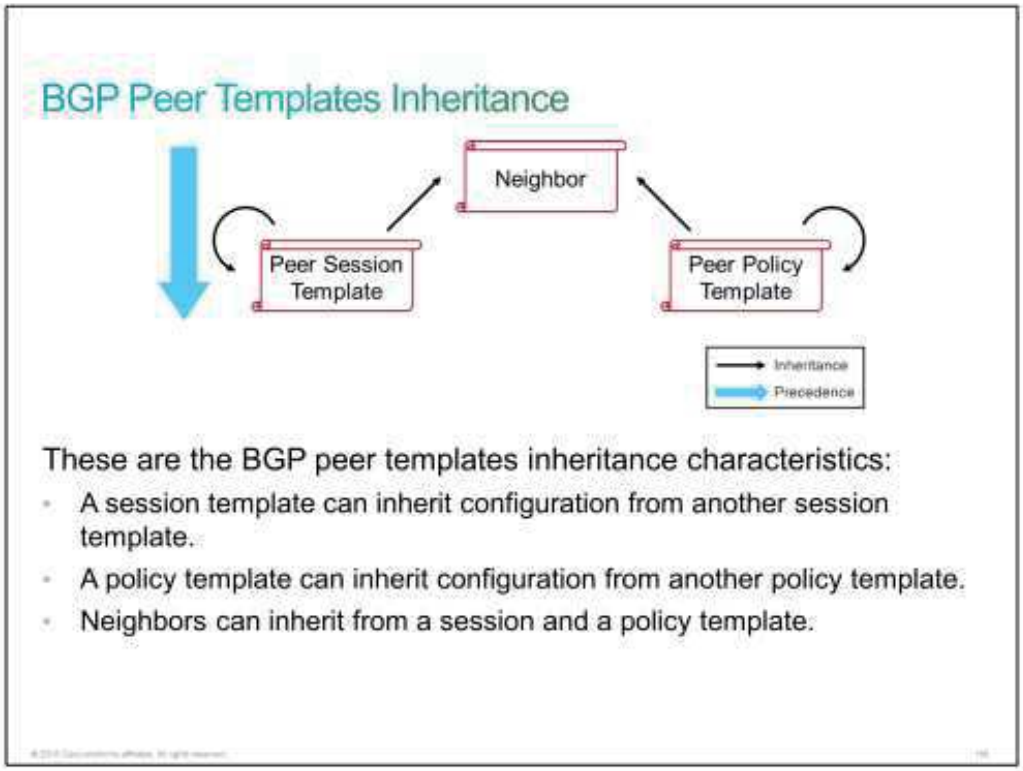
Peer templates improve the flexibility and enhance the capability of neighbor configuration. Peer templates also provide an alternative to peer group configuration and overcome some limitations of peer groups. BGP peer routers using peer templates also benefit from automatic update group configuration. With the configuration of the BGP peer templates and the support of the BGP dynamic update peer groups, the network operator no longer needs to configure peer groups in BGP, and the network benefits from improved configuration flexibility and faster convergence.

Note	The configuration of BGP peer templates does not conflict with or restrict peer group configuration, and peer groups are still supported in Cisco IOS releases that support BGP peer templates. However, a BGP neighbor cannot be configured to work with both peer groups and peer templates. A BGP neighbor can be configured to belong only to a peer group or to inherit policies from peer templates.

https://t.me/learningnets

BGP Peer Templates Inheritance

General session commands can be configured once in a peer session template and then applied to many neighbors through the direct application of a peer session template or through indirect inheritance from a peer session template. The configuration of peer session templates simplifies the configuration of general session commands that are commonly applied to all neighbors within an AS.



Peer session templates support direct and indirect inheritance. A peer can be configured with only one peer session template at a time, and that peer session template can contain only one indirectly inherited peer session template.

Note	If you attempt to configure more than one inherit statement with a single peer session template, an error message will be displayed.

This behavior allows a BGP neighbor to directly inherit only one session template and indirectly inherit up to seven additional peer session templates. This practice allows you to apply a maximum of eight peer session configurations to a neighbor: the configuration from the directly inherited peer session template, and the configurations from up to seven indirectly inherited peer session templates. Inherited peer session configurations are evaluated first and applied starting with the last node in the branch and ending with the directly applied peer session template configuration at the source of the tree. The directly applied peer session template will have priority over inherited peer session template configurations. Any configuration statements that are duplicated in inherited peer session templates will be overwritten by the directly applied peer session template, which means that if a general session command is reapplied with a different value, the subsequent value will have priority and overwrite the previous value that was configured in the indirectly inherited template.

https://t.me/learningnets

Peer policy templates are used to configure BGP policy commands that are configured for neighbors that belong to specific address families. Like peer session templates, peer policy templates are configured once and then applied to many neighbors through the direct application of a peer policy template or through inheritance from peer policy templates. The configuration of peer policy templates simplifies the configuration of BGP policy commands that are applied to all neighbors within an AS.

Like peer session templates, a peer policy template supports inheritance. However, there are minor differences. A directly applied peer policy template can directly or indirectly inherit configurations from up to seven peer policy templates. That is, a total of eight peer policy templates can be applied to a neighbor or neighbor group. Inherited peer policy templates are configured with sequence numbers like route maps. An inherited peer policy template, like a route map, is evaluated starting with the inherit statement with the lowest sequence number and ending with the highest sequence number. However, the difference is that a peer policy template will not collapse like a route map. Every sequence is evaluated, and if a BGP policy command is reapplied with a different value, any previous value from a lower sequence number will be overwritten.

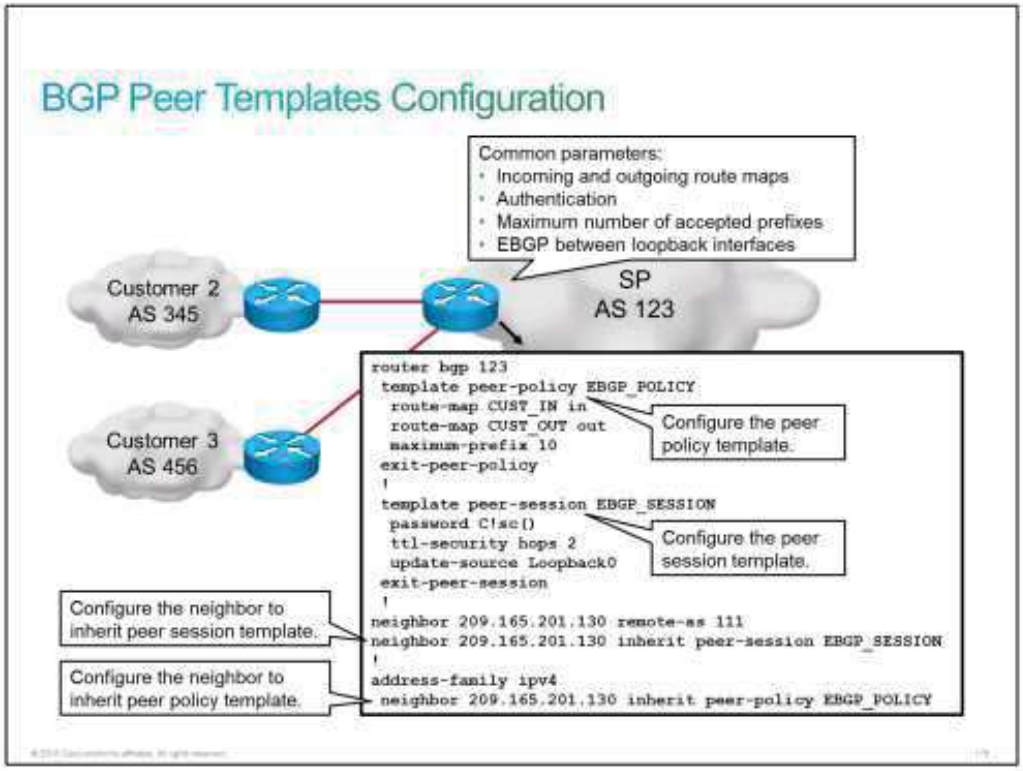
The directly applied peer policy template and the inherit statement with the highest sequence number will always have priority and be applied last. Commands that are reapplied in subsequent peer templates will always overwrite the previous values. This behavior is designed to allow you to apply common policy configurations to large neighbor groups and specific policy configurations only to certain neighbors and neighbor groups, without duplicating individual policy configuration commands.

The configuration of peer policy templates simplifies and improves the flexibility of BGP configuration. A specific policy can be configured once and referenced many times. Because a peer policy supports up to eight levels of inheritance, specific and complex BGP policies can also be created.

https://t.me/learningnets

BGP Peer Templates Configuration

The PE router will be configured with peer templates for EBGP peers. Common parameters are route maps, authentication settings, the maximum number of accepted prefixes, and EBGP session between loopback interfaces. Parameters that are applied to a specific address family will be grouped in a policy template, and other parameters will be grouped in a session template.



The figure shows an example of how to configure peer templates. To create a policy template, first enter BGP router configuration mode. Then use the **template peer-policy** command, followed by the policy template name. Then specify BGP parameters using the same commands as are used for direct configuration on a neighbor. In the example, the policy template that is called EBGP_POLICY is created, which combines routing policies and the maximum number of accepted prefixes.

To create a session template, use the **template peer-session** command, followed by the session template name. Then specify BGP parameters using the same commands as are used for direct configuration on a neighbor. In the example, the session template that is called EBGP_SESSION is created, which combines authentication settings, TTL security, and the EBGP session between loopback interfaces.

To apply a session template to a BGP neighbor, use the **neighbor inherit peer-session** command. To apply a policy template to a BGP neighbor, first enter the proper address family configuration mode and then use the **neighbor inherit peer-policy** command. In the example, the neighbor at 209.165.201.130 is configured to inherit session settings from the EBGP_SESSION session template, and to inherit the IPv4 policy template from the BGP_POLICY policy template.

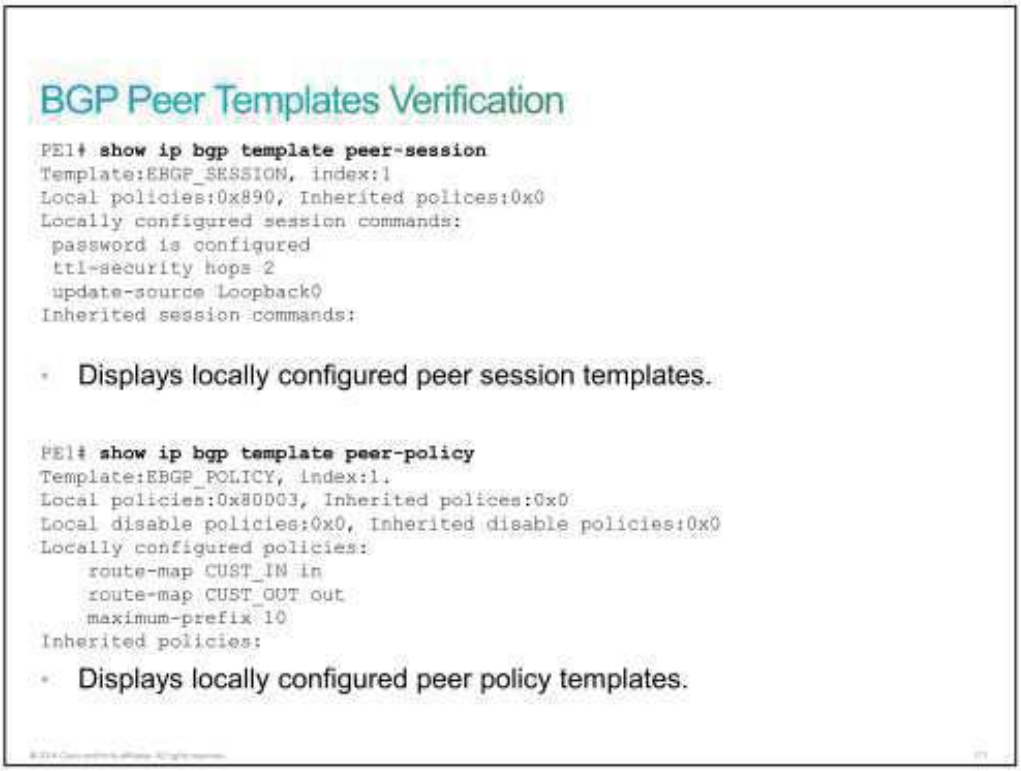
Note	For a complete command reference, refer to the Cisco IOS and IOS XE Software Command Reference guides on www.cisco.com.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

BGP Peer Templates Verification

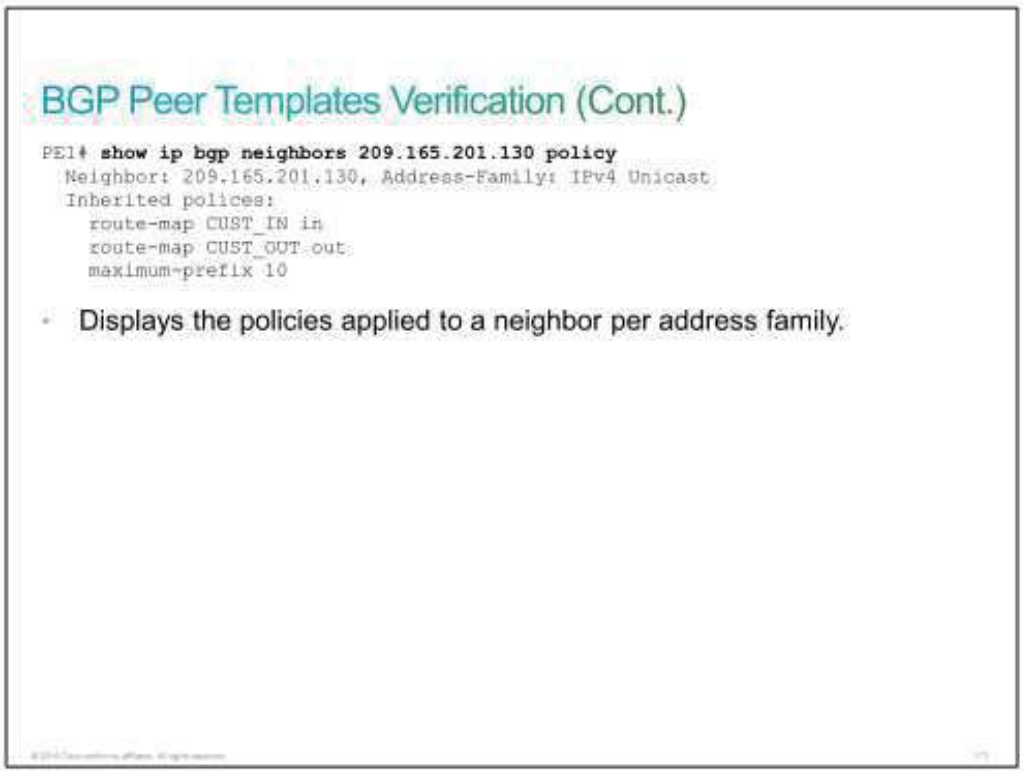
This topic explains BGP peer template verification on Cisco IOS and IOS XE.



To display peer policy template configurations, use the **show ip bgp template peer-session** command.

To display locally configured peer policy templates, use the **show ip bgp template peer-policy** command.

https://t.me/learningnets



To verify policies (directly configured and inherited) that are applied to a BGP neighbor per address family, use the **show ip bgp neighbors policy** command. In the example, the neighbor at 209.165.201.130 inherits route maps and the number of maximum allowed prefixes for the IPv4 address family.

The following **show** output example indicates that the policy is directly configured:

```
PE1# show ip bgp neighbors 10.1.1.1 policy
  Neighbor: 10.1.1.1, Address-Family: IPv4 Unicast
  Locally configured policies:
   maximum-prefix 100
```


https://t.me/learningnets

Summary

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

Summary

- BGP peer groups (Cisco IOS and IOS XE Software) were designed primarily for CPU optimization.
- BGP peer groups (Cisco IOS and IOS XE Software) can also be used for configuration optimization.
- To use BGP peer groups, use the **neighbor** *name* **peer-group** command to create a peer group. Then, assign specific BGP parameters to the peer group.
- BGP peer groups are used on a router with several neighbors that have similar but not completely identical policies.
- BGP dynamic update groups separate BGP update generation from neighbor configuration.
- BGP configuration (Cisco IOS XR Software) and peer templates (Cisco IOS and IOS XE Software) are reusable configuration patterns that support inheritance.

Summary (Cont.)

- IOS XR supports three types of BGP templates:
  - AF groups: contain address family-dependent information
  - session groups: contain address family-independent information
  - neighbor groups: universal
- BGP configuration template inheritance occurs between different template groups.
- To configure the neighbor to inherit configuration from the neighbor group, enter neighbor configuration mode, and use the **use neighbor-group** command, followed by the neighbor group name.
- To display the neighbors and neighbor groups that inherit configuration from a neighbor group, use the **show bgp neighbor-group users** command.

https://t.me/learningnets

Summary (Cont.)

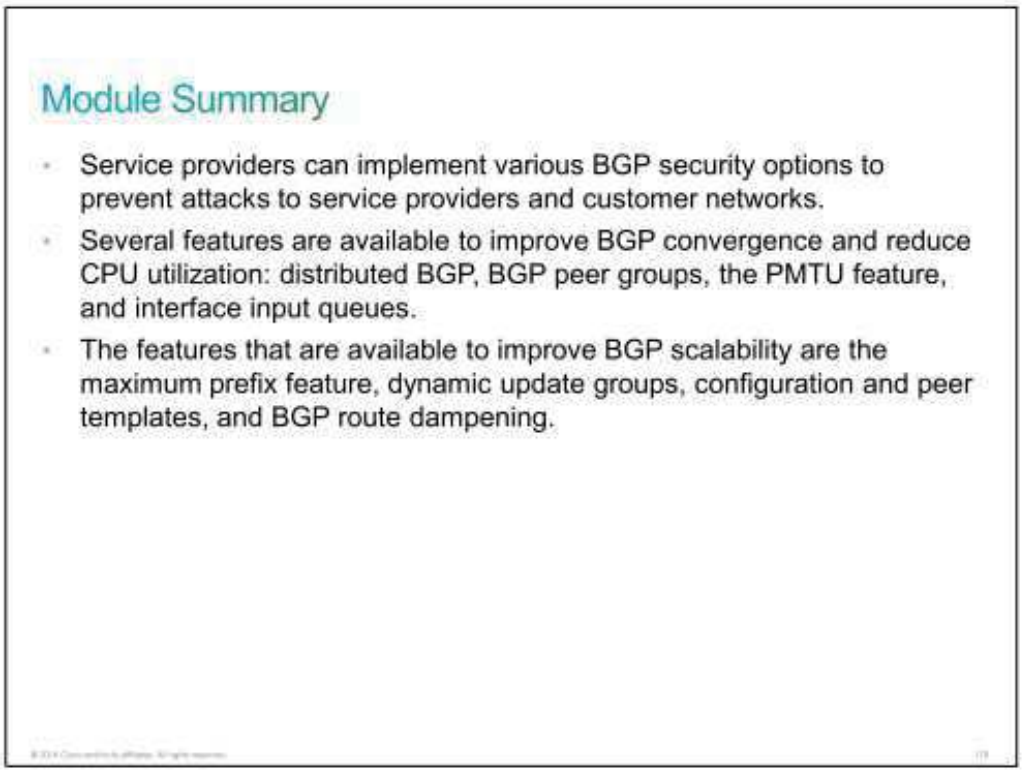
- Cisco IOS and IOS XE Software support two types of BGP templates:
  - Peer session templates: contain configuration common to all address families
  - Peer policy templates: contain configuration applied within a specific address family
- BGP peer session templates support direct and indirect inheritance.
- To create a session BGP peer session template, use the **template peer-session** command, followed by the session template name.
- To display BGP peer policy template configurations, use the **show ip bgp template peer-session** command.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Lesson 4: Module Summary

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



This module discussed BGP security, convergence, and scalability options. To implement BGP security options in a service provider network, the network administrator can implement BGP authentication, a BGP TTL security check, the maximum prefix feature, and CoPP. The network administrator can also implement remote-triggered black-hole filtering to protect the infrastructure from DoS attacks.

To reduce BGP convergence time and CPU utilization, the network administrator can implement distributed BGP (available on Cisco IOS XR Software platforms), BGP peer groups, the path MTU feature, and interface input queues.

BGP deployments should also be scalable. For scalability, network administrators can implement configuration and peer templates, and BGP route dampening.

https://t.me/learningnets

 ©2014 Cisco Systems, Inc.

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 cryptographic algorithm is supported in BGP neighbor authentication on Cisco IOS Software platforms? (Source: Implementing Advanced BGP Operations)

   A. AES
   B. MD5
   C. SHA
   D. 3DES

2. The BGP maximum prefix allows the controlling of the total number of prefixes that can be installed into the routing table. (Source: Implementing Advanced BGP Operations)

   A. true
   B. false

3. Remote-triggered black-hole filtering decreases collateral damage to a service provider during DDoS attacks to customers. (Source: Implementing Advanced BGP Operations)

   A. true
   B. false

4. Which two options can be used to support traffic forwarding during RP switchover? (Choose two.) (Source: Implementing Advanced BGP Operations)

   A. Nonstop Forwarding
   B. Cisco Express Forwarding
   C. Perfect Forward Routing
   D. Nonstop Routing

https://t.me/learningnets

5. Refer to the following Cisco IOS XR Software router output:

```
RP/0/RSP0/CPU0:P1# show bgp summary
BGP router identifier 10.5.1.1, local AS number 64500
BGP generic scan interval 30 secs
BGP table state: Active
Table ID: 0xe0000000   RD version: 15
BGP main routing table version 15
BGP scan interval 60 secs
BGP is operating in DISTRIBUTED mode.
Process  Id   RcvTblVer   bRIB/RIB   LabelVer  ImportVer  SendTblVer
StandbyVer
Speaker  1           15         15         12         12          12
15
bRIB     1           15         15         15         15          15
0
Neighbor        Spk    AS MsgRcvd MsgSent   TblVer  InQ OutQ  Up/Down
St/PfxRcd
10.0.1.1          1 64500    4591    4587       12    0    0    3d04h
4
10.0.2.1          1 64500    4588    4587       12    0    0    3d04h
1
192.168.105.51    1 64505    5050    4594       12    0    0    3d02h
2
```

Which two parameters would indicate that the BGP network has converged? (Choose two.) (Source: Improving BGP Convergence)

A. The TblVer for all neighbors is 12.
B. Spk is set to 1 for all neighbors.
C. The InQ and OutQ for all neighbors is 0.
D. All neighbors are in the Established state and have a number in the PfxRcd column.

6. What is the main task of the BGP scanner process? (Source: Improving BGP Convergence)

A. sends routing updates to BGP neighbors
B. walks the BGP table for routes to enter into the IP routing table
C. confirms the reachability of BGP next hops
D. scans the router configuration to establish and maintain BGP neighbors

7. On the Cisco IOS XR Software platform, how many BGP speaker processes can be enabled with distributed BGP? (Source: Improving BGP Convergence)

A. 4
B. 15
C. 16
D. 32

https://t.me/learningnets

8. What are two potential issues that are caused by modifying the default scan time and advertisement interval on a BGP router? (Choose two.) (Source: Improving BGP Convergence)

   A. Router CPU resources can be exhausted.
   B. Router memory resources can be depleted.
   C. Routing loops are more likely.
   D. BGP could converge faster than the IGP and cause network black holes.

9. Examine the following output.

```
%ROUTING-BGP-5-ADJCHANGE : neighbor 209.165.201.134 Down - Peer exceeding
maximum prefix limit (CEASE notification sent - maximum number of prefixes
reached) (VRF: default)
```

   What is the reason that the BGP session with a neighbor at 209.165.201.134 went down? (Source: Implementing Advanced BGP Operations)

   A. BGP authentication has been enabled.
   B. Routes from the neighbor were dampened too many times.
   C. The neighbor sent more routes than it is allowed to send.
   D. The IP address of the neighbor is not reachable anymore

10. On the Cisco IOS Software router running Software Release 15.0(1), BGP peer group usage is of utmost importance for scalable BGP deployments. (Source: Improving BGP Configuration Scalability)

    A. true
    B. false

11. Which two characteristics accurately describe the function of the BGP Dynamic Update Peer Groups feature? (Choose two.) (Source: Improving BGP Configuration Scalability)

    A. does not provide the operator with time to change the configuration if a mistake is made
    B. separates BGP update generation from peer group configuration
    C. does not require any configuration by the network operator
    D. requires minimal configuration by the network operator

12. Which two features can be used when configuring configuration templates? (Choose two.) (Source: Improving BGP Configuration Scalability)

    A. address family groups
    B. peer session templates
    C. peer policy templates
    D. session groups

https://t.me/learningnets

13. Which two events happen to an EBGP route that has become unreachable when BGP route dampening is used? (Choose two.) (Source: Improving BGP Configuration Scalability)

A. It is removed from the IP routing table.
B. It is removed from the BGP table.
C. It remains in the IP routing table as long as its penalty remains greater than 50 percent of the reuse limit.
D. It is kept in the BGP table and marked as a history entry.

14. Which two methods enable route dampening on a Cisco IOS XR Software router? (Choose two.) (Source: Improving BGP Convergence)

A. globally by enabling route dampening in global configuration mode
B. globally by enabling route dampening under the BGP routing process
C. on specific routes by enabling route dampening on a specific interface
D. by using a route policy in the BGP process to apply route dampening to specific routes

 © 2014 Cisco Systems, Inc.

https://t.me/learningnets

Module Self-Check Answers

Answer Key

1. B
2. B
3. A
4. A, D
5. A, C
6. C
7. B
8. A, B
9. C
10. B
11. B, C
12. A, D
13. A, D
14. B, D

https://t.me/learningnets

https://t.me/learningnets

Module 4: Multicast Overview

Introduction

This module is the entry point to IP multicast services. It presents the functional model of IP multicast and gives an overview of technologies that are present in IP multicasting. The module contains the business, theoretical, and implementation background that is needed by designers, implementers, and operations staff in service provider networks. The module is composed of an introduction to IP multicast concepts, a discussion of distribution trees and protocols, and an explanation of multicast technologies on the data link layer.

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

- List various types of IP multicast applications and explain their requirements
- Explain how protocols are building the IP multicast distribution tree, and review the functions and methods that are performed within the IP multicast-enabled network in order to ensure the delivery path from multicast sources to receivers
- Identify IP multicast issues on a data link layer
- Explain the purpose of the multicast route table and routing protocol supported for mroute table population

https://t.me/learningnets

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Lesson 1: Introducing IP Multicast

Overview

IP multicast is fundamentally changing the way that we live, work, play, and learn by providing innovative solutions that are simple, highly available, virtualized, open, and safe. This bandwidth conservation technology reduces traffic and server loads by simultaneously delivering a single stream of information to thousands of users. Applications that take advantage of multicast technologies include video conferencing, corporate communications, distance learning, and distribution of software, stock quotes, and news.

Since multicast is a different transmission mode from unicast, only protocols that are designed for multicast can be used with multicast. Multicast networks have source, network, and receiver segments. All of these segments must perform specific functions in order to deliver the multicast traffic. In this lesson, you will examine the benefits of IP multicast and IP addressing used for multicast. You will also be introduced to multicast-enabled applications.

Upon completing this lesson, you will be able to:

- Explain the benefits of IP multicasting
- Discuss high-level multicast operations
- Discuss the advantages and disadvantages of multicast over unicast transmission
- Discuss the types of multicast applications
- Discuss IP multicast group addresses
- Discuss IP multicast basic addressing
- Discuss IP multicast RFC 2770 and SSM addressing
- Discuss the multicast SDR concept
- Discuss the IP multicast service model
- Discuss the functions of a multicast network

https://t.me/learningnets

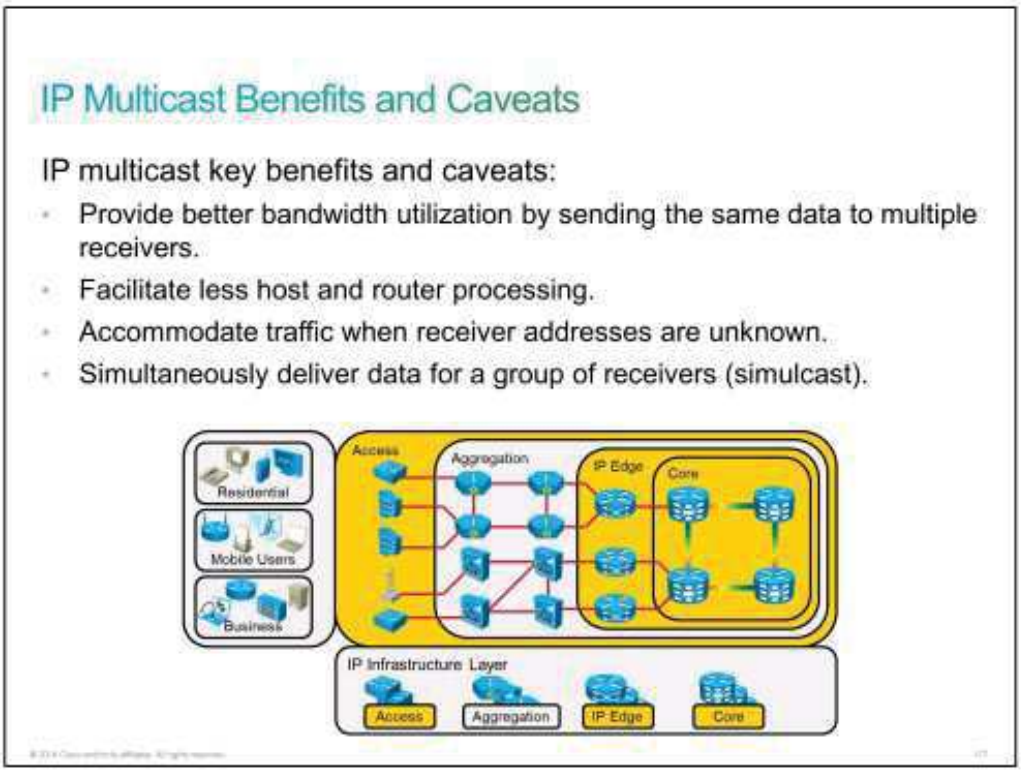
- Discuss the functions of multicast sources and receivers
- Discuss the protocols that are used to support multicast
- Discuss multicast forwarding and RPF check
- Discuss multicast scoping using TTL threshold and address scoping

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

IP Multicast Benefits and Caveats

IP multicast can be used in the access, aggregation, IP edge, and core layers of the Cisco IP Next-Generation Network infrastructure.



Multicast may be used to send the same data packets to multiple receivers. Doing this does not mean the packets are not duplicated for every receiver. Instead, they are sent in a single stream, where downstream routers perform packet multiplication over receiving links.

Routers process fewer packets because they receive only a single copy of the packet. This packet is then multiplied and sent on outgoing interfaces where there are receivers.

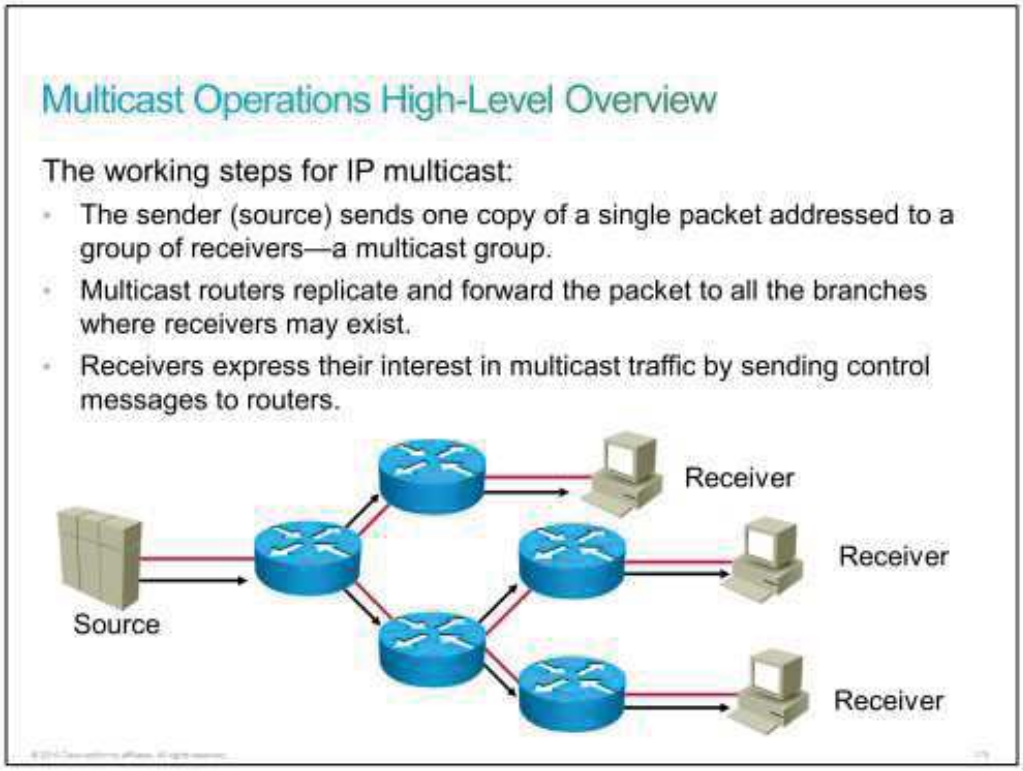
Because downstream routers perform packet multiplication and delivery to receivers, the sender or source of multicast traffic does not have to know the unicast addresses of the receiver.

Simulcast—simultaneous delivery for a group of receivers—may be used for several purposes, including audio and video streaming, news and similar data delivery, and software upgrade deployment.

https://t.me/learningnets

Multicast Operations High-Level Overview

The main idea of multicast is replicating a single packet for multiple receivers.



In multicast, the sender sends only one copy of a single data packet that is addressed to a group of receivers—a multicast group. Downstream multicast routers replicate and forward the data packet to all those branches where receivers exist. Receivers express their interest in multicast traffic by registering at their first-hop router using IGMP for IPv4 multicast or MLD for IPv6 multicast.

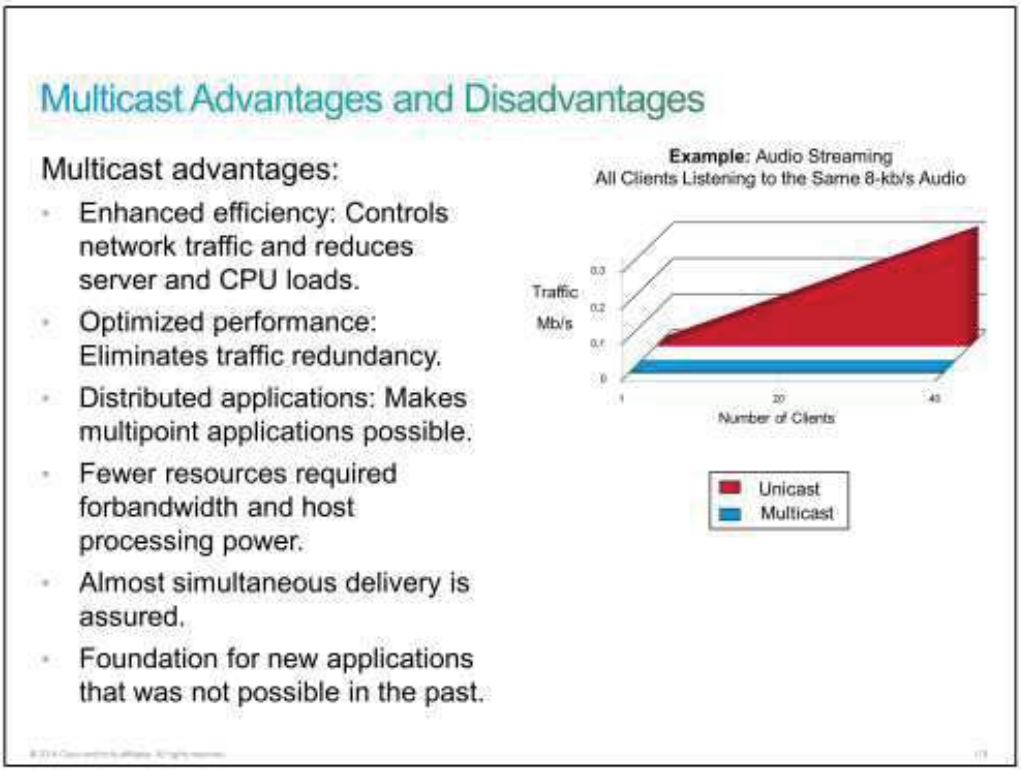
The figure shows a multicast source host transmitting one copy of data, and a network replicating the packet. Routers are responsible for replicating the packet and forwarding it to multiple recipients. Routers replicate the packet at any point where the network paths diverge, and use RPF techniques to ensure the packet is forwarded to the appropriate downstream paths without routing loops. Each packet exists only in a single copy on any given network. The multicast source host may send to multiple receivers simultaneously because it is sending only one packet.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Multicast Advantages and Disadvantages

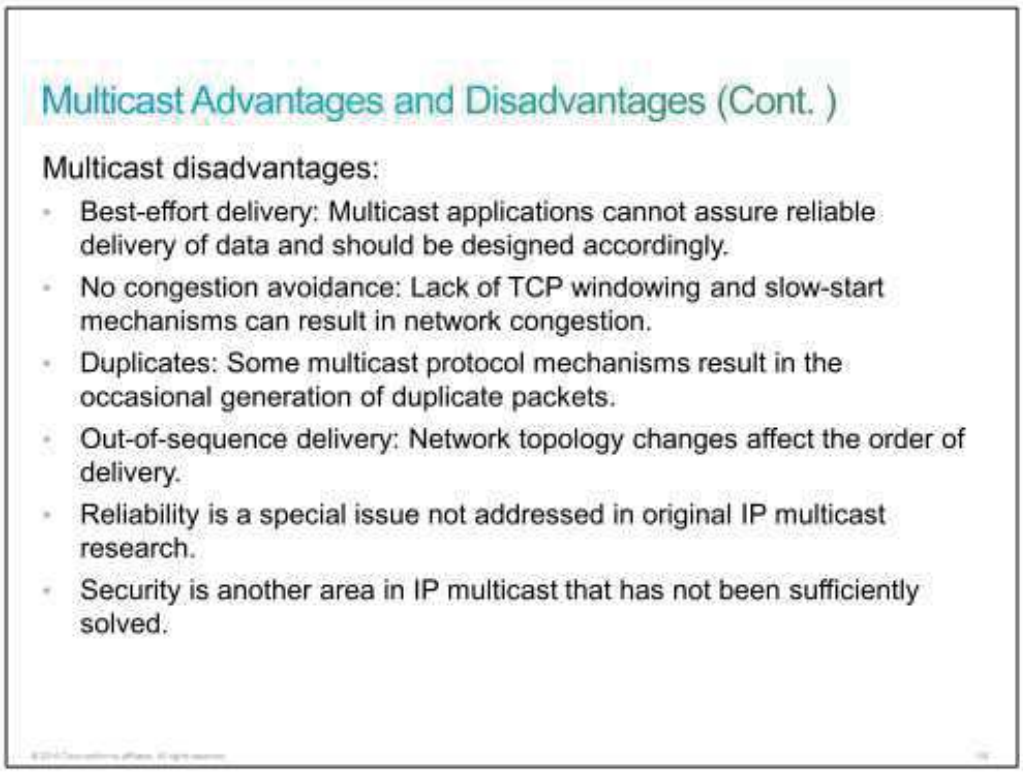
Knowing the advantages and disadvantages of multicast will help you understand why you should choose multicast transmission over unicast transmission in many environments.



Multicast transmission provides many advantages over unicast transmission:

- **Enhanced efficiency**: Available network bandwidth is utilized more efficiently because multiple streams of data are replaced with a single transmission.
- **Optimized performance**: Fewer copies of data require forwarding and processing.
- **Distributed applications**: Multipoint applications will not be possible as demand and usage grows, because unicast transmission will not scale. Traffic level and clients increase at a 1:1 rate with unicast transmission.
- For the equivalent amount of multicast traffic, the sender needs much less processing power and bandwidth.
- Multicast packets do not impose high-bandwidth utilization as unicast packets do, so there is a greater possibility that they will arrive almost simultaneously at the receivers.
- Multicast enables a whole range of new applications that were not possible on unicast (for example, IPTV).

https://t.me/learningnets



Most multicast applications are UDP-based. This foundation results in some undesirable consequences when compared to similar unicast TCP applications.

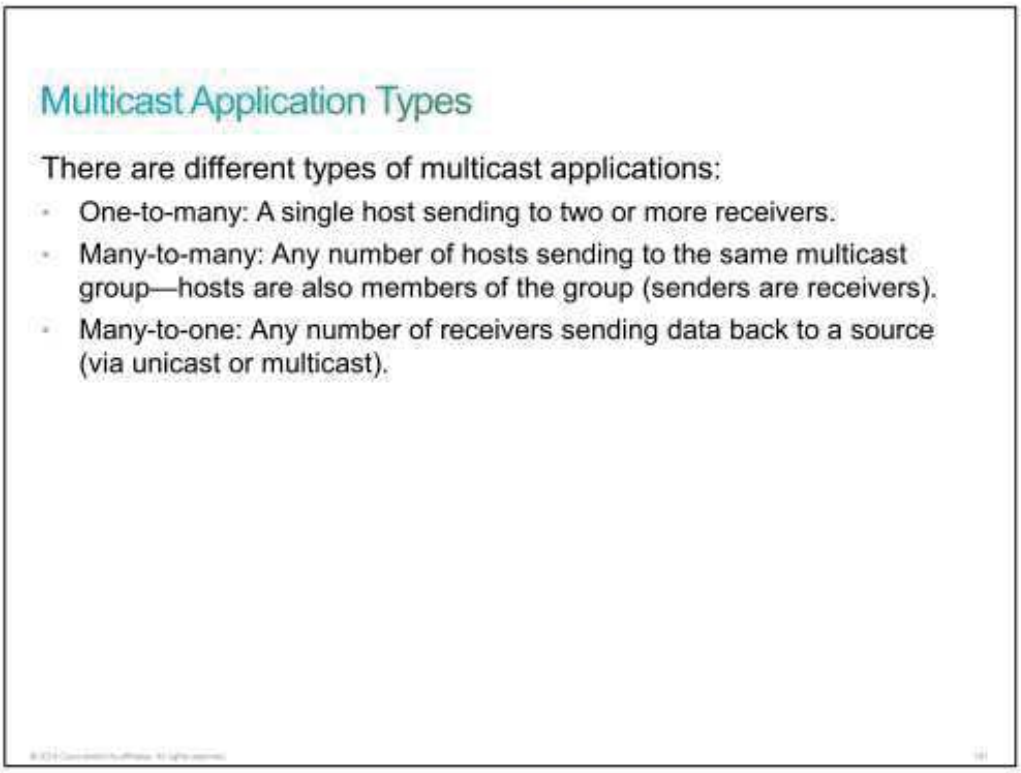
There are also some disadvantages of multicast that need to be considered:

- Best-effort delivery results in occasional packet drops. These losses may affect many multicast applications that operate in real time (for example, video and audio). Also, requesting retransmission of the lost data at the application layer in these not-quite-real-time applications is not feasible. Heavy drops on voice applications result in jerky, missed speech patterns that can make the content unintelligible when the drop rate gets high enough. Sometimes, moderate to heavy drops in video appear as unusual artifacts in the picture. However, even low drop rates can severely affect some compression algorithms. This action causes the picture to become jerky or to freeze for several seconds while the decompression algorithm recovers.
- No transport layer congestion control may result in overall network degradation as the popularity of UDP-based multicast applications grows.
- Duplicate packets may occasionally be generated as multicast network topologies change. Applications must expect occasional duplicate packets to arrive, and they should be designed accordingly.
- Out-of-sequence delivery of packets to the application may also occur during network topology changes or during other network events that affect the flow of multicast traffic.
- UDP has no reliability mechanisms, so reliability issues have to be addressed in multicast applications where reliable data transfer is necessary.
- It remains a challenge to restrict multicast traffic to only a selected group of receivers. Eavesdropping issues are not sufficiently solved yet. Some commercial applications (for example, financial data delivery) will only be possible when reliability and security issues are properly solved.

https://t.me/learningnets

Multicast Application Types

Multicast applications may be one-to-many, many-to-many or many-to-one.



There are different types of multicast applications. Here are two of the most common models:

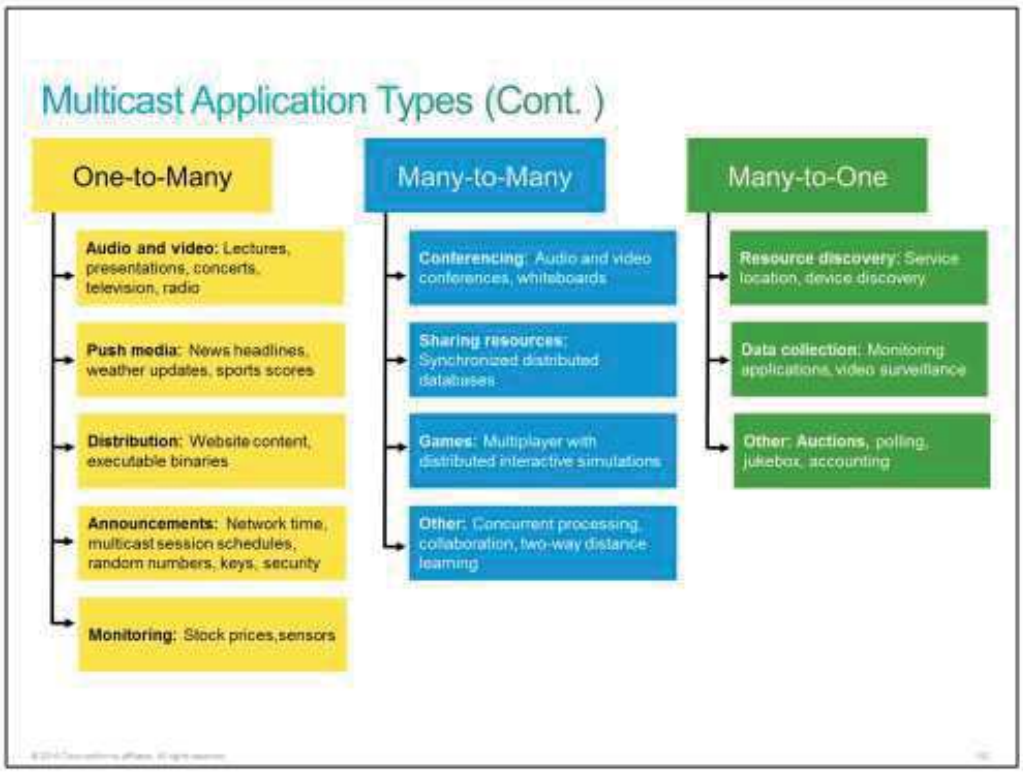
- One-to-many, where one sender sends data to many receivers.
- Many-to-many, where a host can simultaneously be a sender and a receiver.

Other models are also used, especially in financial applications and networks. These include many-to-one, where many receivers are sending data back to one sender, and few-to-many.

Many new multipoint applications are emerging as demand for them grows:

- Real-time applications include live broadcasts, financial data delivery, whiteboard collaboration, and video conferencing.
- Not-real-time applications include file transfer, data and file replication, and VoD.

https://t.me/learningnets



The figure presents major IP multicast application categories.

The one-to-many multicast application, where a single source is sending to an unknown group of receivers, is the simplest type of application. This application may be used for audio and video distribution, push media, announcements, monitoring, and so on. If a one-to-many application needs feedback from receivers, it becomes a many-to-many application. One-to-many applications (for example, file distribution) are tolerant to delays. There are also some applications that are sensitive to jitter (for example, audio and video conferencing). Some are not tolerant to delays at all (for example, real-time monitoring and financial applications).

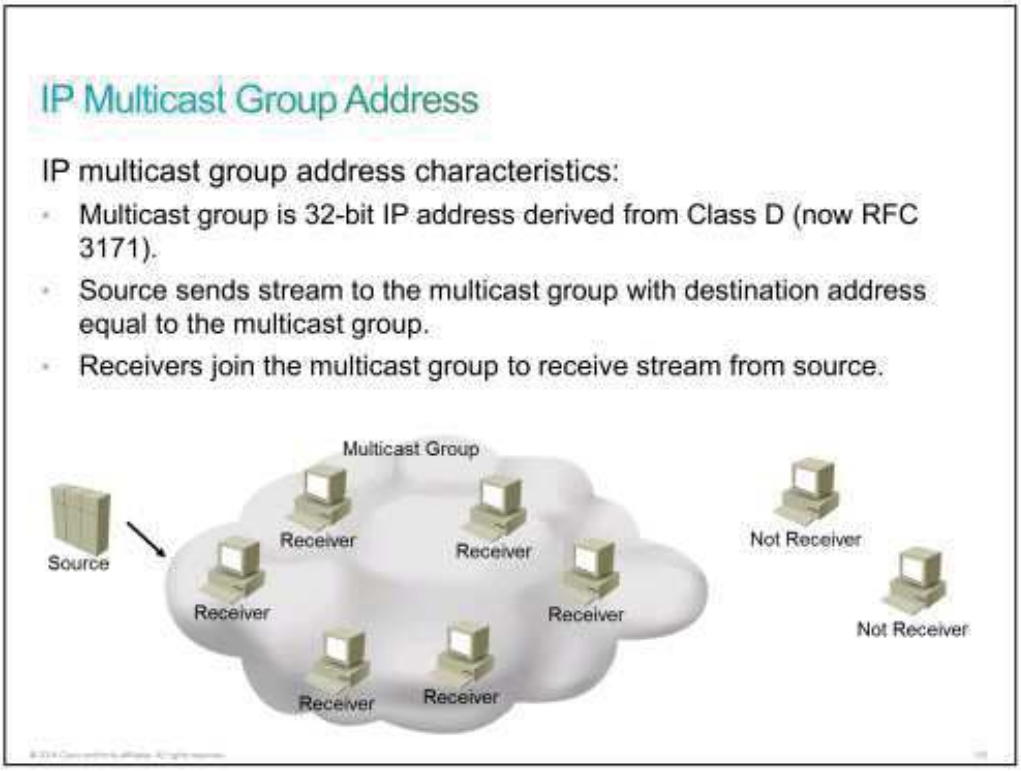
Many-to-many multicast applications are those where two or more receivers also act as senders. Receiving data from several sources increases the complexity of applications and creates different management challenges. Using a many-to-many multicast concept as a foundation, a whole new range of applications may be built (for example, collaboration, concurrent processing, and distributed interactive simulations). Many-to-many applications are intolerant to delays because they are bidirectional. Delays in multimedia conferencing higher than 500 ms may be very disturbing to the audience.

The many-to-one or many-to-few multicast model may be used for resource discovery, data collection, and similar applications. However, building many-to-one applications presents several challenges. Maintaining a multicast state in networks with many senders may cause scaling problems in the network that are transparent to the application. Many senders may cause message implosion problems at the receiver side. There are some solutions to these problems (for example, bidirectional trees in multicast routing), but application developers must be aware of these problems.

https://t.me/learningnets

IP Multicast Group Address

A multicast address is associated with a group of interested receivers.

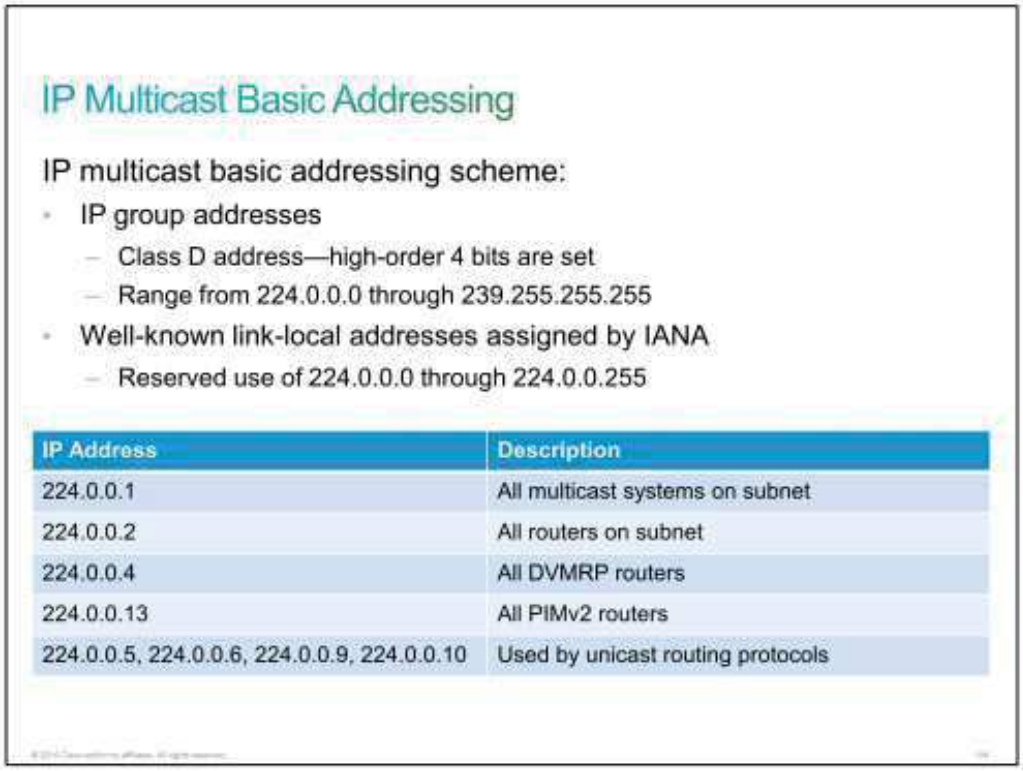


According to RFC 3171, addresses 224.0.0.0 through 239.255.255.255, the former Class D addresses, are designated as multicast addresses in IPv4. The sender sends a single datagram (from its unicast address) to the multicast address, and the intermediary routers take care of making copies and sending them to all receivers that have registered their interest in receiving the multicast data for that multicast address.

https://t.me/learningnets

IP Multicast Basic Addressing

Multicast IP addresses use the Class D address space.



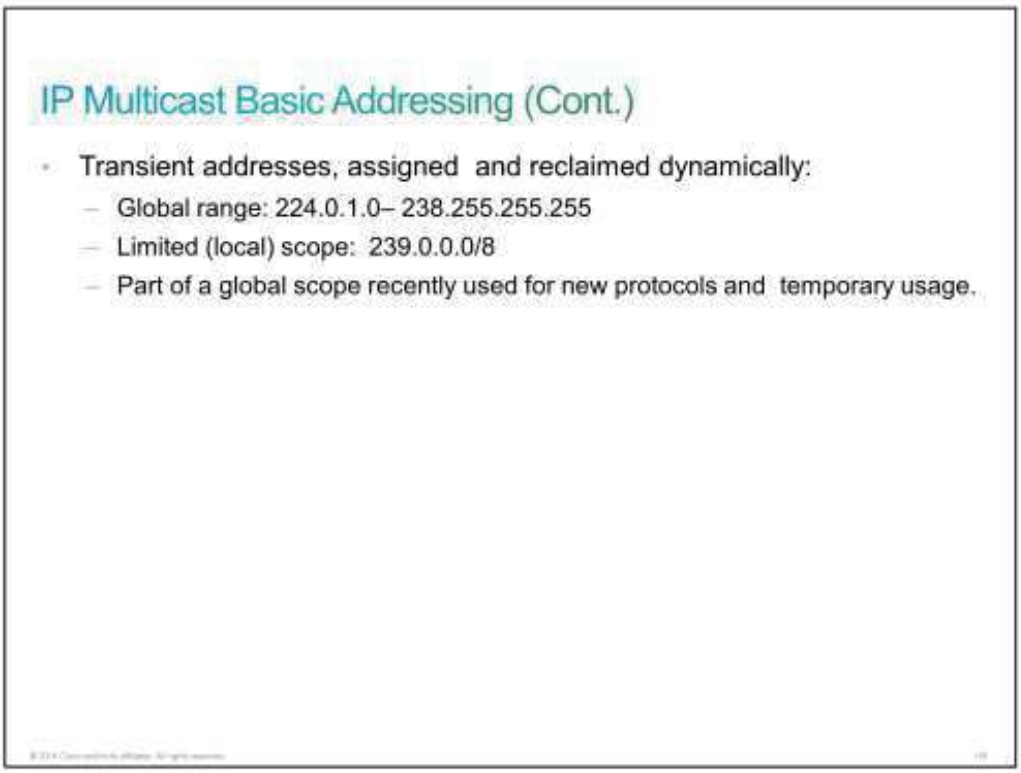
Class D addresses are denoted by the high-order 4 bits set to 1110.

The multicast IP address space is separated into the following address groups:

- Local-scope addresses are addresses 224.0.0.0 through 224.0.0.255 and are reserved by the IANA for network protocol use. Multicasts in this range are never forwarded off the local network, regardless of TTL, and usually the TTL is set to 1. Here are some examples of local multicast addresses:
  - 224.0.0.1 All hosts
  - 224.0.0.2 All multicast routers
  - 224.0.0.3 All DVMRP routers
  - 224.0.0.5 All OSPF routers
  - 224.0.0.6 All OSPF DRs

 © 2014 Cisco Systems, Inc.

https://t.me/learningnets



- Global-scope addresses are addresses 224.0.1.0 through 238.255.255.255, and are allocated dynamically throughout the Internet.
- Administratively scoped addresses are addresses 239.0.0.0 through 239.255.255.255, and are reserved for use inside private domains.

Global-scope multicast addresses and administratively scoped multicast addresses are transient addresses. These addresses are assigned and reclaimed dynamically within applications.

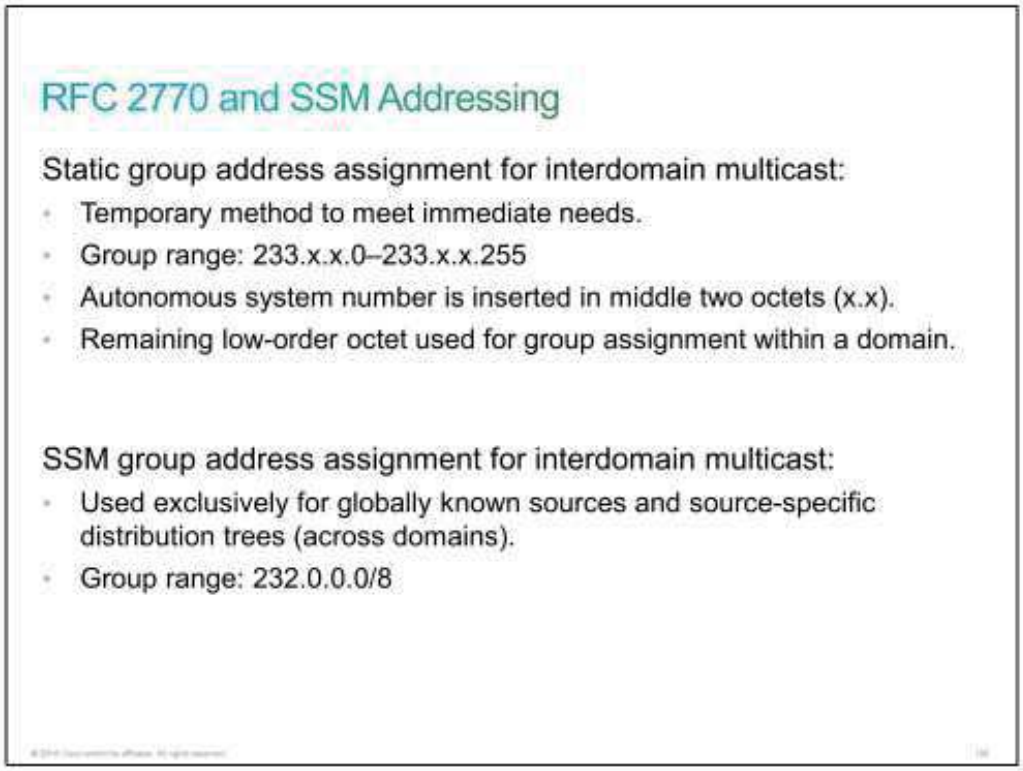
The administratively scoped multicast address space is divided into the following scopes:

- Local scope (239.255.0.0/16, and grows downward to 239.254.0.0/16, 239.253.0.0/16)
- Organization local scope (239.192.0.0/14 with possible expansion to ranges 239.0.0.0/10, 239.64.0.0/10, and 239.128.0.0/10)

https://t.me/learningnets

RFC 2770 and SSM Addressing

RFC 2770 describes static address assignment and SSM describes limiting source addresses for multicast groups.



The growth in demand for interdomain multicast routing led to some interim solutions such as GLOP. (Note that GLOP is not an acronym.) One of the methods for static address allocation for multicast groups is defined in RFC 2770, titled "GLOP Addressing in 233/8":

- Until MASC has been fully specified and deployed, many content providers of the Internet require something at the very least to begin address allocation. This necessity is being addressed with a temporary method of static multicast address allocation.
- The basic concept behind this methodology is as follows:
  - Use the 233.0.0.0/8 address space for static address allocation.
  - The middle two octets of the group address would contain your AS number.
  - The final octet is available for group assignment.

The GLOP address space is a means of assigning addresses in 233.0.0.0/8 based on an AS number, as described in RFC 2770. Basically, a /24 is assigned to each AS based on the 16-bit AS number. The method is not applicable to the newer 32-bit extension AS numbers. For a given AS number, which is converted into two octets (X and Y in this example) in the usual fashion of Internet addresses, the GLOP space is therefore 233.X.Y.0/24. This address range is assumed to be assigned by default and does not have to be explicitly requested. Note that a GLOP allocation only provides 256 globally unique multicast group addresses, which are widely viewed as not enough for large-scale broadcasters.

 © 2014 Cisco Systems, Inc.

https://t.me/learningnets

There are several situations where the sources (multicast senders) have to be globally known. A special range of multicast addresses has been dedicated for those servers, and a special multicast protocol is in development—SSM. SSM is a method of delivering multicast packets in which the only packets that are delivered to a receiver are those originating from a specific source address requested by the receiver. By limiting the source in this manner, SSM reduces demands on the network and improves security. This protocol will always allow the building of the distribution tree that is rooted at the source for any group address from the range 232.0.0.0/8 (232.0.0.1–232.255.255.255).

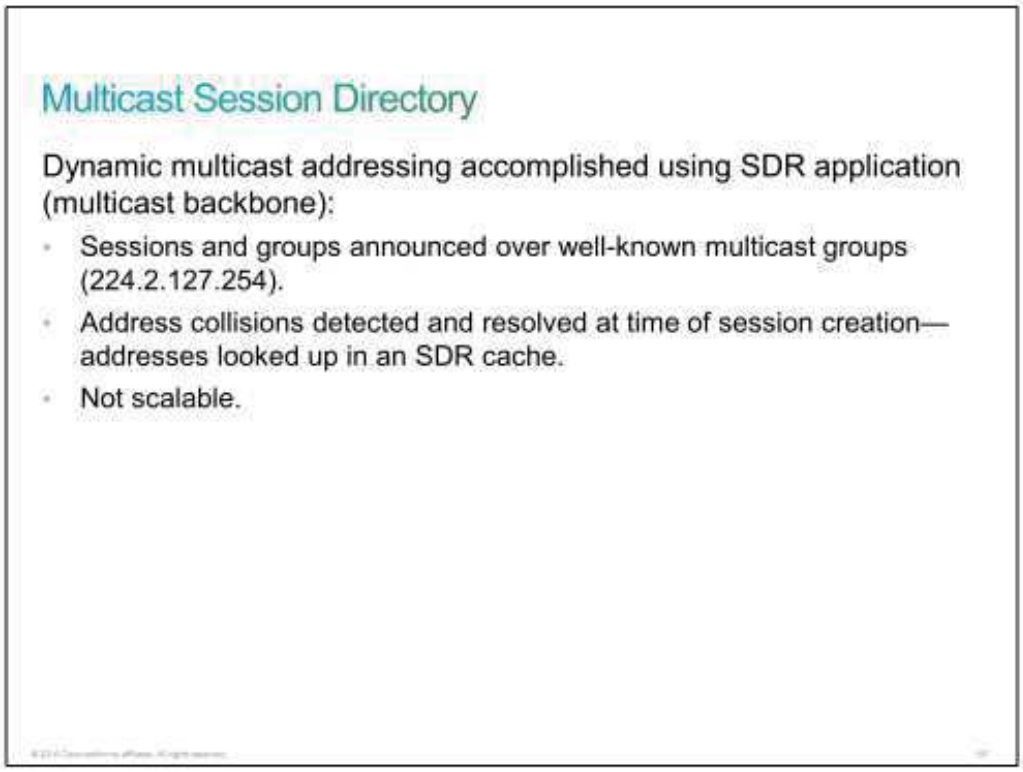
SSM is best understood in contrast to ASM. In the ASM service model, a receiver expresses interest in traffic to a multicast address. The multicast network must discover all multicast sources sending to that address, and route data from all multicast sources to all interested receivers.

In the SSM service model, in addition to the receiver expressing interest in traffic to a multicast address, the receiver expresses interest in receiving traffic only from specific multicast sources sending to that multicast address. This relieves the network of discovering many multicast sources, and reduces the amount of multicast routing information that the network must maintain.

https://t.me/learningnets

Multicast Session Directory

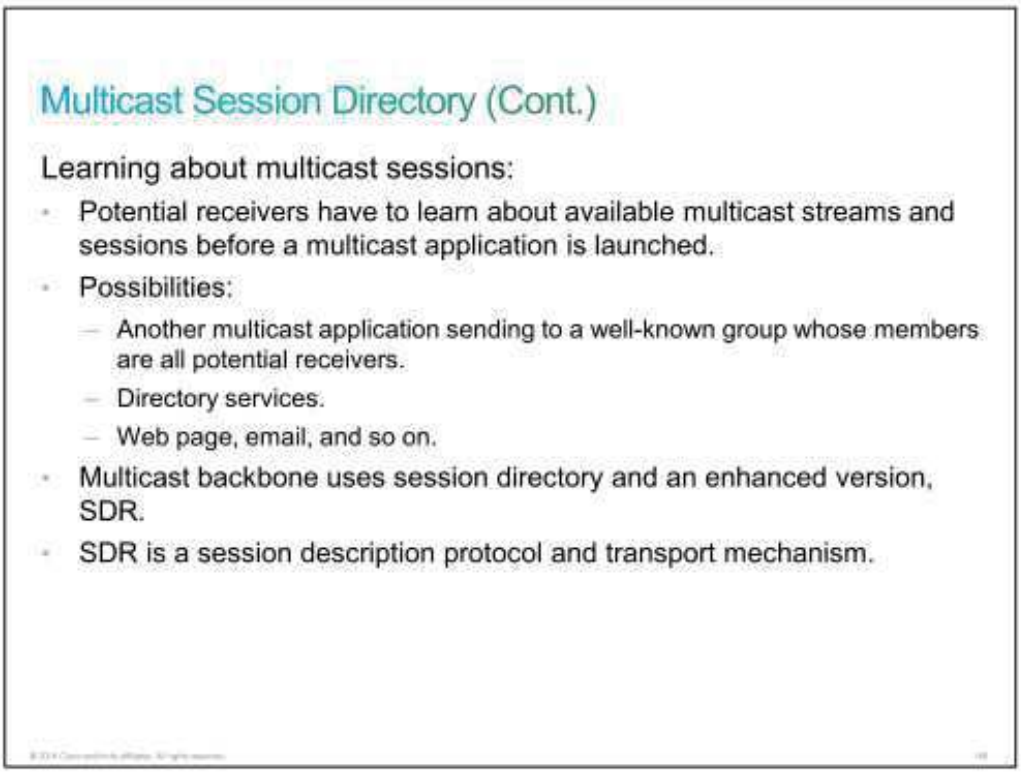
Multicast SDR application is the way to scale dynamic address assignments for multicast groups and to learn about available multicast streams and sessions.



The scalability in multicast address allocation can only be achieved with dynamic group address assignment methods:

- SDR application:
  - Used to pick an unused multicast group address, the SDR application detects collisions in an IP multicast group address assignment at the time new sessions are created.
  - Although it was sufficient for use on the old multicast backbone when the total number of multicast sessions in the Internet was quite low, SDR has severe scaling problems that preclude it from continuing to be used as the number of multicast sessions increases.

https://t.me/learningnets



Whenever a multicast application is started on a receiver, the application has to know which multicast group to join. The application has to learn about the available sessions and streams, which typically map to one or more IP multicast groups.

Applications can learn about the multicast sessions in several ways:

- The application may join a well-known, predefined group, to which announcements about available multicast sessions are sent.
- Some type of directory services may be available, and the application may contact the appropriate directory server.
- The application may be launched from a web page on which the multicast sessions are listed as URLs. Even email may be used.

In the multicast backbone, the session directory application served as a means for announcing available sessions and to assist in creating new sessions. The initial session directory tool was revised, resulting in the SDR tool. SDR is an applications tool that allows the following:

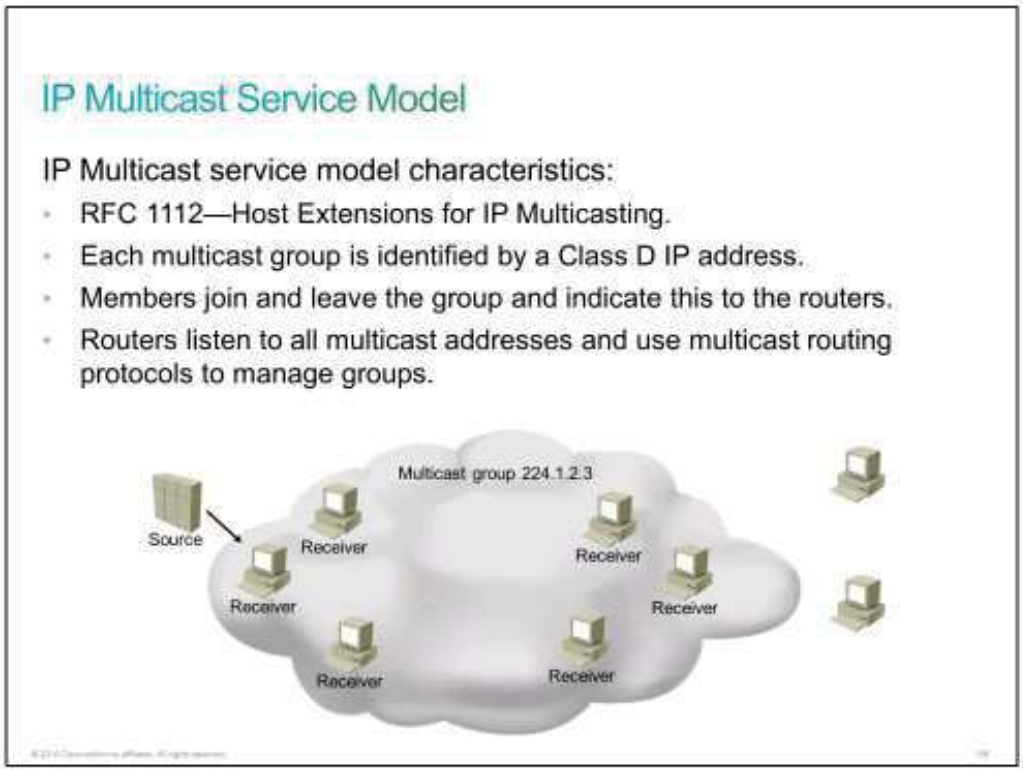
- Session description and its announcement.
- Transport of session announcement via well-known multicast groups (224.2.127.254).
- Creation of new sessions.

SDR uses SAP, which will periodically multicast a session announcement packet describing a particular session. SAP announcement packets can be received by a multicast receiver by joining the well-known group 224.2.127.254. A user can then select the option to receive traffic for a multicast group.

https://t.me/learningnets

IP Multicast Service Model

IP multicast service models consist of three main components: senders send to a multicast address, receivers express an interest in a multicast address, and routers deliver traffic from the senders to the receivers.



RFC 1112 specifies the host extensions for IP to support multicast:

- IP multicast allows hosts to join a group that receives multicast packets.
- It allows users to dynamically register (join or leave multicast groups) based on the applications they use.
- It uses IP datagrams to transmit data.

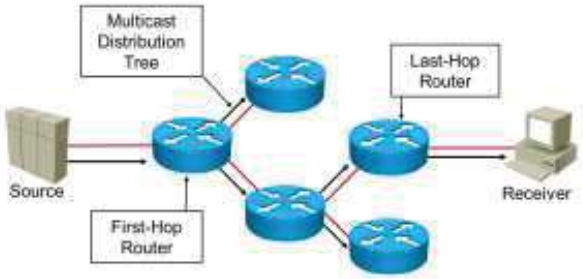
The Class D IP addresses (224.0.0.0 to 239.255.255.255) define each multicast group. Addresses are allocated dynamically and represent receiver groups, not the individual hosts.

Receivers may dynamically join or leave an IPv4 multicast group at any time using IGMP messages, and may dynamically join or leave an IPv6 multicast group at any time using MLD messages. Messages are sent to the last-hop routers, which manage group membership.

Routers use multicast routing protocols—for example, DVMRP, MOSPF, CBT, PIM, and so on—to efficiently forward multicast data to multiple receivers. The routers listen to all multicast addresses and create multicast distribution trees, which are used for multicast packet forwarding.

https://t.me/learningnets

IP Multicast Service Model (Cont.)



- Multicast network routers are distinct from source and receiver segments.
- Sources simply start sending data without any indication and first-hop routers forward data.
- Receivers report their membership to last-hop routers.
- Last-hop (leaf) routers communicate group membership to the network.

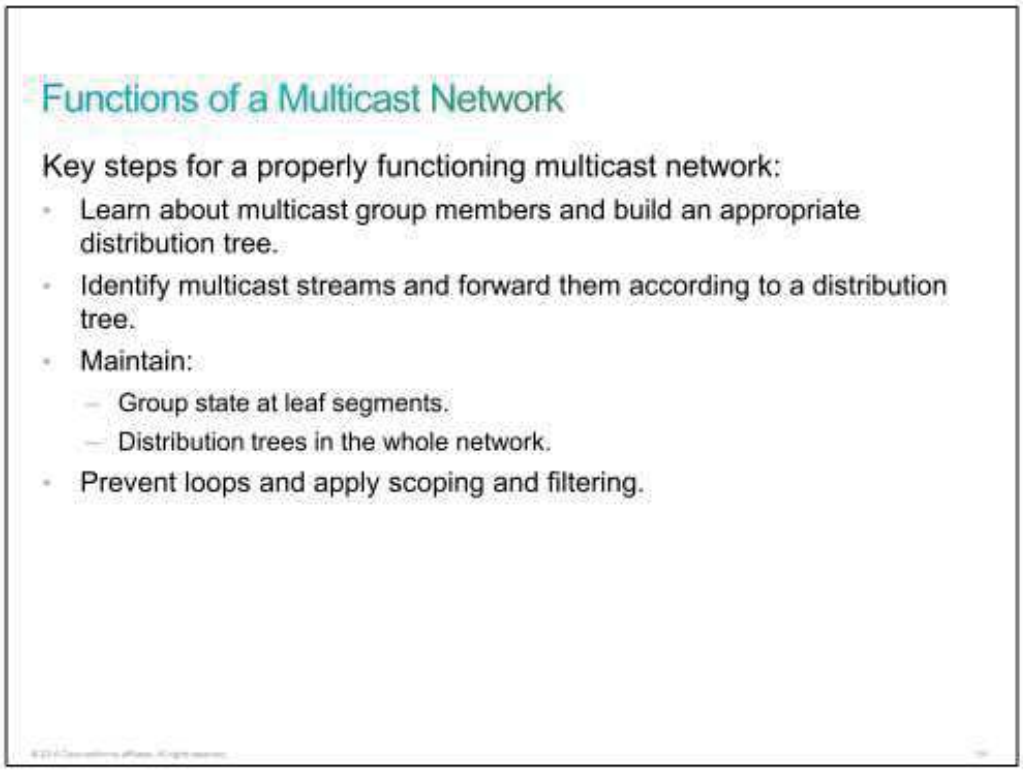
Routers identify multicast traffic and forward the packets from senders toward the receivers. When the source becomes active, it starts sending the data without any indication. First-hop routers, to which the sources are directly connected, start forwarding the data to the network. Receivers that are interested in receiving IPv4 multicast data register to the last-hop routers using IGMP membership messages. Last-hop routers are those routers that have directly connected receivers. Last-hop routers forward the group membership information of their receivers to the network, so that the other routers are informed about which multicast flows are needed.

The figure shows a multicast source that is connected to a first-hop router, which forwards multicast packets into the network. Packets traverse the SPTs on their way to the receivers toward the last-hop router.

https://t.me/learningnets

Functions of a Multicast Network

For a multicast network to function properly, a multicast distribution tree should be built and maintained by network multicast routers.



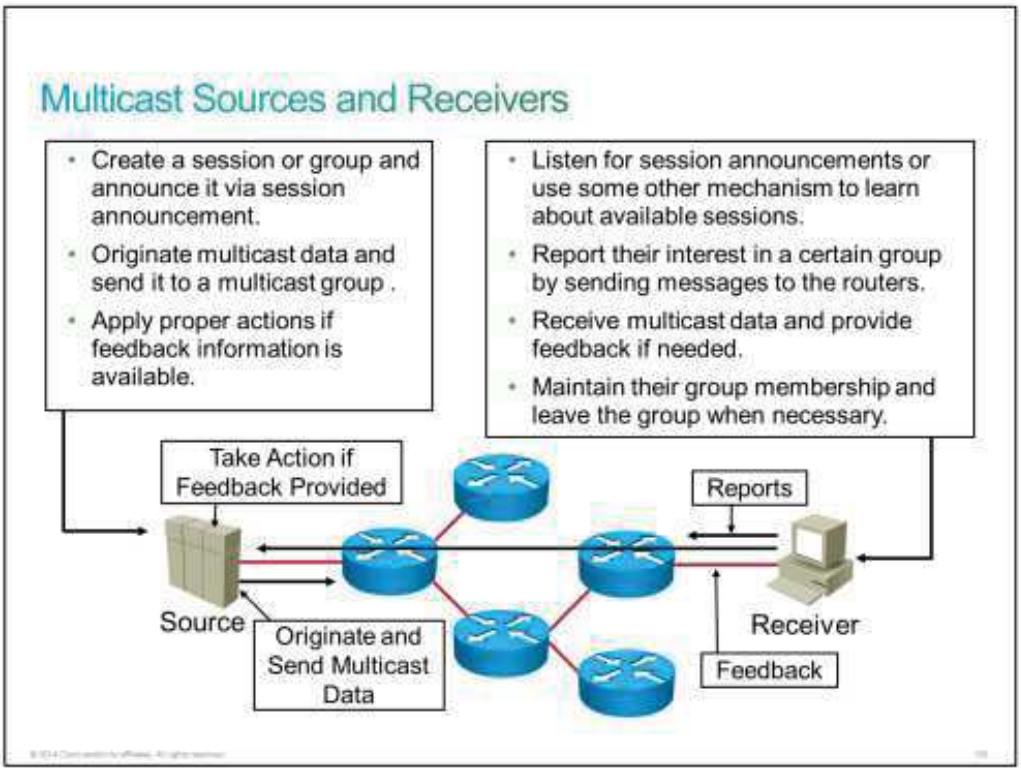
To build appropriate multicast distribution trees, the multicast network routers must learn about their multicast-enabled neighbors. As they build the multicast distribution trees, the routers start forwarding multicast traffic according to network needs. During normal operation, routers maintain the multicast distribution trees and multicast group state at leaf segments. The routers also prevent loops and apply scoping or filtering.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Multicast Source and Receivers

Multicast sources create a session and announce it via session announcements. These sources originate multicast data and send it to a multicast group. If there is any feedback that is received from the receivers, they apply proper actions. For example, they can change data encoding to compensate for lower bandwidth.



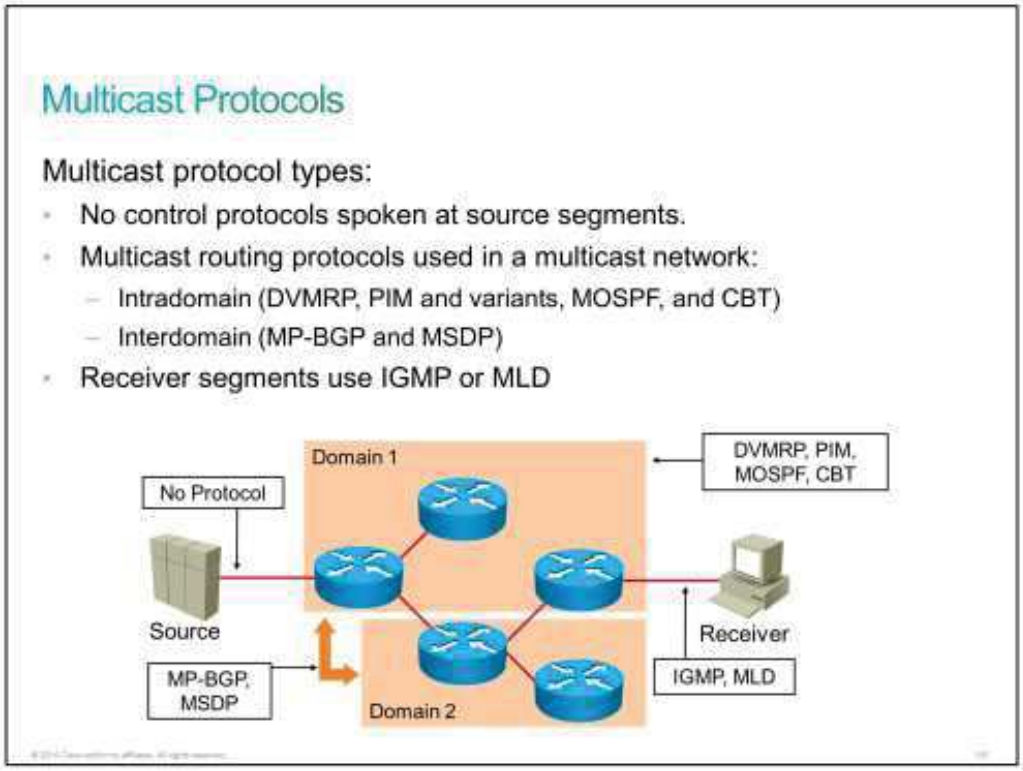
Multicast receivers listen to the session announcements or use some other mechanism to learn about the available multicast sessions. Receivers may select certain multicast groups and report their interest by sending a host membership report to the last-hop router. The last-hop router then starts forwarding traffic for a specified multicast group to the receivers. After the receivers start receiving multicast data, they maintain their group membership and may also provide feedback to the source about the received data. When receivers want to stop receiving certain multicast data, they send a leave message to the router.

Unicast packets are delivered to a specific recipient on an Ethernet or IEEE 802.3 subnet by setting a specific Layer 2 MAC address on the Ethernet packet address. Broadcast packets make use of a broadcast MAC address (FF:FF:FF:FF:FF:FF), which includes setting the broadcast or multicast bit in the address. Multicast packets are delivered by using the Ethernet MAC address range 01:00:5e:00:00:00–01:00:5e:7f:ff:ff. This address allows for 23 bits of available address space. The first octet (01) includes the broadcast or multicast bit. The lower 23 bits of the 28-bit multicast IP address are mapped into the 23 bits of available Ethernet address space. This configuration creates ambiguity in delivering packets. If two hosts on the same subnet each subscribe to a different multicast group whose address differs only in the first five bits, Ethernet packets for both multicast groups will be delivered to both hosts. This scenario, in turn, would require the network software in the hosts to discard the unnecessary packets.

https://t.me/learningnets

Multicast Protocols

Multicast protocols may differ depending on where in a multicast network they are implemented.



There is no multicast protocol for source registering used between the source and the first-hop router, such as IGMP, which is used between the last-hop router and the receiver.

Inside the multicast network, various multicast routing protocols are used. The multicast routing protocols may be separated into these two groups:

- **Intradomain**: DVMRP, PIM, MOSPF, and CBT
- **Interdomain**: MP-BGP in combination with MSDP

Between the last-hop router and the receivers, IGMP or MLD are used. Receivers use IGMP (for IPv4) or MLD (for IPv6) to report their multicast group membership to the router.

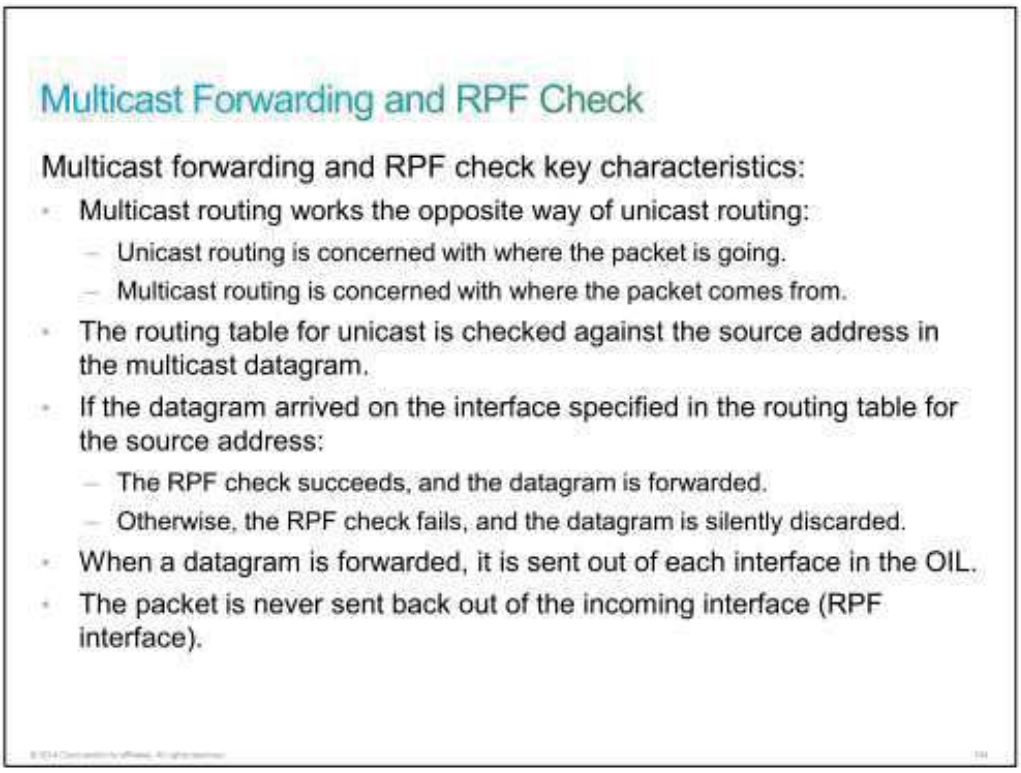
Note	The DVMRP, MOSPF, and CBT protocols are outside of the scope of this course and will not be discussed further.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Multicast Forwarding and RPF Check

In unicast routing, when the router receives the packet, the decision whether to forward the packet is made depending on the destination address of the packet. In multicast routing, where to forward the multicast packet depends on where the packet came from.



Multicast routers must know the origin of the packet in addition to its destination, which is the opposite of unicast routing. In multicast origination, the source IP address denotes the known multicast source, and the destination IP address denotes a group of unknown receivers. Multicast sources use the group address as the IP destination address in their data packets. The receiver uses this group address to inform the network that they are interested in receiving the packets that are sent to that group.

Multicast routing uses the RPF mechanism to prevent forwarding loops and to ensure that the multicast traffic is using the shortest path from the source to the receivers.

When a router receives a multicast packet, it checks its routing tables (usually the unicast routing table) to determine whether the interface on which the multicast packet was received provides the shortest path back to the multicast source. If the interface provides the shortest path to the source, the router forwards the packet. If the multicast packet arrived on an interface that does not provide the shortest path to the source, it is silently discarded.

After the router receives a multicast packet, it performs an RPF check. If the RPF check succeeds, the packet is forwarded. Otherwise, it is silently discarded.

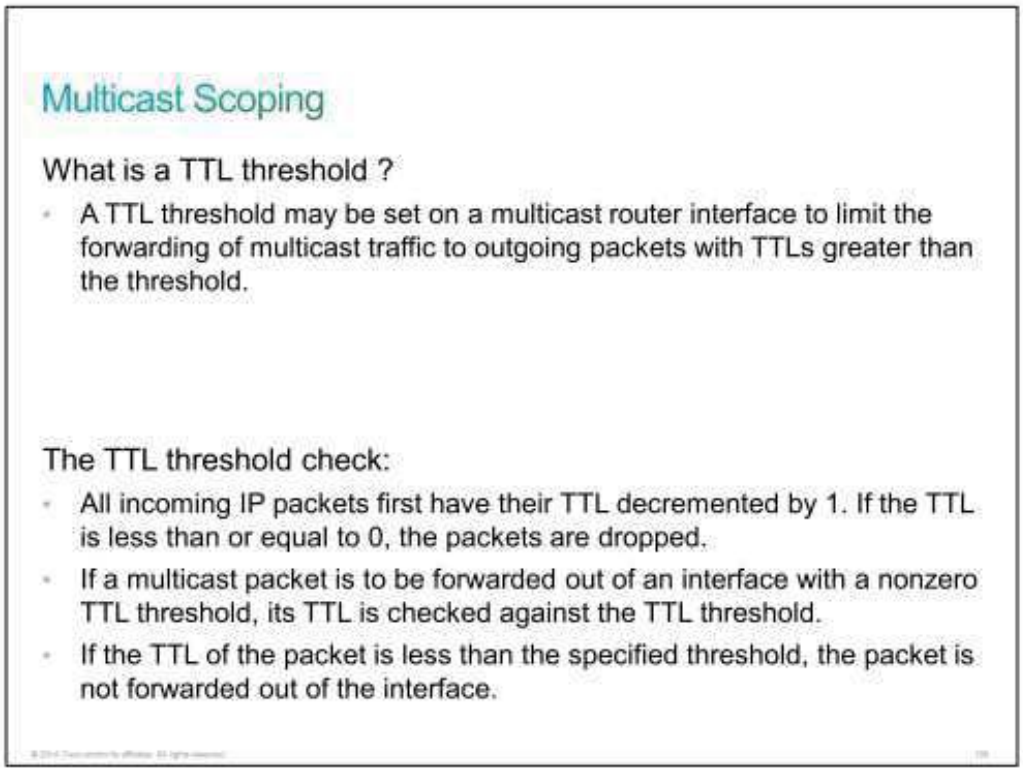
The multicast packet is forwarded out of each interface that is in the OIL. OIL entries list the interfaces of the multicast neighbors that are downstream of the current router.

The incoming interface (or RPF interface) on which the packet was received is never in the OIL. Therefore, the multicast packet is never forwarded back out of the RPF interface.

https://t.me/learningnets

Multicast Scoping

Most multicast applications include the ability to define an initial TTL value other than the default 255. This leads to a common method for bounding the multicast traffic to within a certain well-defined part of the network.



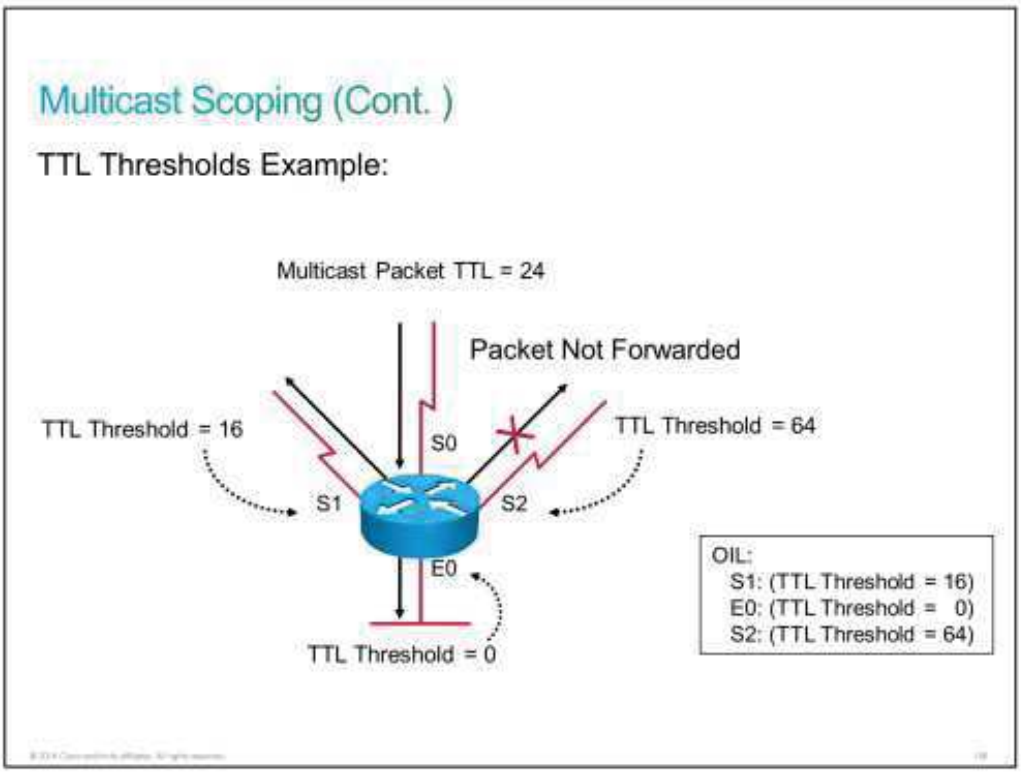
Sometimes people want to define the delivery boundaries for certain multicast traffic. They may not want all of their multicast traffic to be heard all over their network or outside of their network boundaries. Another possibility is that they may not want some external multicast traffic to enter their network. To achieve this goal, they may use TTL thresholds.

TTL thresholds may be set on a multicast router interface to limit the forwarding of multicast traffic to outgoing packets with TTLs greater than the TTL threshold. A TTL threshold of 0 implies that no threshold has been set.

As a multicast packet arrives, the TTL is decremented by 1. If the resulting TTL is less than or equal to 0, it is dropped. If a multicast packet is forwarded out of an interface with a TTL threshold configured, then its TTL is checked against the configured TTL threshold.

If a packet TTL is less than or equal to the specified TTL threshold, the packet is not forwarded out of the interface. If a packet TTL is greater than the specified TTL threshold, the packet is forwarded out of the interface. Typically, TTL thresholds are set on the boundary multicast routers to make sure traffic does not cross where it should not.

https://t.me/learningnets



The figure shows an example of multicast scoping using TTL thresholds.

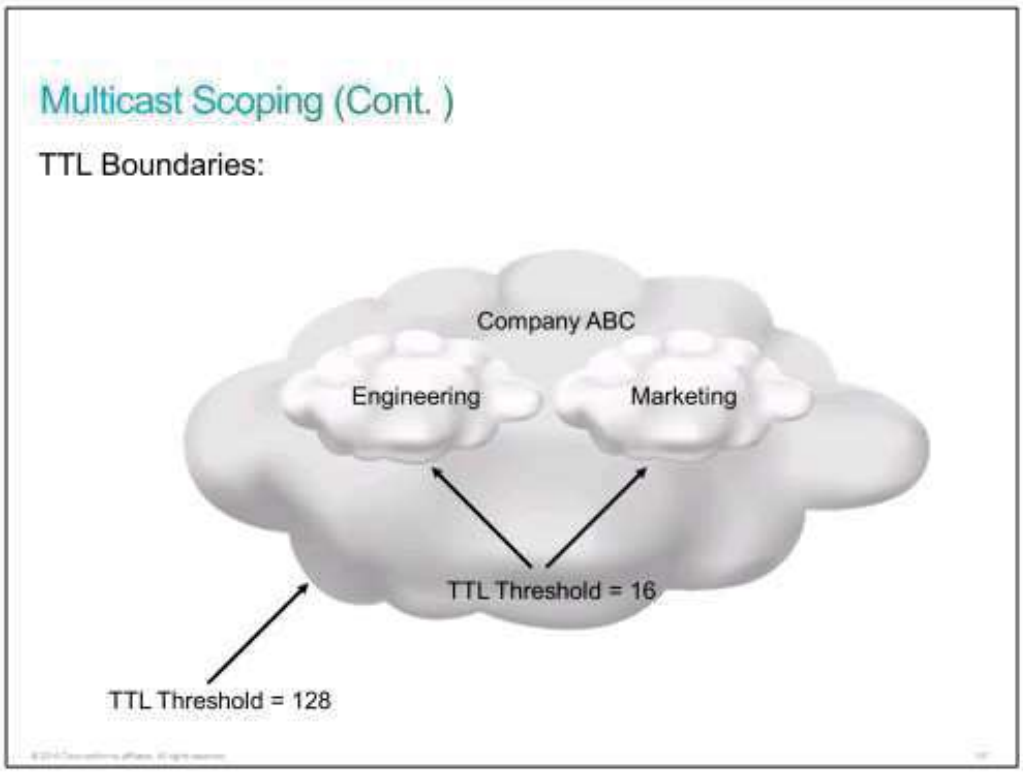
Interface Serial1 has a TTL threshold that is set to 16. Interface Serial2 has a TTL threshold that is set to 64. Interface Ethernet0 has a TTL threshold that is set to 0. This configuration means that the latter interface does not have any TTL threshold set.

The router receives a multicast packet with TTL 24 on interface Serial0. The TTL is first decremented by the normal router IP packet processing to 23. Before sending the packet on all interfaces in the OIL, the router checks for TTL thresholds that are set on those interfaces.

Because the TTL threshold on interfaces Serial1 (TTL threshold = 16) and Ethernet0 (TTL threshold = 0) is less than the TTL of the multicast packet, the packet is forwarded.

Interface Serial2 has a TTL threshold that is set to 64, which is greater than the TTL of the multicast packet, so the packet is not forwarded over interface Serial2.

https://t.me/learningnets



In the figure, the engineering and marketing departments may prevent department-related multicast traffic from leaving their network by using a TTL of 16 for their multicast sessions. Therefore, their multicast sources have to send all the multicast traffic with a TTL less than or equal to 16. To achieve this, you would configure a TTL threshold of 16 on the boundary router between the engineering and marketing departments. Any multicast packets with a TTL value less than or equal to 16 will be dropped, in order to prevent the multicast packets from reaching the other department.

Similarly, company ABC may prevent private multicast traffic from leaving its network by using a TTL of 128 for its multicast sessions.

An example of typical TTL values used for controlling multicast scope follows:

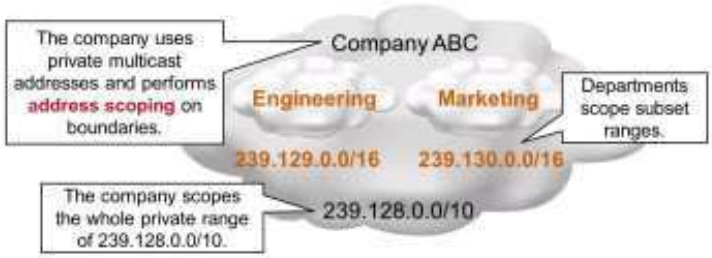
Scope	Initial TTL value
Local segment	1
Site, department, or division	16
Enterprise	128
Global	255

https://t.me/learningnets

Multicast Scoping (Cont.)

TTL scoping characteristics:

- TTL scoping depends on the settings of the TTL, which are sometimes unknown or unpredictable.
- An alternative is address scoping, which allows multicast boundaries to be established per group address.
- Traffic that does not match the address is not accepted on an incoming interface or not forwarded to an outgoing interface.



Although TTL scoping seems easy to implement, there are many problems with it. If you use TTL scoping with broadcast-and-prune multicast protocols, the router—which discards the multicast packet—will not be able to prune any upstream sources. It is also impossible to configure overlapping zones with TTL scoping.

Address scoping allows the multicast boundaries to be established per multicast group address, and is much more flexible than TTL scoping. The multicast traffic that does not match the address within a scope is dropped on an incoming interface, as well as on an outgoing interface.

If you use address scoping for overlapping zones, you must use separate address spaces within these zones, which makes the administration of multicast addresses difficult.

In the figure, the company uses the private multicast address range 239.128.0.0/10 and performs address scoping on its boundaries. Departments within the company use the subset of that range. The engineering and marketing departments use the subset ranges 239.129.0.0/16 and 239.130.0.0/16, respectively.

https://t.me/learningnets

Summary

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

Summary

- Multicast sends a single packet to multiple receivers, thus reducing network load closer to the source.
- Two of the most common multicast application models are one-to-many and many-to-many.
- Multicast reduces load on the network, but adds complexity when compared with unicast.
- Applications may learn about the sessions in several ways:
  - By joining a well-known group.
  - By using directory services.
  - By using a URL to launch an application.
- Multicast IP addresses use the Class D address space.
- SDR is a session description protocol and transport mechanism.
- In a multicast network, there are sources, receivers and routers.

Summary (Cont.)

- A multicast network learns about multicast group members and multicast streams, and builds an appropriate distribution tree between sources and receivers.
- Sources are the source of multicast streams, receivers are the destination.
- Various multicast routing protocols are used inside the multicast network.
- The multicast network performs RPF checks to prevent multicast traffic loops.
- Address scoping or TTL scoping is used to restrain multicast traffic to specific network segments.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Lesson 2: Defining Multicast Distribution Trees and Forwarding

Overview

Multicast traffic forwarding is accomplished by multicast-capable routers. These routers create distribution trees that control the IP multicast traffic path in order to deliver traffic to all receivers. Multicast traffic, flows from the source to the multicast group over a distribution tree that connects all of the sources to all of the receivers in the group.

Key concepts in IP multicast include an IP multicast group address, a multicast distribution tree, and receiver-driven tree creation. Sources and receivers use an IP multicast group address to send and receive content. Sources use the group address as the IP destination address in their data packets. Receivers use this group address to inform the network that they are interested in receiving the packets destined to this multicast group. The IPv4 receivers use IGMP message to join a group.

Once the receivers join a particular IP multicast group, a multicast distribution tree is constructed for that group. Most widely used protocol for this activity is PIM. Data packets sent from senders to a multicast group reaches to all receivers by using multicast distribution trees.

This lesson presents an overview of the techniques that are involved in building of multicast distribution trees and gives an overview of IP multicast protocols. In this lesson you will learn about IP multicast forwarding protocols, building of distribution trees, and reporting of group membership.

Upon completing this lesson, you will be able to:

- Discuss the role of the RPF check in building multicast distribution trees
- Discuss source or shortest-path distribution trees and shared distribution trees
- Discuss the (S,G) and (*,G) notations
- Identify multicast routing protocols
- Provide a high-level overview of PIM dense and sparse modes

https://t.me/learningnets

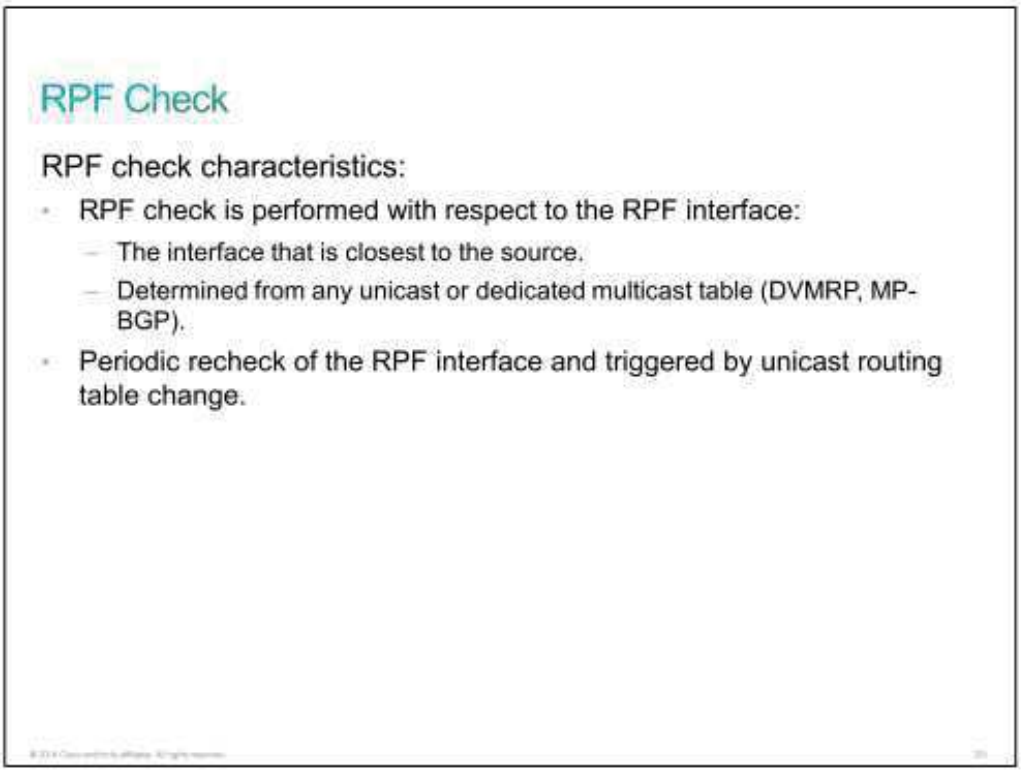
- Discuss intradomain multicast routing protocols
- Discuss interdomain multicast routing protocols
- Identify multicast high-availability options
- Discuss multicast NSF with SSO operations
- Discuss PIM triggered joins
- Provide an overview of IGMP
- Provide an overview of IGMPv1
- Provide an overview of IGMPv2
- Provide an overview of IGMPv3

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

RPF Check

Routers perform an RPF check to ensure that arriving multicast packets were received through the interface that is on the most direct path to the source which sent the packets.

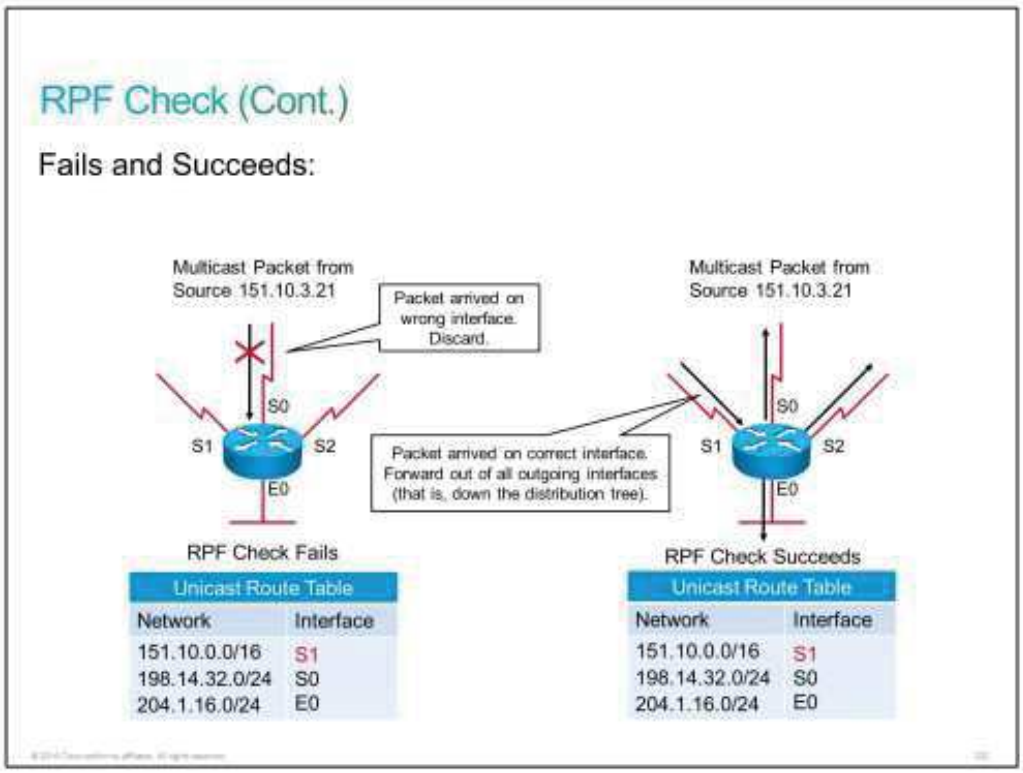


An RPF check is always performed regarding the incoming interface—the RPF interface. The RPF check will succeed if the incoming interface is the shortest path to the source.

The RPF interface is determined either by the underlying unicast routing protocol or the dedicated multicast routing protocol—for example, DVMRP or MP-BGP.

The multicast routing protocol relies on underlying unicast routing tables. A change in the unicast routing table immediately triggers RPF recheck on most modern routers.

https://t.me/learningnets



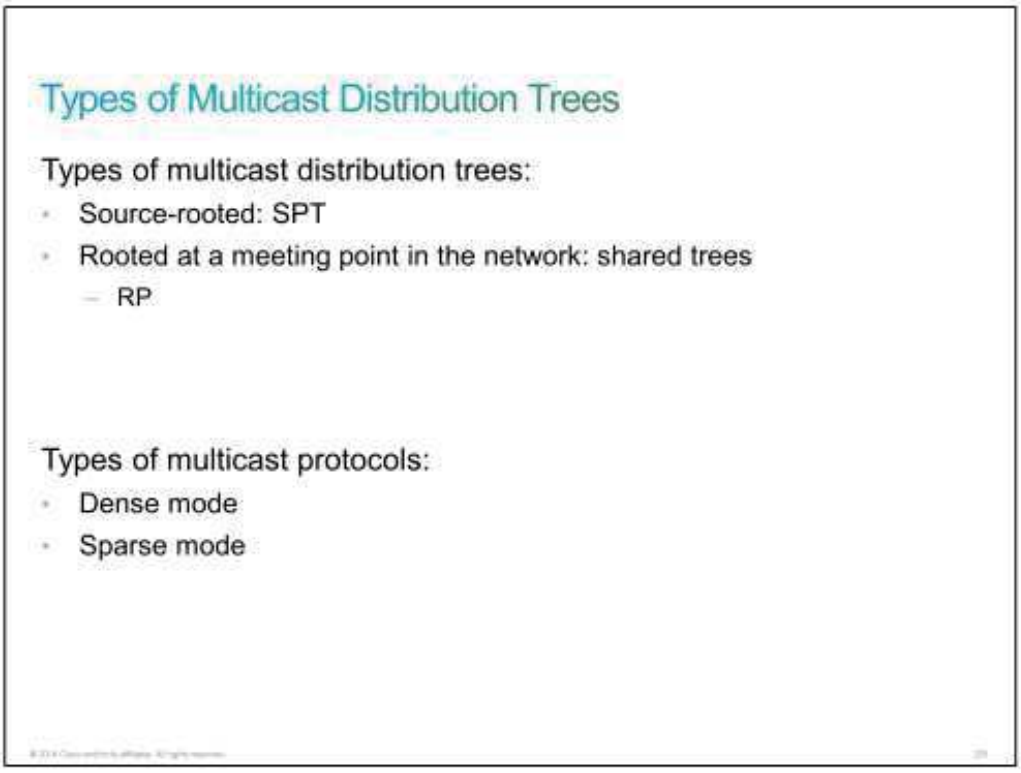
The left figure illustrates a failed RPF check. The router in the left figure receives a multicast packet from source 151.10.3.21 on interface Serial0. The router performs the RPF check by examining the unicast routing table. The unicast routing table indicates that interface Serial1 is the shortest path to the network 151.10.0.0/16. Because interface Serial0 is not the shortest path to the network from which the packet from the source 151.10.3.21 arrived, the RPF check fails, and the packet is discarded.

The right figure illustrates a successful RPF check. The router in the right figure receives a multicast packet from source 151.10.3.21 on interface Serial1. The router performs the RPF check by looking into the unicast routing table. The unicast routing table indicates that interface Serial1 is the shortest path to the network 151.10.0.0/16. Because interface Serial1 is the shortest path to the network from which the packet from the source 151.10.3.21 arrived, the RPF check succeeds. With this success, the packet is forwarded on every interface in the OIL. In the right figure, the OIL for the current multicast packet consists of interfaces Serial2 and Ethernet0, so the packet is forwarded on these interfaces.

https://t.me/learningnets

Types of Multicast Distribution Trees

Multicast distribution trees define the path which multicast traffic flows from the source to the receivers. There are two types of multicast distribution trees—source-rooted or SPTs, and shared trees.

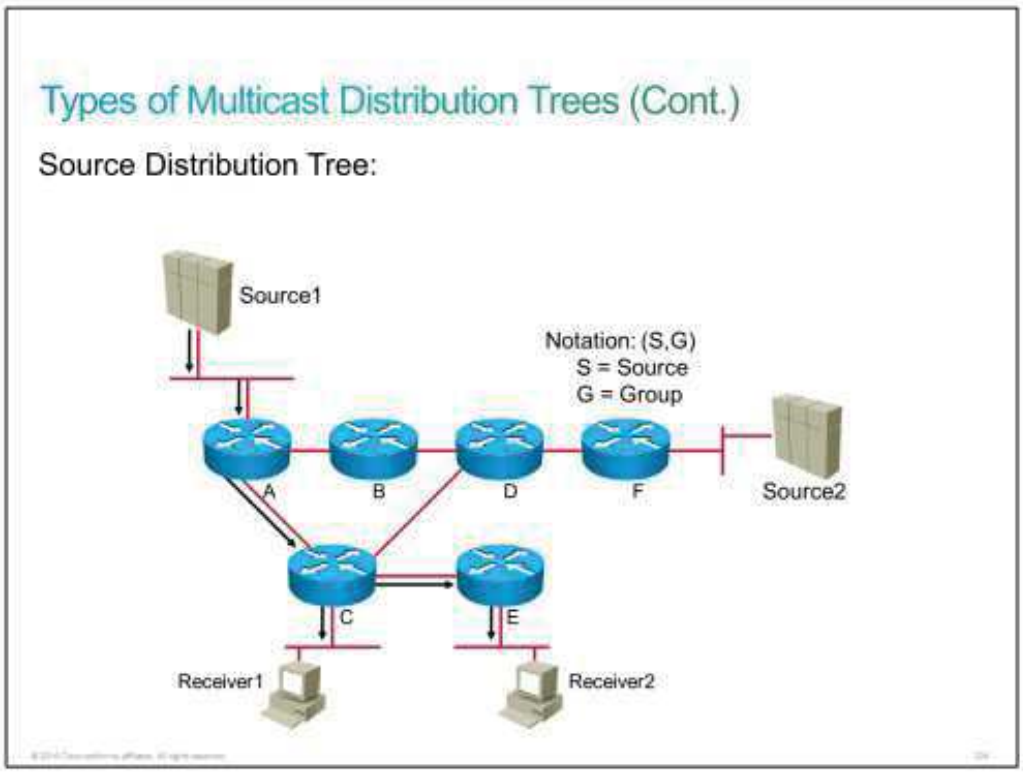


With a source-rooted tree, a separate tree is built for each source to all members of multicast group. Because the source-rooted tree takes a direct or shortest path from the source to its receivers. It is also called an SPT.

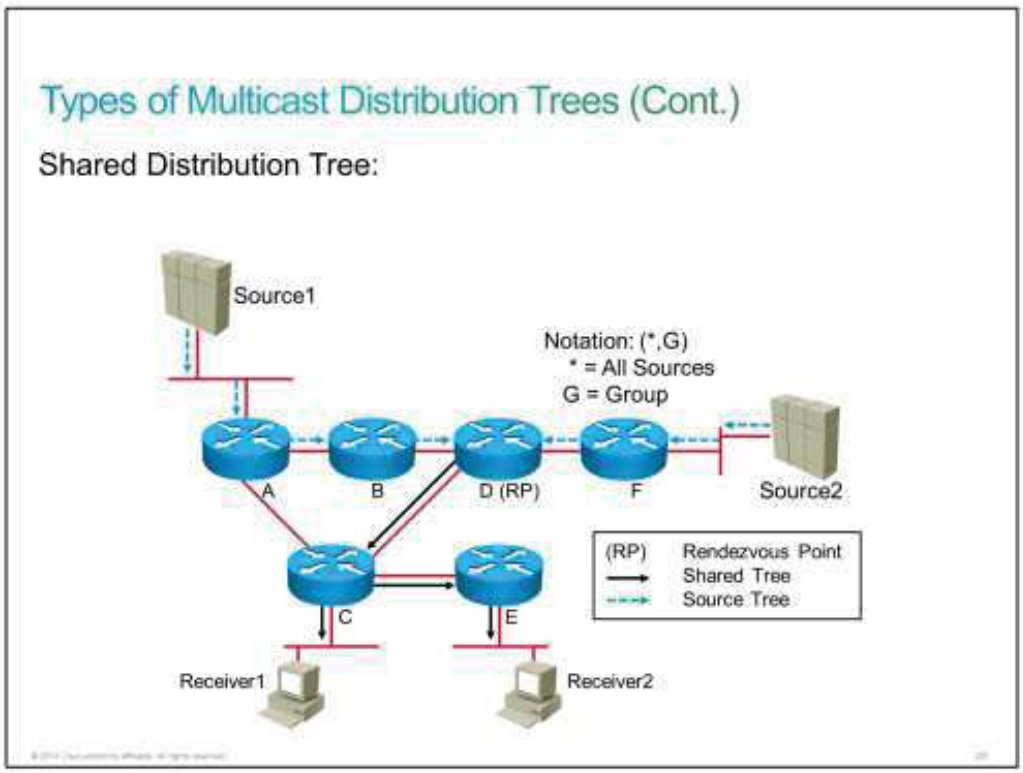
Shared-tree protocols create multicast forwarding paths that rely on a central core router. This router serves as a RP between multicast sources and destinations. Sources initially send their multicast packets to the RP, which, in turn, forwards data through a shared tree to the members of the group. A shared tree is less efficient than an SPT (because the paths between the source and the receivers are not necessarily the shortest), but it is less demanding on routers (in terms of memory and CPU usage).

There are two types of multicast routing protocols: dense mode protocols and sparse mode protocols. Dense mode protocols flood multicast traffic to all parts first, then prune the flows where there are no receivers by using a periodic flood-and-prune mechanism. Sparse mode protocols use an explicit join mechanism, distribution trees are built on demand by explicit tree join messages. Join messages are sent by routers which have directly connected receivers.

https://t.me/learningnets



The figure shows an SPT between Source1 and Receiver1 and 2. It is appropriately assumed that the path between the source and receivers over routers A, C, and E is the path with the lowest cost. Packets are forwarded down the SPT according to the pairs of source and group addresses. For this reason, the forwarding state that is associated with the SPT is referred to by the notation (S,G) (pronounced *S comma G*). In this notation, S is the IP address of the source, and G is the multicast group address. A separate SPT is built for every source S sending to group G.



 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

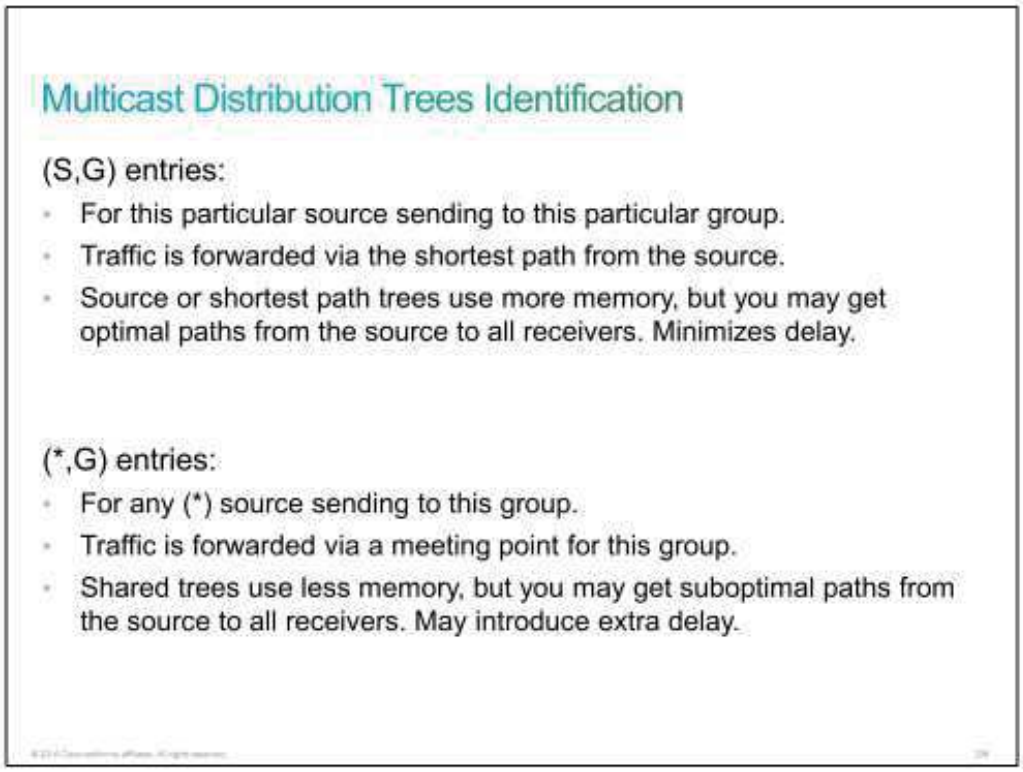
The figure shows a shared distribution tree. Router D is the root of this shared tree, which is built from router D to routers C and E toward Receiver1 and Receiver2. In PIM, the root of the shared tree is called an RP. Packets are forwarded down the shared distribution tree to the receivers. The notation (*,G) (pronounced *star comma G*) identifies the default forwarding state for the shared tree. The * symbol represents a wildcard entry, meaning that any source can be plugged into the notation, and G is the multicast group address.

The figure also shows traffic flow on two source-rooted trees in addition to the shared distribution tree. Source1 and Source2 are sending multicast packets toward an RP via the source-rooted trees. From the RP, the multicast packets are flowing via a shared distribution tree toward Receiver1 and Receiver2.

https://t.me/learningnets

Multicast Distribution Tree Identification

It is important to be aware of the key differences between multicast distribution trees for better understanding of which multicast routing protocol is deployed in the network.



The multicast forwarding entries that appear in multicast forwarding tables may be read in the following way:

- **(S,G)**: This notation indicates that a source S is sending to the group G.
- **(*,G)**: This notation indicates that any source (*) is sending to the group G. These entries reflect the shared tree, but are also created (in Cisco routers) for any existing (S,G) entry.

SPT (S,G) state entries use more router memory because there is an entry for each sender and group pair. However, because the traffic is sent over the optimal path to each receiver, the delay in packet delivery is minimized.

(*,G) state entries for shared distribution trees consume less router memory, but you may get suboptimal paths from a source to receivers, thus introducing an extra delay in packet delivery.

https://t.me/learningnets

Multicast Protocols Overview

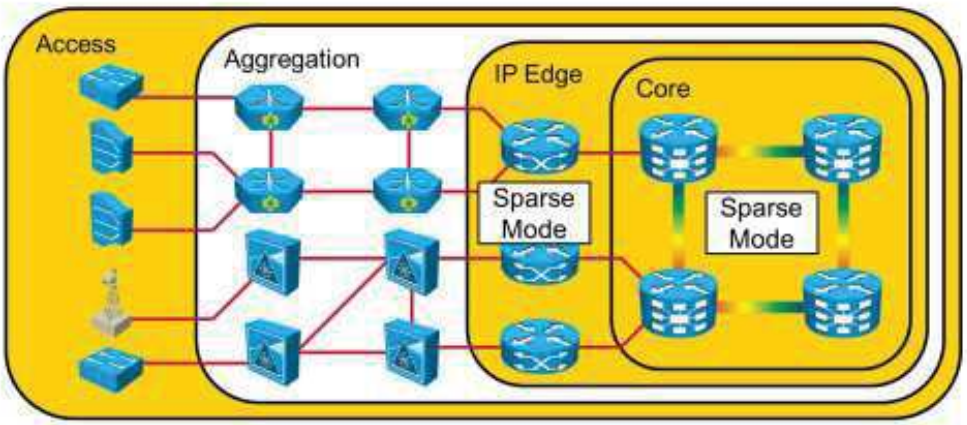
There are two types of multicast routing protocols: dense mode and sparse mode.

Multicast Protocols Overview

Multicast Protocols in the Cisco IP NGN Infrastructure Layer

- Dense mode (DVMRP, MOSPF, PIM-DM):
  - Uses the push model.
  - Initial traffic flooded to all branches of the distribution tree.
  - Branches without receivers get pruned.
  - Displays flood-and-prune behavior (typically every 3 minutes).
- Sparse mode (PIM-SM, CBT):
  - Uses the pull model (join behavior).
  - Branches without receivers never get the traffic.
  - Last-hop routers pull the traffic from the meeting point or from the source.

Multicast Protocols Overview (Cont.)



https://t.me/learningnets

Dense mode protocols use the push model, where multicast traffic is flooded throughout the network. Dense mode protocols are DVMRP, MOSPF, and PIM-DM. Dense mode protocols are not likely to be used in the service provider network.

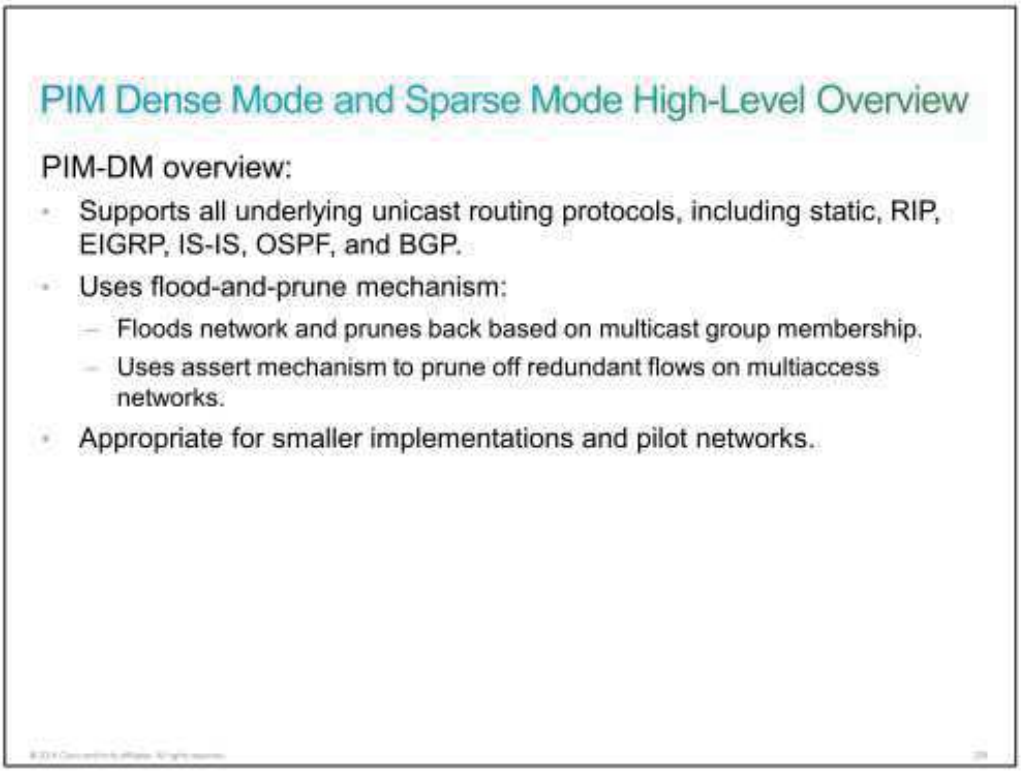
Sparse mode routing protocols are a better choice because they use the pull model. By using the pull model, the protocol enables multicast traffic to be sent only where it is requested. This kind of behavior may be described as explicit join behavior, because the source assumes that no receivers are interested unless the receivers explicitly join the group. Sparse mode protocols are PIM-SM and CBT. Sparse mode protocols maybe used in the access, aggregation, IP edge, and core Cisco IP NGN infrastructure layers of the service provider network.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

PIM Dense Mode and Sparse Mode High-Level Overview

PIM-DM is defined with RFC 3973. The objective of developing this protocol was to provide intradomain multicast routing independent of the underlying unicast routing protocol. Underlying routing can be implemented using static routes, RIP, EIGRP, IS-IS, OSPF, and BGP.



PIM-DM is similar to DVMRP because they both use the flood-and-prune mechanism. However, unlike DVMRP, PIM-DM does not have prebuilt truncated broadcast trees when using its own routing protocol. Instead, PIM-DM uses the RPF-check mechanism to accept incoming multicast traffic and then initially floods it to all PIM-DM neighbors. PIM-DM neighbors that do not have downstream receivers send back prune messages to shut off the flow of unwanted traffic. On multiaccess networks, PIM uses an assert mechanism to elect a designated forwarder for a particular source and to shut off redundant flows of traffic. In PIM-DM, the combination of prune messages and the assert mechanism combine to build SPTs.

PIM-DM is appropriate for small implementations and trial networks, and is interoperable with DVMRP.

https://t.me/learningnets

PIM Dense Mode and Sparse Mode High-Level Overview (Cont.)

PIM-SM Overview:

- Works with any of the underlying unicast routing protocols.
- Supports both source trees and shared trees.
- Based on an explicit pull model and uses an RP:
  - Senders are registered with RP by first-hop router.
  - Receivers are joined to the shared tree (rooted at the RP) by last-hop router.
- Appropriate for:
  - Large-scale deployment for both densely and sparsely populated groups in the enterprise.
  - Optimal choice for all production networks, regardless of size and membership density.
- Optimizations and derivatives:
  - BIDIR-PIM, SSM

PIM-SM is described in RFC 4601. Similar to PIM-DM, PIM-SM operates independently of underlying unicast protocols. PIM-SM uses shared distribution trees rooted at the RP, but it may also switch to the source-rooted distribution tree.

PIM-SM is based on an explicit pull model. Therefore, the traffic is forwarded only to those parts of the network that need it.

PIM-SM uses an RP to coordinate the forwarding of multicast traffic from a source to its receivers. Senders register with the RP and send a single copy of multicast data through it to the registered receivers. Group members are joined to the shared tree by their local designated router. A shared tree that is built this way is always rooted at the RP.

PIM-SM is appropriate for the wide-scale deployment of both densely and sparsely populated groups in the enterprise network. It is the optimal choice for all production networks, regardless of size and membership density.

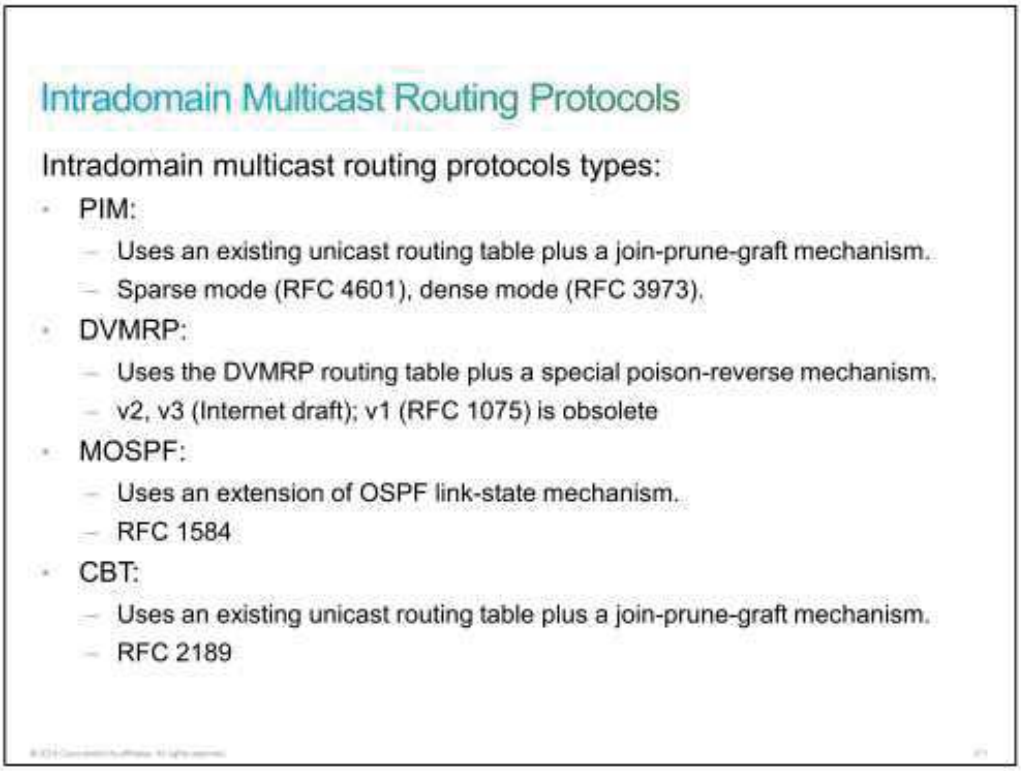
There are many optimizations and enhancements to PIM, including the following:

- BIDIR-PIM mode is designed for many-to-many applications.
- SSM is a variant of PIM-SM that only builds source-specific SPTs and does not need an active RP for source-specific groups. The address range 232.0.0.0/8 is used.

https://t.me/learningnets

Intradomain Multicast Routing Protocols

Multicast distribution trees are built in various ways, depending upon the multicast routing protocol employed.



The following protocols constitute the intradomain multicast routing protocols:

- **PIM** comes in two versions: PIM-SM and PIM-DM. Much of the effort of the IETF group that is working toward a multicast protocol is focused on PIM-SM. PIM uses the underlying unicast routing table (in any unicast routing protocol), in addition to the following messages:
  - **Join:** Routers add their interfaces or send PIM join messages upstream to establish themselves as branches of the tree when they have interested receivers downstream.
  - **Prune:** Routers prune their interfaces or send PIM prune messages upstream to remove them from the distribution tree when they no longer have interested receivers downstream.
  - **Graft:** Routers send PIM graft messages upstream when they have a pruned interface and have already sent PIM prune messages upstream, but have received an IGMP host report for the same group that was pruned. Routers must reestablish themselves as branches of the distribution tree because they have newly interested receivers attached to their routes.
- **DVMRP** evolved from version 1 to version 3. DVMRP version 1 (RFC 1075) is obsolete and is never used. DVMRP version 2 is an old Internet draft and is the implementation that was used throughout the multicast backbone. DVMRP version 3 is the current Internet draft, although most vendors have not completely implemented it. DVMRP uses a special RIP-like multicast routing table in addition to the following:
  - **Poison reverse:** The special metric of infinity (32) and the originally received metric are used to signal that the router must be placed on the distribution tree for the source network.

https://t.me/learningnets

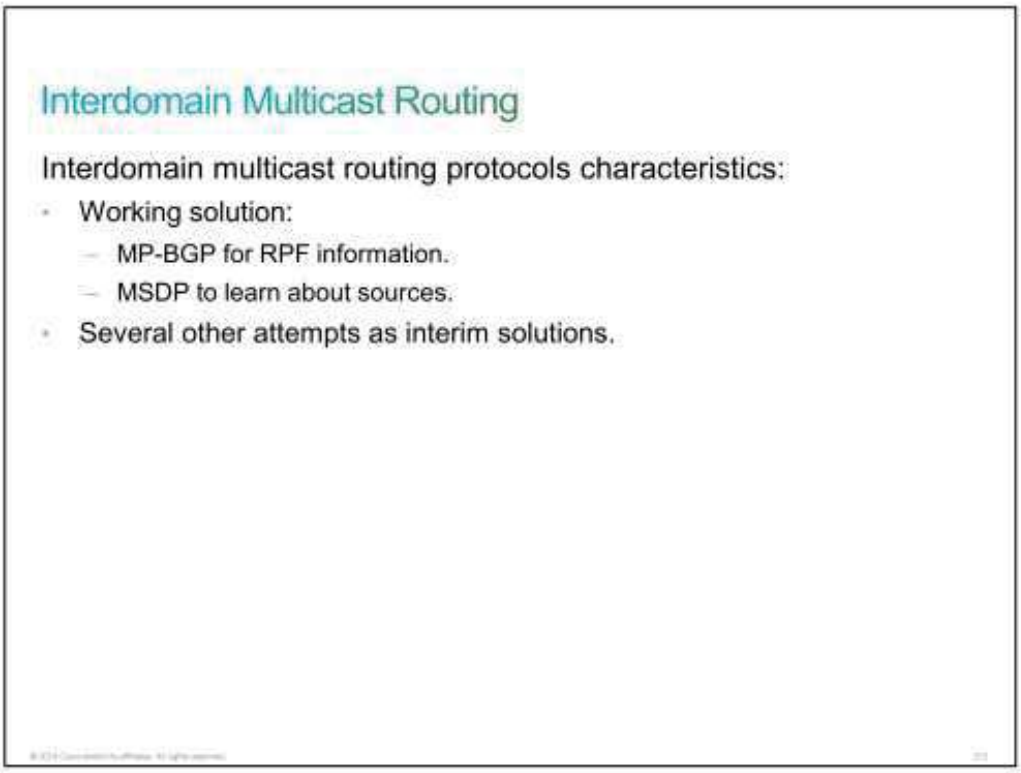
- **Prunes and grafts:** Routers send prunes and grafts up the distribution tree, just as they do in PIM-DM.

- **MOSPF** is defined in RFC 1584. Because of the nature of OSPF, it is not scalable when the number of groups in an area increases or when multicast groups are very active. MOSPF uses the underlying OSPF unicast routing protocol link-state advertisements (LSAs) to advertise the existence of directly connected receivers. This information is then used to build (S,G) trees. Each router maintains an up-to-date image of the topology of the entire network (more explicitly, of an entire area).
- **CBT** is defined in RFC 2189, but it does not have significant usage or industry support, nor is it a primary focus of the IETF working groups. CBT uses the underlying unicast routing table and the join-prune-graft mechanisms (much like PIM-SM).

https://t.me/learningnets

Interdomain Multicast Routing Protocols

MP-BGP is the only working solution that solves a part of the interdomain multicast routing challenge.



MP-BGP allows autonomous systems to exchange multicast RPF information as MP-BGP multicast NLRI. Actually, MP-BGP is only an extension of the BGP unicast routing protocol.

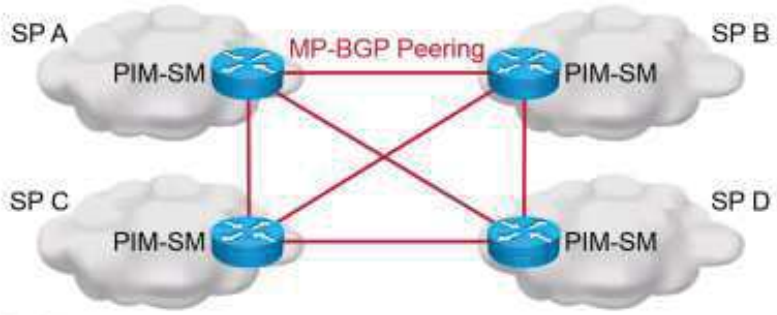
Because MP-BGP does not build multicast distribution trees, an additional protocol must be used to augment PIM-SM. That protocol is the MSDP, which is used to distribute information about the existence of active sources inside one PIM-SM domain to other PIM-SM domains. When routers inside one PIM-SM domain have learned about active sources in other PIM-SM domains, (S,G) joins are sent between domains to interconnect sources and receivers in distant domains via interdomain branches of the SPT.

https://t.me/learningnets

Interdomain Multicast Routing Protocols (Cont.)

MP-BGP:

- PIM-SM used within and across domains.
- MP-BGP used between domains for source network information (RPF checks).

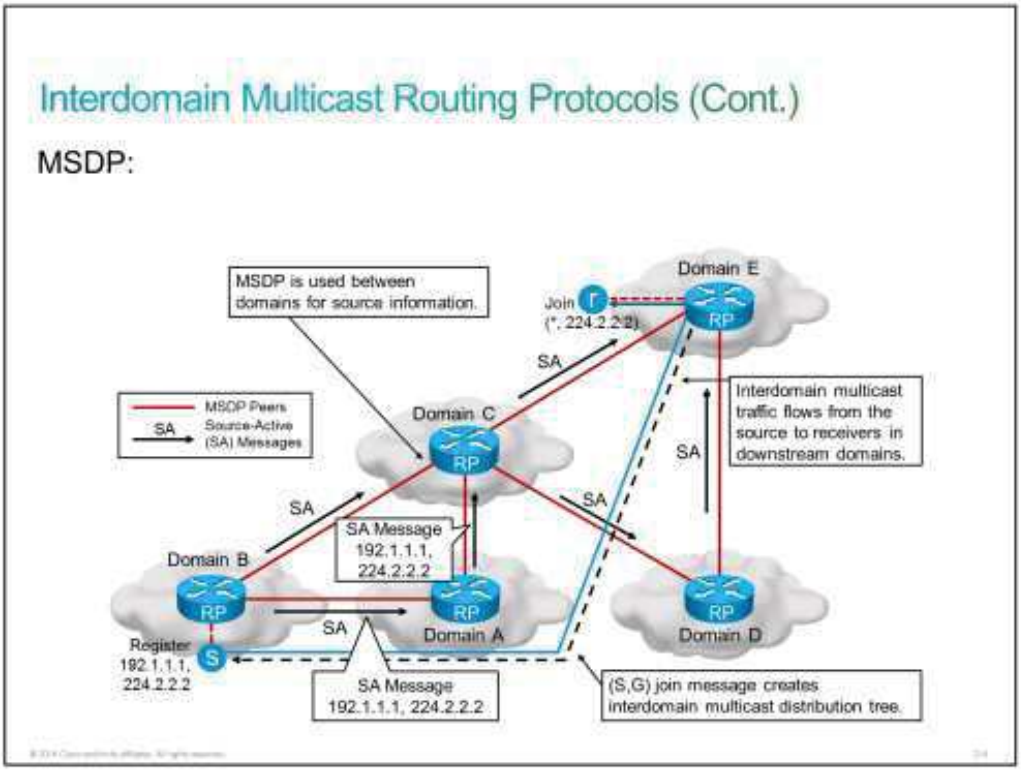


The figure illustrates the first component of the interim solution, which is currently the only working solution for interdomain multicast routing. Traditional BGP that was used for peering between domains was enhanced with multiprotocol extensions to allow other types of NLRI to be exchanged beyond just IPv4 unicast NLRI. This enhancement allows multicast RPF NLRI to also be exchanged.

Routers on the borders of domains thus establish MP-BGP peering and exchange a separate set of multicast RPF NLRI. MP-BGP uses this multicast RPF NLRI information to build and maintain a multicast BGP table, which contains the unicast prefixes used for RFP checking. The multicast BGP table is then used by the routers to perform RPF checks on multicast traffic arriving from other autonomous systems. Because PIM-SM is using this RPF information to determine which way to send (S,G) joins, it also affects how the interdomain branches of SPTs are built. Having two separate sets of NLRI (one for unicast routing and one for multicast RPF checking) allows for incongruent unicast and multicast topologies. Interdomain unicast traffic paths may differ from interdomain multicast traffic paths.

Using MP-BGP, service providers can also implement a different routing policy for unicast traffic versus multicast traffic.

https://t.me/learningnets



The second component that is needed for interdomain multicast routing is MSDP. This protocol works with PIM-SM and allows RPs in one domain to announce their sources to other domains using source-active messages. Routers that act as RPs establish MSDP peering and exchange information on the active sources and groups to which they are sending.

When information is received in a domain with active receivers for the group, an appropriate (S,G) join is initiated (by the RP of the domain) across domains toward the source. With this mechanism, downstream domains pull the traffic from the sources in upstream domains. When (S,G) joins travel toward the source, the distribution tree is built across domains. As soon as the distribution tree is built across domains, the traffic from the active source starts flowing to downstream receivers.

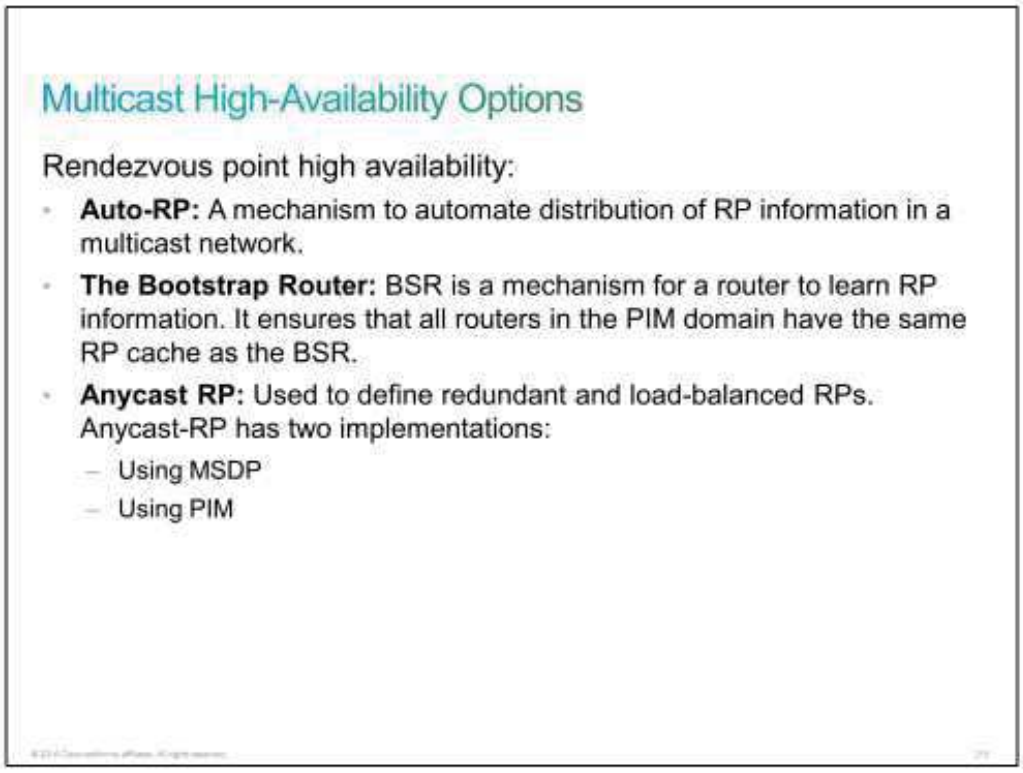
The only working solution today for interdomain multicast is PIM-SM with MP-BGP and MSDP.

In the figure, the 192.1.1.1 source in domain B starts sending multicast traffic for the 224.2.2.2 group, the first-hop router in domain B sends a register message to the RP in domain B, the RP in domain B then sends a source-active message to all its MSDP peers.

https://t.me/learningnets

Multicast High-Availability Options

Multicast high-availability options reduce the reconvergence time of the multicast control plane to a level that is transparent to most multicast-based applications.

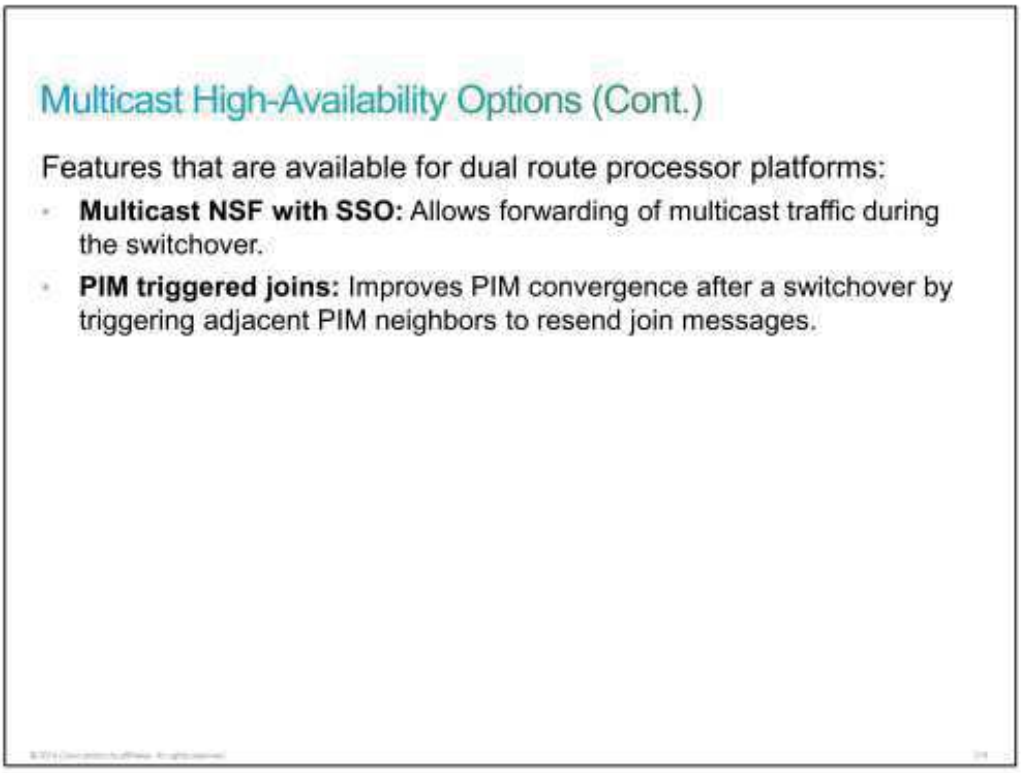


The first aspect of high availability for multicast is redundancy of rendezvous points. The following mechanisms can be used to achieve redundancy of rendezvous points:

- **Auto-Rendezvous Point (Auto-RP)** is a mechanism to automate distribution of RP information in a multicast network. The Auto-RP mechanism operates using two basic components, the candidate RPs and the RP mapping agents.
  - **Candidate RPs** advertise their willingness to be an RP via "RP-announcement" messages. These messages are periodically sent to a reserved well-known group 224.0.1.39 (CISCO-RP-ANNOUNCE)
  - **RP mapping agents** join group 224.0.1.39 and map the RPs to the associated groups. The RP mapping agents advertise the authoritative RP-mappings to another well-known group address 224.0.1.40 (CISCO-RP-DISCOVERY). All PIM routers join 224.0.1.40 and store the RP-mappings in their private cache.
- **Bootstrap router** is a mechanism for a router to learn RP information. It ensures that all routers in the PIM domain have the same RP cache as the BSR. It is possible to configure the BSR to help select an RP set from BSR candidate RPs. The function of the BSR is to broadcast the RP set to all routers in the domain.
- **Anycast RP** is used to define redundant and load-balanced RPs. Anycast-RP has two implementations:
  - Using MSDP described in RFC3446.
  - Using PIM described in RFC 4610.

https://t.me/learningnets

Multicast high-availability options reduce the reconvergence time of the multicast control plane to a level that is transparent to most multicast-based applications.



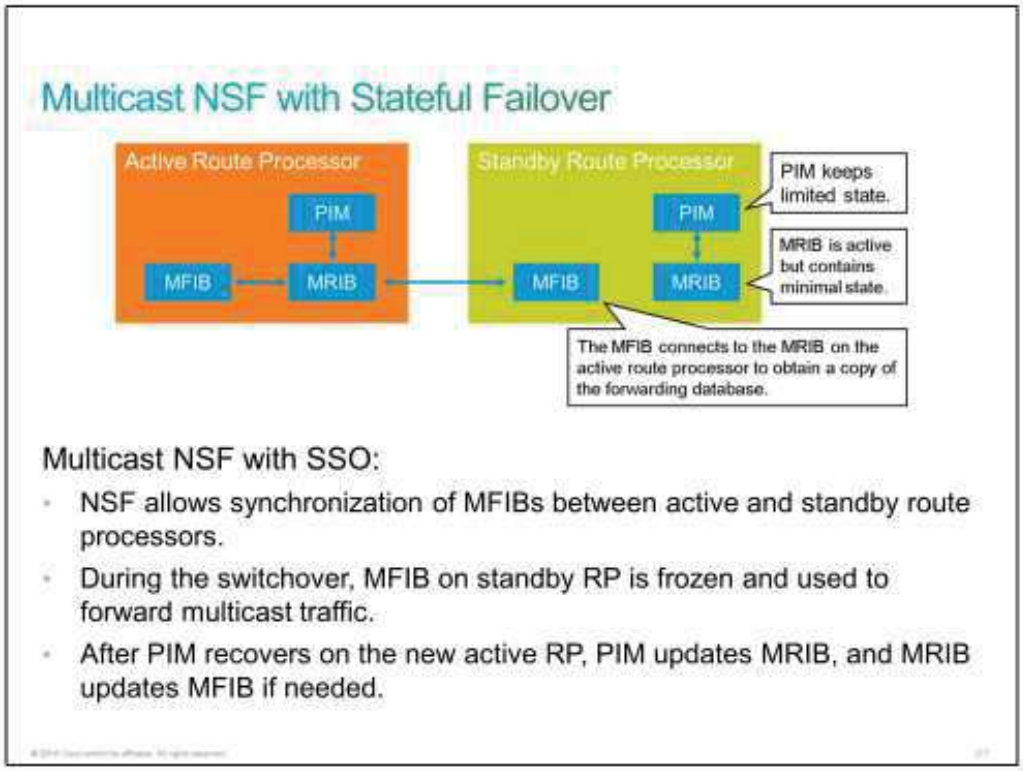
Two additional options are used for multicast high availability on platforms that support dual RPs:

- **Multicast Nonstop Forwarding with Stateful Switchover**: Multicast NSF with SSO allows forwarding of multicast traffic during a switchover.
- **PIM triggered joins**: Improves PIM convergence after a switchover by triggering PIM neighbors to resend join messages for joined groups.

https://t.me/learningnets

Multicast NSF with Stateful Failover

The Cisco Non-Stop Forwarding feature for multicast enhances high availability of multicast packet forwarding.



NSF prevents hardware or software failures on the control plane from disrupting the forwarding of existing packet flows through the router.

With the multicast NSF, the contents of the MFIB are frozen during a control plane failure. Afterward, PIM attempts to recover normal protocol processing and state before the neighboring routers time out the PIM hello neighbor adjacency for the problematic router. Routes in MFIB are marked as stale after entering NSF, and traffic continues to be forwarded (based on those routes) until NSF completion. Upon completion, the MRIB notifies the MFIB, and MFIB performs a mark-and-sweep to synchronize MFIB with the current MRIB route info.

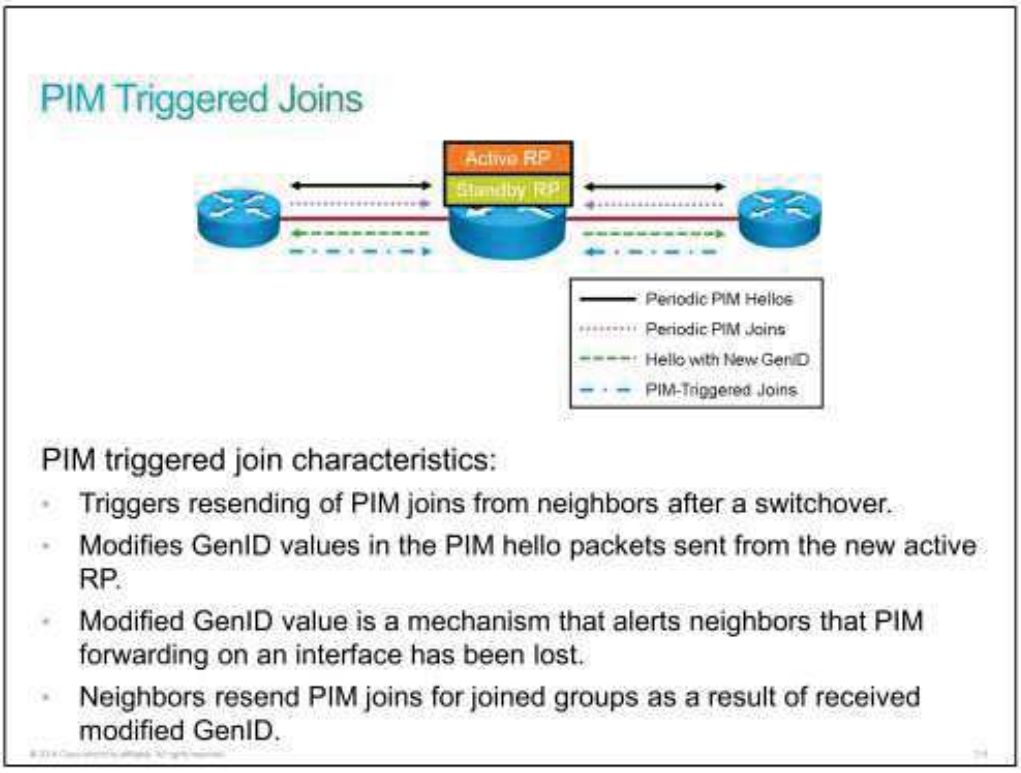
To enable multicast NSF on the Cisco IOS XR Software router, use the following commands:

```
multicast-routing
 nsf
```


https://t.me/learningnets

PIM Triggered Joins

The PIM triggered joins feature is a high-availability multicast enhancement that improves the reconvergence of multicast routing table after an RP switchover.



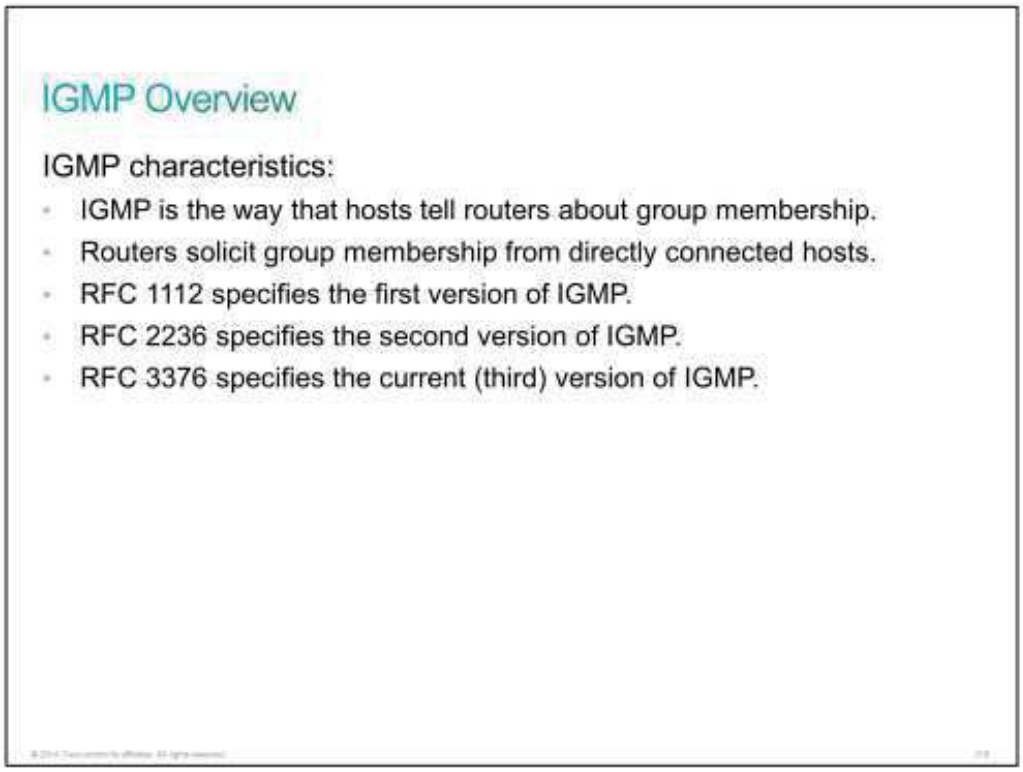
In the event of an RP switchover, this feature utilizes the PIM-SM GenID value as a mechanism to trigger adjacent PIM neighbors on an interface to send PIM join messages for all (*, G) and (S, G) entries that use that interface as an RPF interface, immediately reestablishing those states on the newly active RP. A GenID is a randomly generated 32-bit value that is regenerated each time PIM forwarding is started or restarted on an interface.

After an RP switchover, all instances of PIM that are running on the newly active RP will modify the value of the GenID that is included in PIM hello messages sent to adjacent PIM neighbors. When an adjacent PIM neighbor receives a PIM hello message on an interface with a new GenID, the PIM neighbor will process the modified GenID as an indication that the PIM neighbor has gone down. A modified GenID, therefore, is a mechanism to alert all adjacent PIM neighbors that PIM forwarding on that interface has been lost, which then triggers adjacent PIM neighbors to send PIM joins for all (*, G) and (S, G) states that use that interface as an RPF interface.

https://t.me/learningnets

IGMP Overview

The primary purpose of the IGMP is to permit hosts to communicate their desire to receive IPv4 multicast traffic to the IP multicast router on the local network. This action, in turn, permits the IP multicast router to join the specified multicast group and to begin forwarding the multicast traffic onto the network segment.



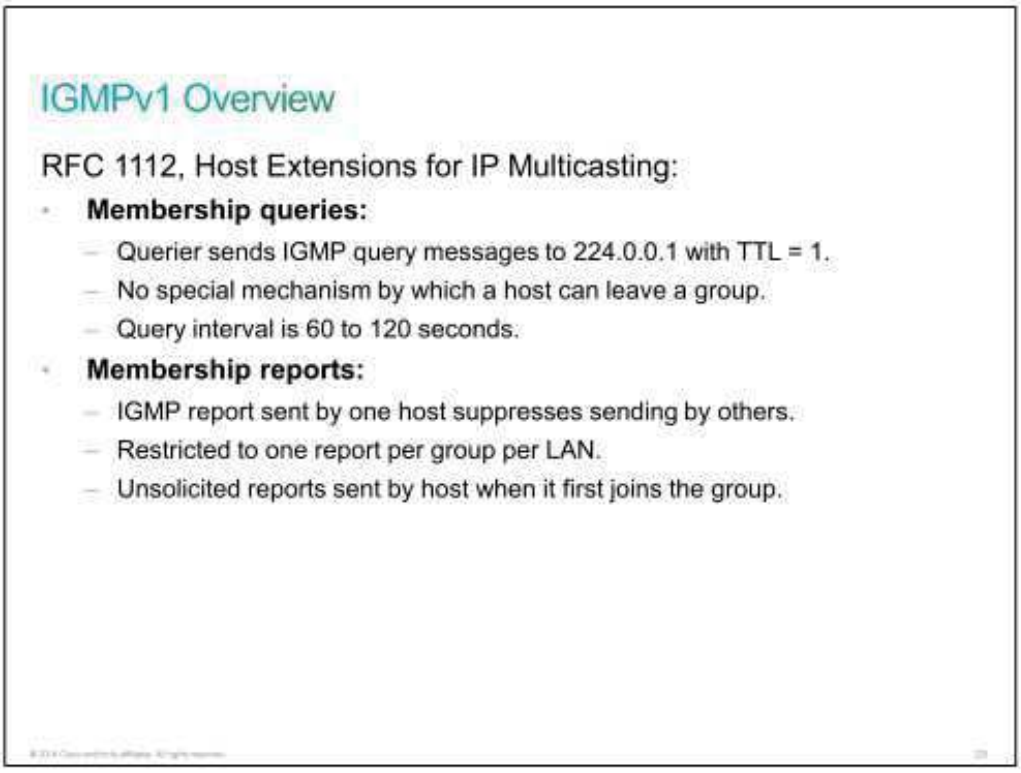
The initial specification for IGMPv1 was documented in RFC 1112. From that time, many problems and limitations with IGMPv1 have been discovered, which has led to the development of the IGMPv2 specifications. IGMPv2 is defined in RFC 2236.

Even before IGMPv2, work on the next generation of IGMP—IGMPv3 (RFC 3376)—had already begun.

https://t.me/learningnets

IGMPv1 Overview

RFC 1112 specifies IGMP as a protocol used by IP hosts to report multicast group membership to their first-hop multicast routers.



Multicast routers periodically (usually every 60 to 120 seconds) send membership queries to the all-hosts multicast address (224.0.0.1) to solicit the multicast groups that are active on the local network.

Hosts that want to receive specific multicast group traffic send membership reports. Membership reports, with a TTL of 1, are sent to the multicast address of the group from which the host wants to receive traffic. Hosts either send reports asynchronously (when they want to first join a group—unsolicited reports) or in response to membership queries. In the latter case, the response is used to maintain the group in an active state so that traffic for the group continues to be forwarded to the network segment.

In IGMPv1, there is no election of an IGMP querier. If more than one router on the segment exists, all the routers send periodic IGMP queries. IGMPv1 has no special mechanism by which the hosts can leave the group. If the hosts are no longer interested in receiving multicast packets for a particular group, they simply do not reply to the IGMP query packets sent from the router. The router continues sending query packets. If the router does not hear a response in three IGMP queries, the group times out and the router stops sending multicast packets on the segment for the group. If the host later wants to receive multicast packets after the timeout period, the host simply sends a new IGMP join to the router, and the router begins to forward the multicast packets again.

After a multicast router sends a membership query, there may be many hosts that are interested in receiving traffic from specified multicast groups. To suppress a membership report storm from all group members, a report suppression mechanism is used among group members. Report suppression saves CPU time and bandwidth on all systems.

Because membership query and report packets have only local significance, the TTL of these packets is always set to 1. TTL also has to be set to 1 because forwarding of membership reports from a local subnet may cause confusion on other subnets.

https://t.me/learningnets

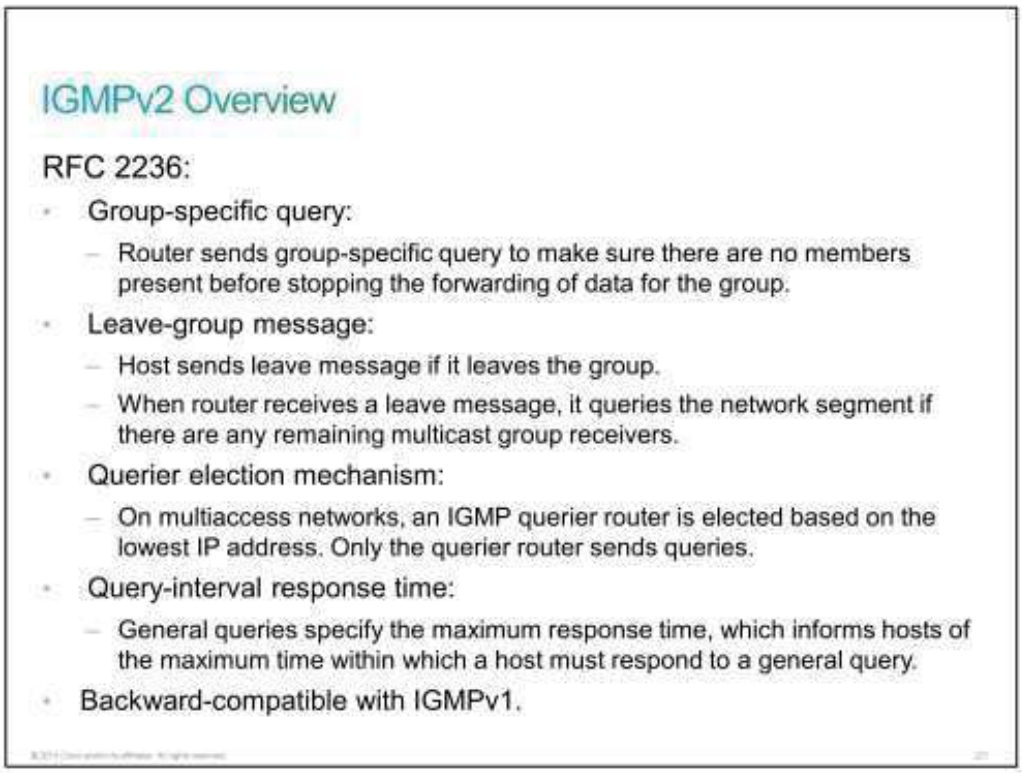
If multicast traffic is forwarded on a local segment, there has to be at least one active member of that multicast group on a local segment.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

IGMPv2 Overview

Because of some limitations that were discovered in IGMPv1, work was begun on IGMPv2 in an attempt to remove these limitations. Most of the changes between IGMPv1 and IGMPv2 were made primarily to address the issues of leave and join latencies, in addition to address ambiguities in the original protocol specification.



Some important changes were made in revising IGMPv1 to IGMPv2:

- Group-specific queries
- Leave-group message
- Querier election mechanism
- Query-interval response time

A group-specific query that was added in IGMPv2 allows the router to query its members only in a single group instead of all groups. This action is an optimized way to quickly find out if any members are left in a group without asking all groups for a report. The difference between the group-specific query and the membership query is that a membership query is multicast to the all-hosts address (224.0.0.1), whereas a group-specific query for group G is multicast to the group G multicast address.

A leave-group message allows hosts to tell the router that they are leaving the group. This information reduces the leave latency for the group on the segment when the member who is leaving is the last member of the group.

https://t.me/learningnets

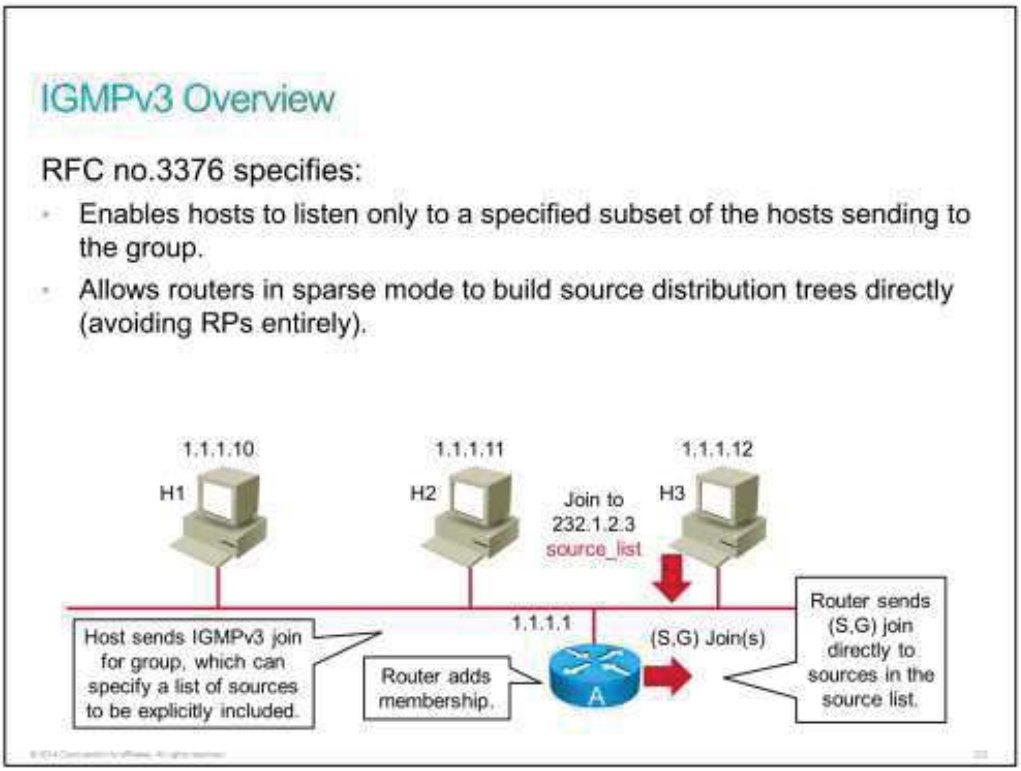
Unlike IGMPv1, IGMPv2 has a querier election mechanism. The lowest unicast IP address of the IGMPv2-capable routers will be elected as the querier. By default, all IGMP routers are initialized as queriers, but must immediately relinquish that role if a lower-IP-address query is heard on the same segment. The query-interval response time has been added to control the burstiness of reports. This time is set in queries to convey to the members how much time they have to respond to a query with a report. IGMPv2 is backward-compatible with IGMPv1.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

IGMPv3 Overview

The main intention of IGMPv3 is to allow hosts to indicate that they only want to receive traffic from a particular source within a multicast group. This enhancement will allow routers in sparse mode to build a source distribution tree directly, thereby avoiding the RP.



The figure shows IGMPv3 operation. Host H3 sends a join message with an explicit request for joins to sources in the source list. The router then uses a source list to issue the correct (S,G) joins rather than sending (*,G) joins to RPs. When used in combination with SSM, which uses a 232.0.0.0/8 address range, the router must not initiate any (*,G) join for those groups.

https://t.me/learningnets

Summary

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

Summary

- RPF is used in multicast to ensure loop-free routing.
- Multicast distribution trees can be rooted at the source or at rendez-vous point.
- There are two notations in multicast for source – group pairings:
  - (S,G) source routed trees group one particular source with one group.
  - (*,G) RP routed trees group any source with a particular group.
- Different protocols can be used for routing, but PIM is most widely used.
- PIM can be used in dense or in sparse mode.
- PIM among other protocols is used for intradomain routing.
- Interdomain routing is accomplished with MP-BGP or MSDP.

Summary (Cont.)

- In addition to regular high availability enhancements, multicast requires RP redundancy.
- NSF and SSO also support multicast forwarding.
- Modified GenID value after a switchover triggers resending of PIM joins from neighbors.
- Between receivers and designated routers IGMP is used to join groups.
- Multicast routers periodically send membership queries to the all-hosts multicast address.
- A group-specific query that was added in IGMPv2 allows the router to query its members only in a single group instead of all groups.
- The main intention of IGMPv3 is to allow hosts to indicate that they only want to receive traffic from a particular source within a multicast group.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Lesson 3: Defining Multicast on the LAN

Overview

To forward IP multicast on a LAN, the IP multicast Layer 3 address needs to be translated to a Layer 2 MAC address. But first, the multicast receivers need to signal to the gateways and switches which IP multicast traffic they need. The default behavior for a switch is to flood the IP multicast traffic out to all ports. The redundant IP multicast traffic can cause problems for the switch, as it can use many CPU cycles on the switch and congest the network.

This lesson represents an introduction to IP multicast issues on a data link layer. It explains the methods of mapping network layer multicast addresses to data link layer addresses, and it lists the mechanisms and protocols for constraining multicast streams in a LAN environment. The learner will identify the multicast-related issues that are associated with the reality of the data link layer versus the logical network layer.

Because of its unrestricted forwarding, IP multicast traffic can consume many resources on a switch.

This lesson explains IGMP snooping and how it can constrain the multicast flood of traffic in a switched environment. It explains how to configure and troubleshoot IGMP snooping, and identifies the most relevant implementation issues.

Upon completing this lesson, you will be able to:

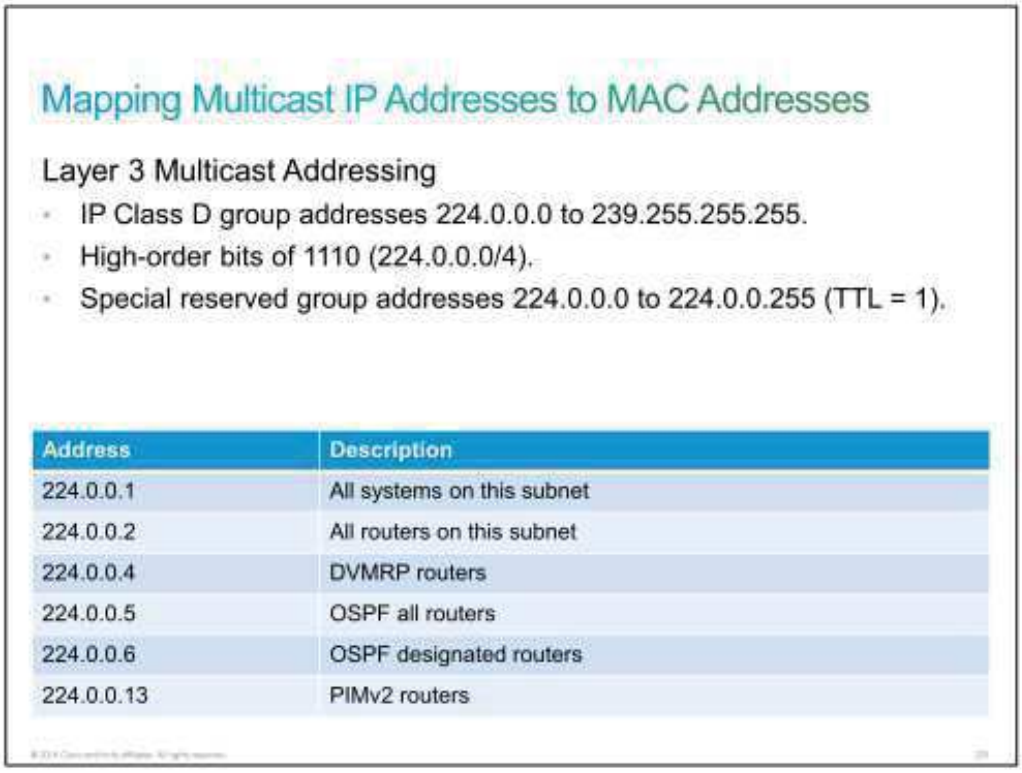
- Describe the algorithms that map IP multicast addresses to MAC addresses
- Discuss how a Layer 2 switch forwards multicast traffic
- Describe how to enable and configure IGMP
- Discuss IGMP join-group and static-group configuration options
- Discuss the IGMPv3 host stack feature and its configuration
- Explain IGMP snooping and its configuration
- Explain IGMP fast-leave in a switch
- Explain PIM snooping and its configuration

https://t.me/learningnets

https://t.me/learningnets

Mapping Multicast IP Addresses to MAC Addresses

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.



The addresses between 224.0.0.0 and 239.255.255.255, or all addresses with high-order bits set to 1110, make up the Class D range. These addresses are used for multicast purposes.

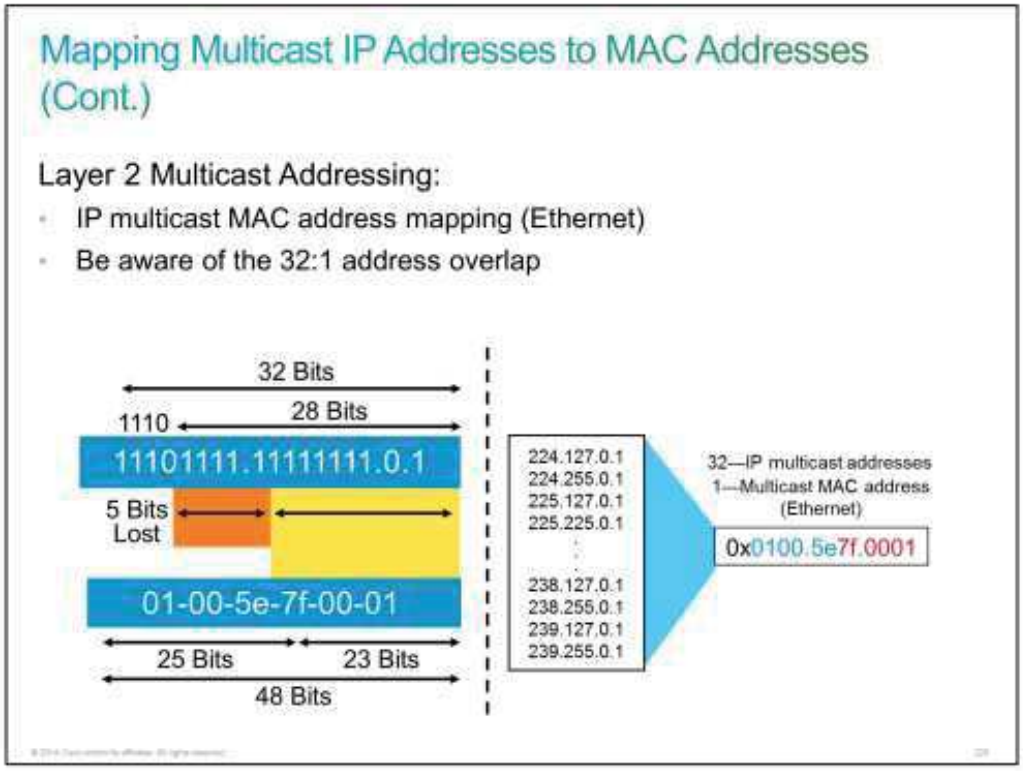
IANA is the authority responsible for the assignment of reserved Class D addresses. Other interesting reserved addresses in the range 224/8 are as follows:

- **224.0.0.1:** All systems on this subnet.
- **224.0.0.2:** All routers on this subnet.
- **224.0.0.4:** DVMRP routers.
- **224.0.0.5:** OSPF, all routers (RFC 1583).
- **224.0.0.6:** OSPF, designated routers, and BDRs (RFC 1583).
- **224.0.0.9:** RIPv2 routers.
- **224.0.0.13:** All PIMv2 routers.
- **224.0.1.39:** CISCO-RP-ANNOUNCE (Auto-RP).
- **224.0.1.40:** CISCO-RP-DISCOVERY (Auto-RP).

https://t.me/learningnets

IPv6 multicast addresses have the prefix ff00::/8. IPv6 multicast addresses are generally formed from four bit groups:

- Prefix (8 bits)
- Flags (4 bits)
- Scope (4 bits)
- Group ID (112 bits)



The translation between IP multicast and MAC addresses for Ethernet is achieved by mapping the low-order 23 bits of the IP (Layer 3) multicast address into the low-order 23 bits of the IEEE (Layer 2) MAC address. In the MAC address, the low-order bit (0x01) in the first octet indicates that the packet is a Layer 2 multicast packet. The 0x01005e prefix (known as the vendor code) has been reserved for use in mapping Layer 3 IP multicast addresses into Layer 2 MAC addresses. This loss of five bits of information when mapping Layer 3 multicast addresses to Layer 2 MAC addresses was not originally intended.

When Steve Deering was doing seminal research on IP multicast, he requested 16 OUIs to map all 28 bits of the Layer 3 IP multicast address into unique Layer 2 MAC addresses. An OUI is the high 24 bits of a MAC address that the IEEE assigns to an organization. Therefore, one OUI provides 24 bits of unique MAC addresses to the organization. However, because of budget constraints, Deering was able to access only one-half of one OUI. Therefore, Deering had only 23 bits of MAC address space with which to map the 28 bits of the IP multicast address. The 25th bit in the multicast MAC address should be set to 0.

Because there are 28 bits of unique address space for an IP multicast address (32 minus the first 4 bits containing the 1110 Class D prefix), and there are only 23 bits mapped into the IEEE MAC address, there are 5 bits of overlap (28 – 23 = 5, and 2 to the fifth power is 32). So there is a 32:1 overlap of Layer 3 addresses to Layer 2 addresses. Therefore, several Layer 3 addresses may map to the same Layer 2 multicast address.

https://t.me/learningnets

For example, the following 32 IP multicast addresses map to the same Layer 2 multicast of 01-00-5e-0a-00-01. The IP multicast addresses in binary = 1110 **xxxx x**000 1010 0000 0000 0000 0001 where **xxxx x** are the 5 overlapping bits:

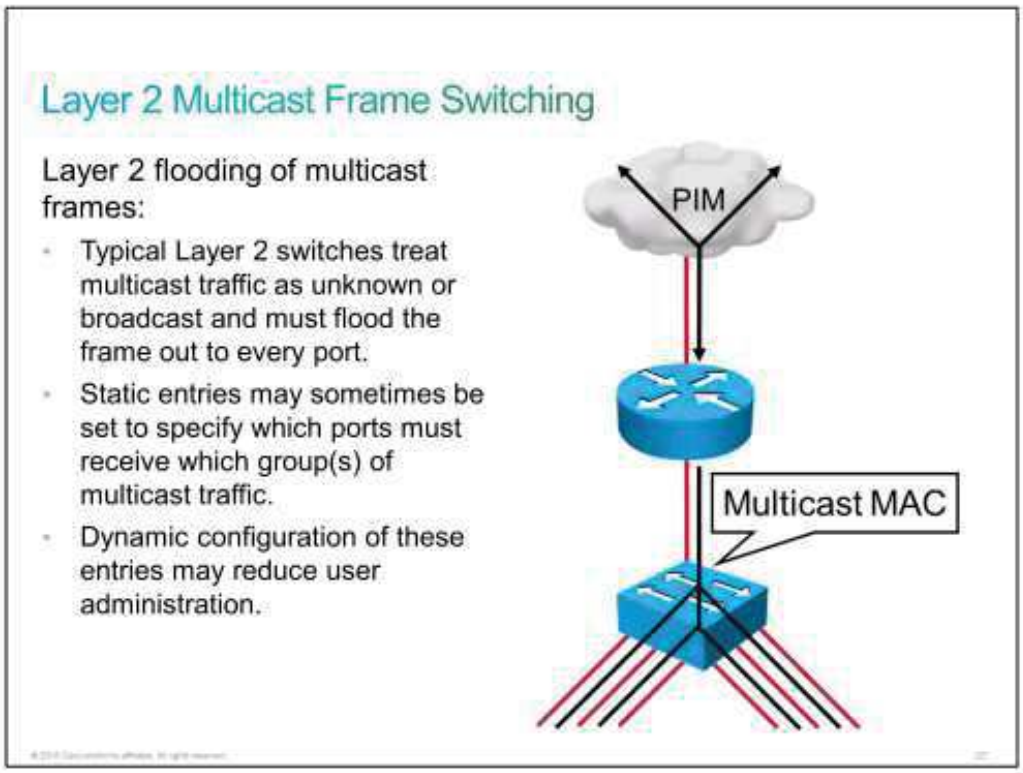
1110 **0000 0**000 1010 0000 0000 0000 0001 = 224.10.0.1
1110 **0000 1**000 1010 0000 0000 0000 0001 = 224.138.0.1
1110 **0001 0**000 1010 0000 0000 0000 0001 = 225.10.0.1
1110 **0001 1**000 1010 0000 0000 0000 0001 = 225.138.0.1
1110 **0010 0**000 1010 0000 0000 0000 0001 = 226.10.0.1
1110 **0010 1**000 1010 0000 0000 0000 0001 = 226.138.0.1
1110 **0011 0**000 1010 0000 0000 0000 0001 = 227.10.0.1
1110 **0011 1**000 1010 0000 0000 0000 0001 = 227.138.0.1
1110 **0100 0**000 1010 0000 0000 0000 0001 = 228.10.0.1
1110 **0100 1**000 1010 0000 0000 0000 0001 = 228.138.0.1
1110 **0101 0**000 1010 0000 0000 0000 0001 = 229.10.0.1
1110 **0101 1**000 1010 0000 0000 0000 0001 = 229.138.0.1
1110 **0110 0**000 1010 0000 0000 0000 0001 = 230.10.0.1
1110 **0110 1**000 1010 0000 0000 0000 0001 = 230.138.0.1
1110 **0111 0**000 1010 0000 0000 0000 0001 = 231.10.0.1
1110 **0111 1**000 1010 0000 0000 0000 0001 = 231.138.0.1
1110 **1000 0**000 1010 0000 0000 0000 0001 = 232.10.0.1
1110 **1000 1**000 1010 0000 0000 0000 0001 = 232.138.0.1
1110 **1001 0**000 1010 0000 0000 0000 0001 = 233.10.0.1
1110 **1001 1**000 1010 0000 0000 0000 0001 = 233.138.0.1
1110 **1010 0**000 1010 0000 0000 0000 0001 = 234.10.0.1
1110 **1010 1**000 1010 0000 0000 0000 0001 = 234.138.0.1
1110 **1011 0**000 1010 0000 0000 0000 0001 = 235.10.0.1
1110 **1011 1**000 1010 0000 0000 0000 0001 = 235.138.0.1
1110 **1100 0**000 1010 0000 0000 0000 0001 = 236.10.0.1
1110 **1100 1**000 1010 0000 0000 0000 0001 = 236.138.0.1
1110 **1101 0**000 1010 0000 0000 0000 0001 = 237.10.0.1
1110 **1101 1**000 1010 0000 0000 0000 0001 = 237.138.0.1
1110 **1110 0**000 1010 0000 0000 0000 0001 = 238.10.0.1
1110 **1110 1**000 1010 0000 0000 0000 0001 = 238.138.0.1
1110 **1111 0**000 1010 0000 0000 0000 0001 = 239.10.0.1
1110 **1111 1**000 1010 0000 0000 0000 0001 = 239.138.0.1

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

https://t.me/learningnets

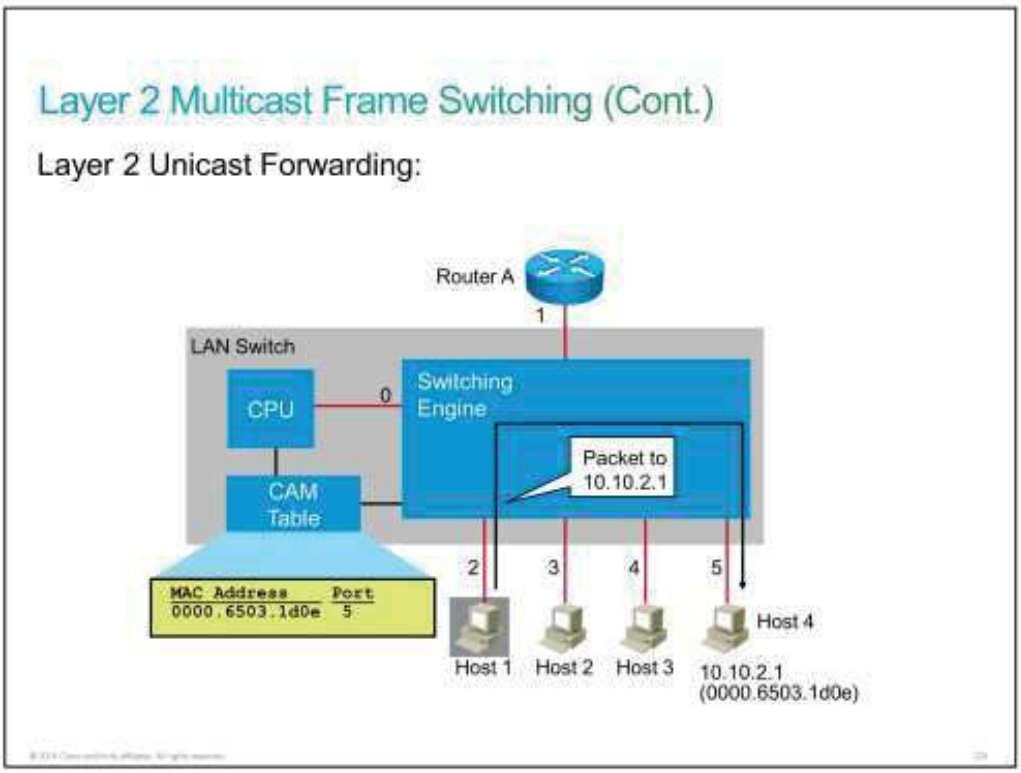
Layer 2 Multicast Frame Switching

For most Layer 2 switches, multicast traffic is treated like an unknown MAC address or broadcast frame. This causes the frame to be flooded out to every port within a VLAN. However, IP multicast hosts may join and be interested in only specific multicast groups. On most Layer 2 switches, all ports forward this traffic, resulting in wasted bandwidth on both the segments and on the end stations.



Cisco switches circumvent this behavior by using the CLI to program the switch to manually associate a multicast MAC address with various ports. For example, manually associating a multicast address with ports 5, 6, and 7 causes only those ports to receive multicast traffic that is destined for that multicast group. This method works acceptably. However, IP multicast hosts dynamically join and leave groups by using IGMP or MLD to signal to the multicast router. This static method of entering the multicast information is not very scalable. Dynamic configuration of the forwarding tables of the switches might be more effective and might reduce user administration.

https://t.me/learningnets



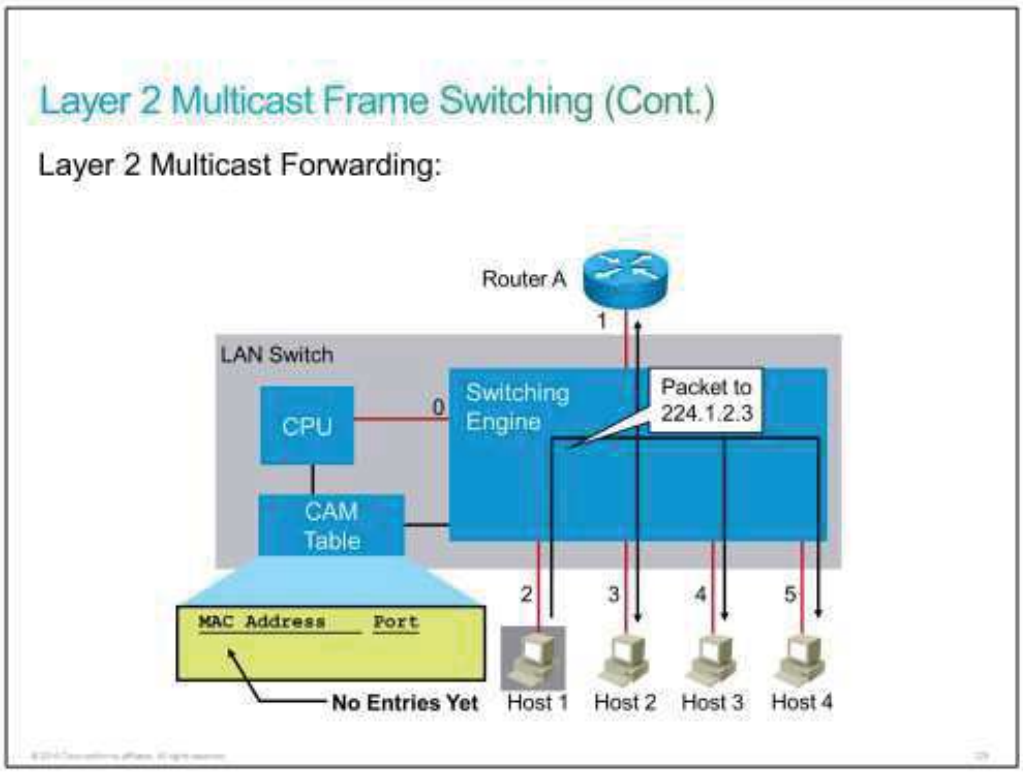
Most Layer 2 switches consist of the following components:

- **Switching engine:** The switching engine is used to switch packets from the input port to the output ports under the control of the CAM table.
- **CAM table:** The CAM table contains information that is used to control the operation of the switching engine. Each entry in this table contains a Layer 2 destination MAC address and output ports where packets that are addressed to this destination must be switched.
- **CPU:** The CPU is used to populate the CAM table with destination MAC addresses so that the switching engine may switch packets efficiently. The CPU "learns" the ports that are associated with a particular MAC address by watching arriving traffic that is sent by hosts.

In this figure, the switch has learned the port (port 5) that is associated with the Host 4 MAC address (0000.6503.1d0e). The CPU stores this information in the CAM table.

Because of a CAM table entry, the switching engine switches packets that arrive with the Host 4 MAC address as the destination to port 5. Because the packets are IP packets, the appropriate MAC address was determined by the ARP mechanism.

https://t.me/learningnets



In this figure, Host 1 sends a multicast packet on port 2 of the switch. Because the switch does not have an entry in the CAM table that matches the destination multicast MAC address, it treats the packet as an unknown or broadcast packet and floods it out to all ports.

Many multicast switching solutions have been developed to improve the behavior of the switches when they receive multicast frames. These solutions include:

- Cisco Group Management Protocol
- IGMP snooping
- GMRP
- RGMP
- PIM snooping

Cisco Group Management Protocol is a Cisco proprietary protocol that runs between a multicast router and a switch. This protocol enables the Cisco multicast router receiving the IGMP messages that are sent by hosts to inform the switch about the information that is contained in the IGMP packet.

With IGMP snooping, the switch intercepts IGMP messages from the host itself and updates its MAC table accordingly. To implement IGMP snooping without suffering switch performance degradation, it is necessary to make the switch Layer 3-aware. This action is typically accomplished by using Layer 3 ASICs.

GMRP requires software components to run on both the switch and on the host. GMRP uses GARP to register and propagate multicast membership information in a switching domain. GARP is a Layer 2 transport mechanism that lets switches and end systems propagate useful information throughout the switching domain.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

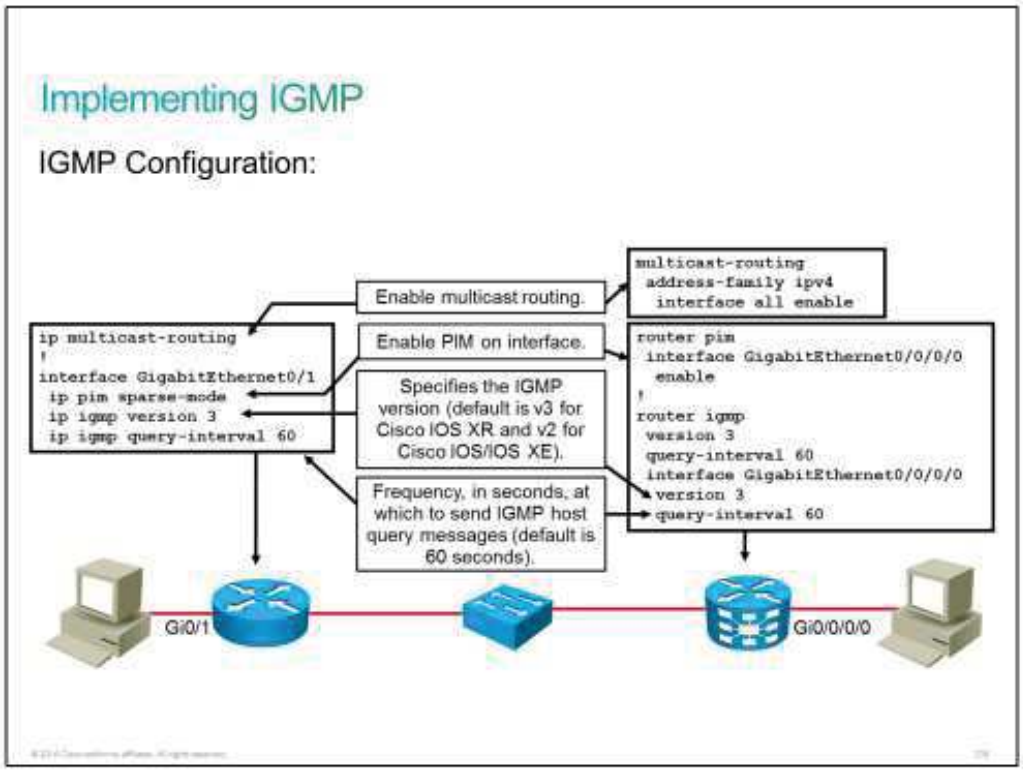
RGMP is a Cisco proprietary protocol that is used between multicast routers and switches to restrict multicast packet forwarding in switches to only those routers where the multicast packets are needed. The purpose of RGMP is to reduce the multicast router overhead. RGMP is designed for backbone-switched networks where multiple high speed routers are interconnected. RGMP and Cisco Group Management Protocol are both Cisco proprietary protocols, but are mutually exclusive. Enabling one disables the other. RGMP does work with IGMP snooping, while IGMP snooping only helps prevent hosts from receiving unwanted multicast traffic, IGMP snooping does not prevent a multicast router from receiving unwanted multicast traffic. Using RGMP, a router can tell a switch to only forward specific multicast group traffic to the router, but not multicast traffic from all groups. This scenario is useful when you have a switch that is upstream from a multicast router and the hosts downstream from the multicast router only need traffic from a particular multicast group, but the upstream switch on the far side of the multicast router does not know that. By using RGMP between the multicast router and the upstream switch, the multicast router can send RGMP messages to the upstream switch to join or leave a multicast group.

In networks where a Layer 2 router interconnects several routers—such as an IXP—the router floods IP multicast packets on all multicast router ports by default, even if there are no multicast receivers downstream. With PIM snooping enabled, the router restricts multicast packets for each IP multicast group to only those multicast router ports that have downstream receivers joined to that group. When you enable PIM snooping, the router learns which multicast router ports need to receive the multicast traffic within a specific VLAN by listening to the PIM hello messages, PIM join and prune messages, and bidirectional PIM-designated forwarder-election messages.

https://t.me/learningnets

Implementing IGMP

On the Cisco IOS and IOS XE platforms, IGMP is enabled by default. On Cisco IOS XR, you have to enable multicast routing as the first step.



Use the following Cisco IOS XR commands to enable multicast routing:

- **multicast-routing**
- **address-family ipv4**
- **interface all enable**

You can set the IGMP version on a per-interface basis using the Cisco IOS/IOS XE **ip igmp version** interface command, or Cisco IOS XR **version** router IGMP command (you can also use the same command under a particular IGMP interface). The default setting is version 2 for Cisco IOS/IOS XE and version 3 for Cisco IOS XR. To configure the frequency at which Cisco IOS/IOS XE/IOS XR Software sends IGMP host query messages, use the Cisco IOS/IOS XE **ip igmp query-interval** interface command or the Cisco IOS XR **query-interval** router IGMP command (you can also use the same command under a particular IGMP interface).

Note	To enable IGMP on an interface, you must enable PIM on the interface first. To enable PIM on an interface, use the **`ip pim sparse-mode`** command on the Cisco IOS and IOS XE router. On the Cisco IOS XR router, first enter router igmp configuration mode, then enter interface configuration mode and use the **`enable`** command.

https://t.me/learningnets

Implementing IGMP (Cont.)

```
RP/0/RSP0/CPU0:PE7# show igmp interface GigabitEthernet0/0/0/0
GigabitEthernet0/0/0/0 is up, line protocol is up
  Internet address is 192.168.107.70/24
  IGMP is enabled on interface
  Current IGMP version is 3
  IGMP query interval is 60 seconds
  IGMP querier timeout is 125 seconds
  IGMP max query response time is 10 seconds
  Last member query response interval is 1 seconds
  IGMP activity: 12 joins, 7 leaves
  IGMP querying router is 192.168.107.70 (this system)
```

Displays IGMP interface settings.

```
RP/0/RSP0/CPU0:PE7# show igmp groups
IGMP Connected Group Membership
Group Address   Interface              Uptime    Expires   Last Reporter
224.0.0.2       GigabitEthernet0/0/0/0   2d19h     never     10.7.1.1
```

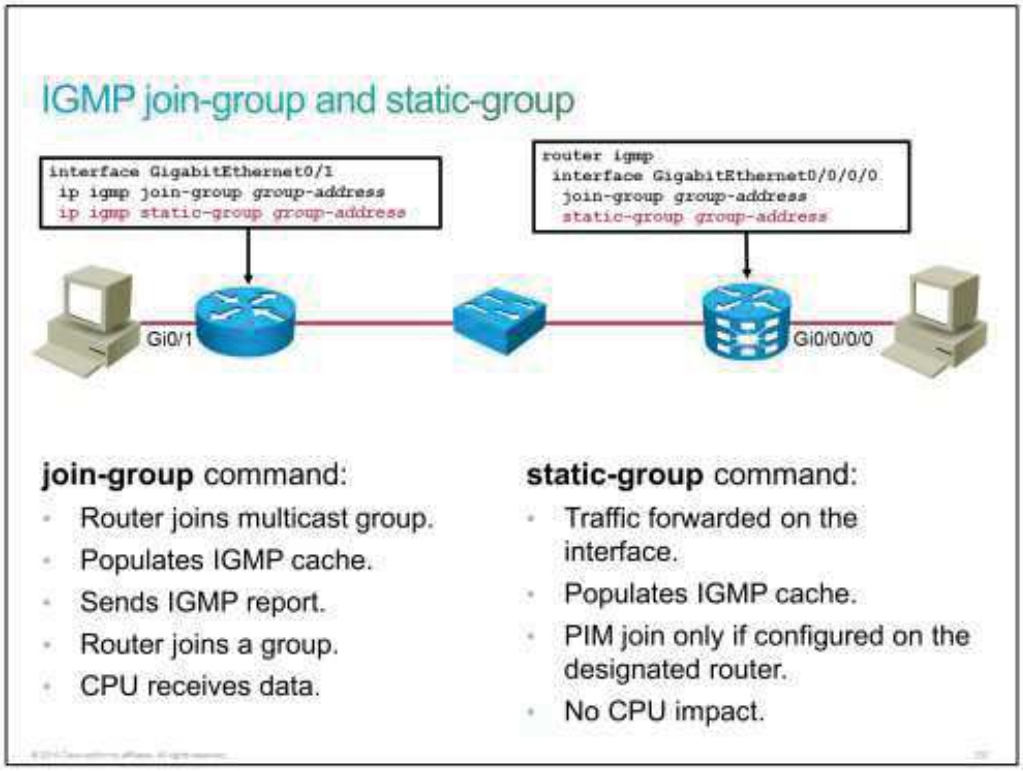
Displays the IGMP groups that are registered on the router.

To display IGMP interface settings, use the Cisco IOS/IOS XE **show ip igmp interface** or the Cisco IOS XR **show igmp interface** command. To view the IGMP groups that are registered on the router, use the Cisco IOS/IOS XE **show ip igmp groups** or the Cisco IOS XR **show igmp groups** command.

https://t.me/learningnets

IGMP join-group and static-group

If you want the router to join a multicast group you have to configure this with proper CLI commands.



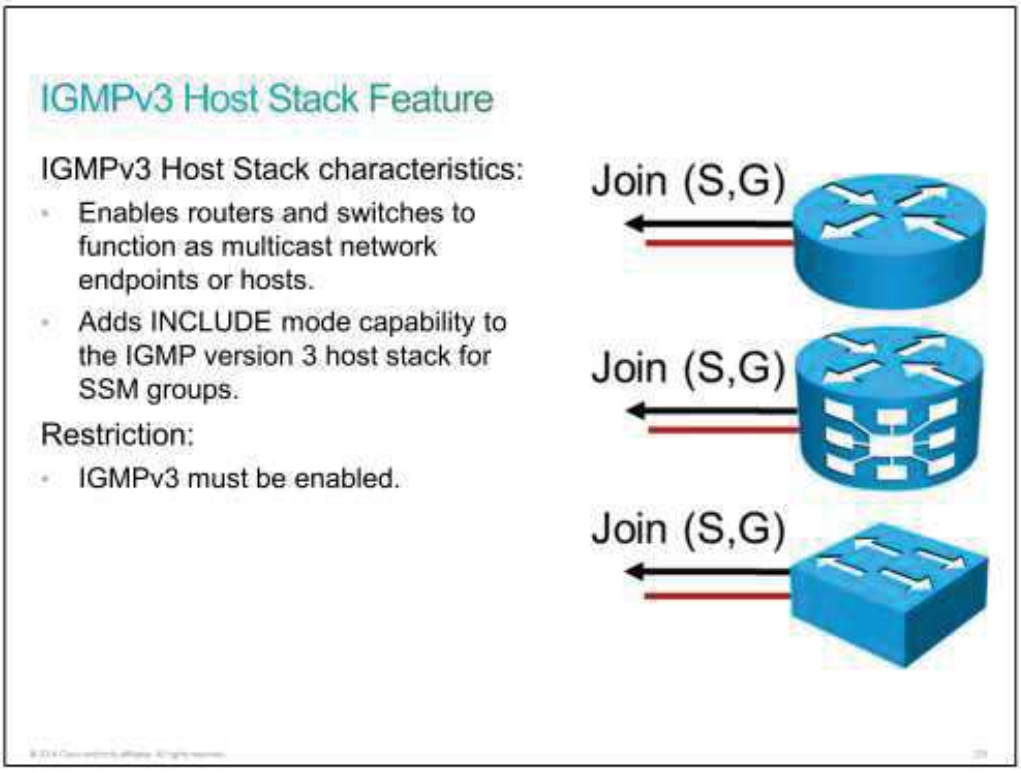
To have the router join a multicast group, use the Cisco IOS/IOS XE **ipigmp join-group** interface configuration command, or the Cisco IOS XR **join-group** command under IGMP configuration mode. This command causes the router to join a group and start listening for the multicast traffic for that group. This command causes the router CPU to process the received multicast packets, and 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/IOS XE **ip igmp static-group** interface configuration command or the Cisco IOS XR **static-group** command under IGMP configuration mode. If the router on which this command is configured is the designated router, it will generate a PIM join. This command has no impact on the router CPU, as it does not process the multicast packets. This command causes the router to forward multicast traffic down an interface.

https://t.me/learningnets

IGMPv3 Host Stack Feature

The IGMPv3 host stack feature enables routers and switches to function as multicast network endpoints or hosts. The feature adds INCLUDE mode capability to the IGMP version 3 host stack for SSM groups.



The Cisco IOS/IOS XE **ip igmp join-group** or Cisco IOS XR **join-group** command will be accepted, but IGMPv3 reports will not be sent if IGMP version 3 is not configured on the interface.

An IGMPv3 report is sent when one of the following events occurs:

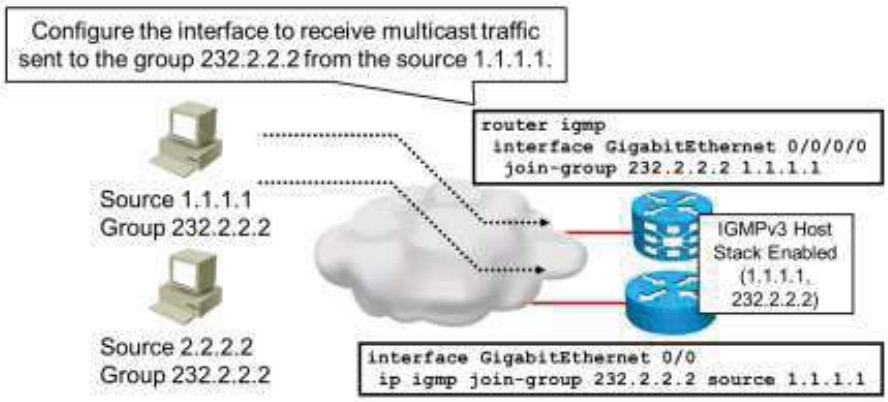
- When a source join group is configured and there is no existing state for this group and source.
- When no join group source is configured and there is an existing state for this group and source.
- When a query is received.

https://t.me/learningnets

IGMPv3 Host Stack Feature (Cont.)

Enabling the IGMPv3 Host Stack:

- Applications can leverage SSM as the preferred method.
- Assists in troubleshooting.



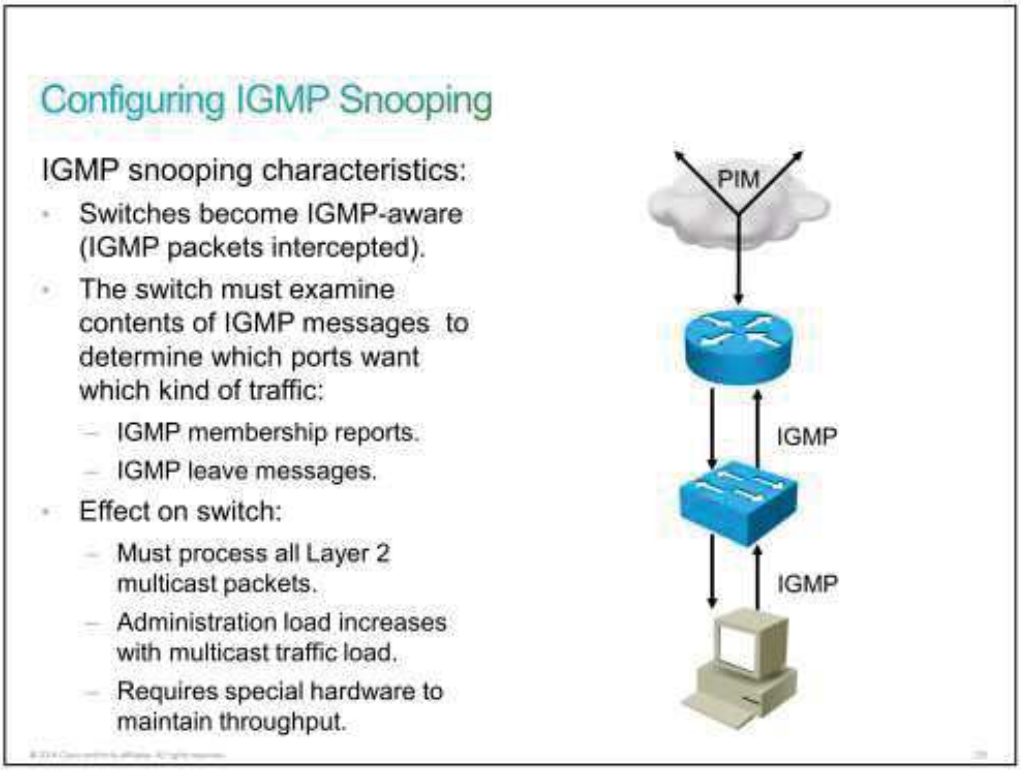
The IGMPv3 host stack feature enables routers and switches to function as a multicast receiver to aid in multicast troubleshooting. This figure shows how to configure the interface of the router to receive multicast traffic, sent to the group 232.2.2.2, only from the source 1.1.1.1.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Configuring IGMP Snooping

IGMP snooping is designed to control multicast propagation scope in a switched environment.



IGMP snooping is used in the Cisco IP NGN infrastructure access layer of the service provider network. As its name implies, IGMP snooping switches become IGMP-aware and listen in on the IGMP conversations between hosts and routers. This activity requires the processor in each switch to identify and intercept a copy of all IGMP packets that are flowing between routers and hosts, and vice versa. It includes the following IGMP packets:

- IGMP membership reports
- IGMP leaves

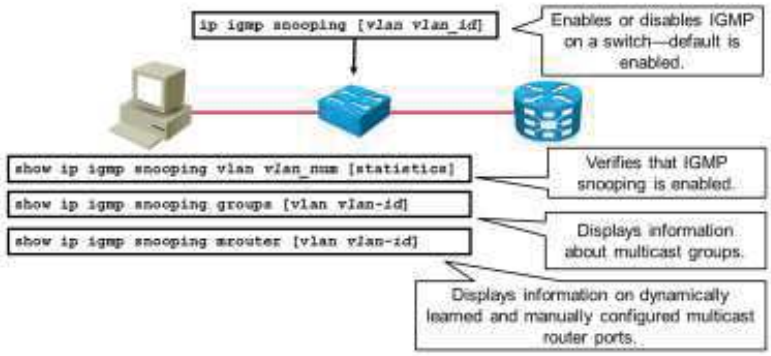
If IGMP snooping is not carefully implemented, a switch may have to intercept all Layer 2 multicast packets to identify IGMP packets. This action may have a significant impact on switch performance. Proper designs require special hardware (Layer 3 ASIC) to avoid this problem, which may directly affect the overall cost of the switch. Thus, switches must effectively become Layer 3-aware to avoid serious performance problems because of IGMP snooping.

https://t.me/learningnets

Configuring IGMP Snooping (Cont.)

Configuring IGMP Snooping:

- IGMP snooping in switches enables automatic detection of multicast groups.
- Configuration is needed on switches only: transparent to routers and multicast hosts.



IGMP snooping enables switches to listen to IGMP membership reports and IGMP leave messages. IGMP snooping must be configured on switches only, and is transparent to the routers and multicast hosts.

IGMP snooping may not be enabled if Cisco Group Management Protocol or GMRP is enabled on the same switch. However, these technologies may be combined in the same Layer 2 network.

On a Cisco switch, IGMP snooping is globally enabled by default. Use the **ip igmp snooping** Cisco IOS command to enable IGMP snooping on a switch. Using the **vlan** option available on switches that support Cisco IOS Software, you can enable IGMP snooping selectively on a specific VLAN. Ensure that Cisco Group Management Protocol is disabled before enabling IGMP snooping.

Multicast MAC addresses may also be manually added using the **mac-address-table static** Cisco IOS command. The information entered statically supersedes the dynamically learned information.

To verify IGMP snooping status, use the **show ip igmp snooping vlan | statistics | mrouter** Cisco IOS/IOS XE command. The options may differ depending on the platform.

Use the **show ip igmp snooping groups** command to display the IGMP snooping multicast table for the switch or the multicast information. Use the command with the **vlan** keyword to display the multicast table for a specified multicast VLAN or specific multicast information.

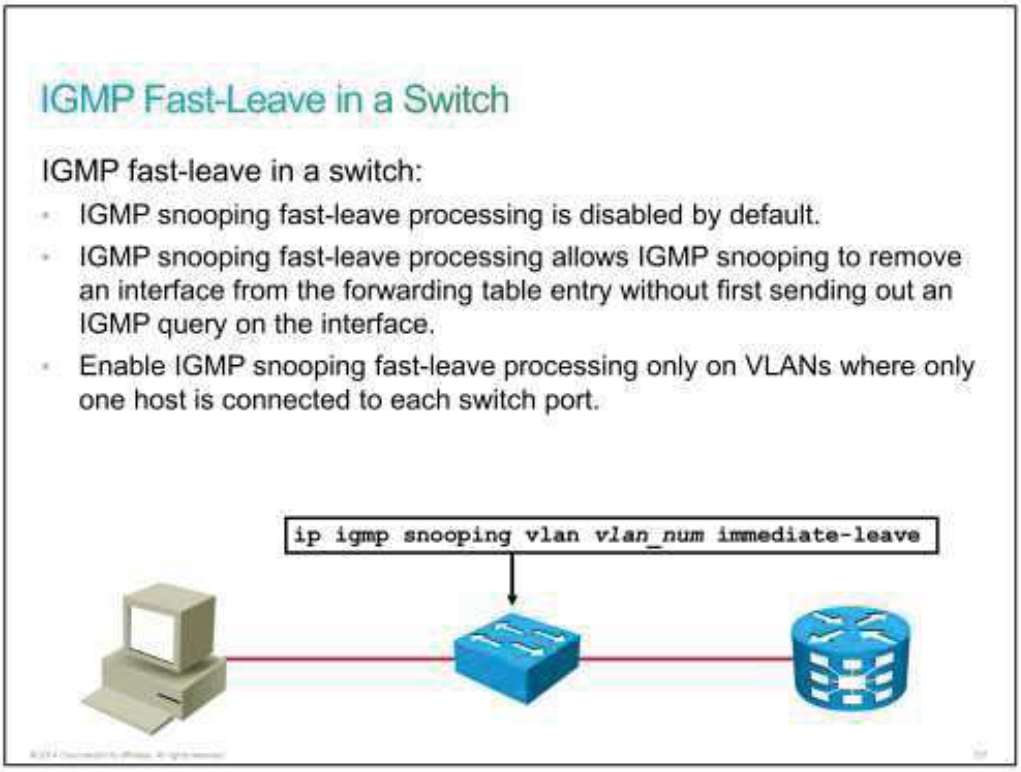
Use the **show ip igmp snooping mrouter** command to display the IGMP snooping dynamically learned and manually configured multicast router ports for the switch or for the specified multicast VLAN.

 © 2014 Cisco Systems, Inc.

https://t.me/learningnets

IGMP Fast-Leave in a Switch

Using normal IGMP leave processing, when IGMP snooping receives a group-specific IGMPv2 leave message from a host, it sends out an IGMP query to determine if any other devices connected to that interface are interested in traffic for the specific multicast group. If IGMP snooping does not receive an IGMP join message in response to the query, it assumes that no other devices connected to the interface are interested in receiving traffic for this multicast group, and it removes the interface from its forwarding table entry for that multicast group. If the leave message was from the only remaining interface with hosts interested in the group, and IGMP snooping does not receive an IGMP join in response to the IGMP query, it removes the group entry and relays the IGMP leave to the multicast router.



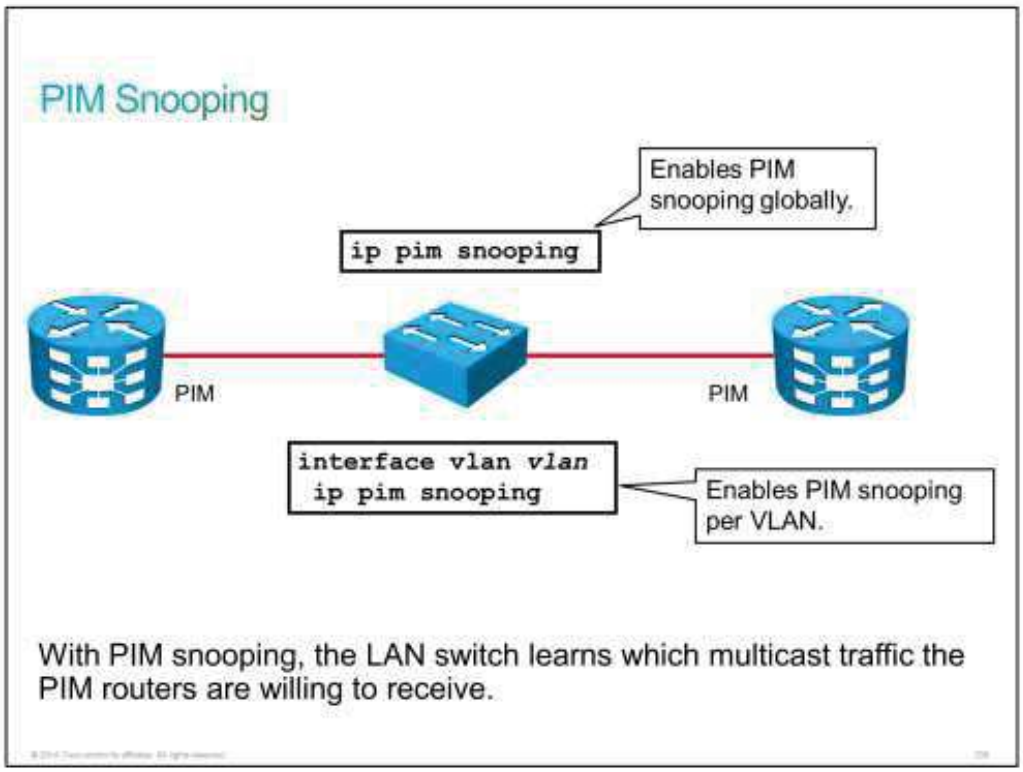
IGMP snooping fast-leave processing allows IGMP snooping to remove an interface from the forwarding table entry without first sending out an IGMP query on the interface. Upon receiving a group-specific IGMPv2 leave message, IGMP snooping immediately removes the interface from the forwarding table entry for that multicast group.

Note	Use fast-leave processing only on VLANs where only one host is connected to each switch port. If fast-leave is enabled in VLANs where more than one host is connected to a switch port, the multicast traffic of some hosts might be dropped inadvertently. IGMP snooping fast-leave processing is supported only with IGMP version 2 and 3 hosts. IGMP version 3 fast-leave processing is enabled by default.

https://t.me/learningnets

PIM Snooping

In networks where a Layer 2 switch interconnects several routers, such as an IXP, the switch floods IP multicast packets on all multicast router ports by default, even if there are no multicast receivers downstream.



If PIM snooping is enabled, the switch restricts multicast packets for each IP multicast group to only those multicast router ports that have downstream receivers joined to that group. When you enable PIM snooping, the switch learns which multicast router ports need to receive the multicast traffic within a specific VLAN by listening to the PIM hello messages, PIM join and prune messages, and bidirectional PIM designated forwarder-election messages.

Use the **ip pim snooping** command in global configuration mode to enable PIM snooping globally. To enable PIM snooping on a per-VLAN basis, first enter VLAN interface configuration mode and then use the **ip pim snooping** command.

Note	This feature supported on Cisco 6500 series switches,7200 series routers and ME3600X/ME3800X series switches.

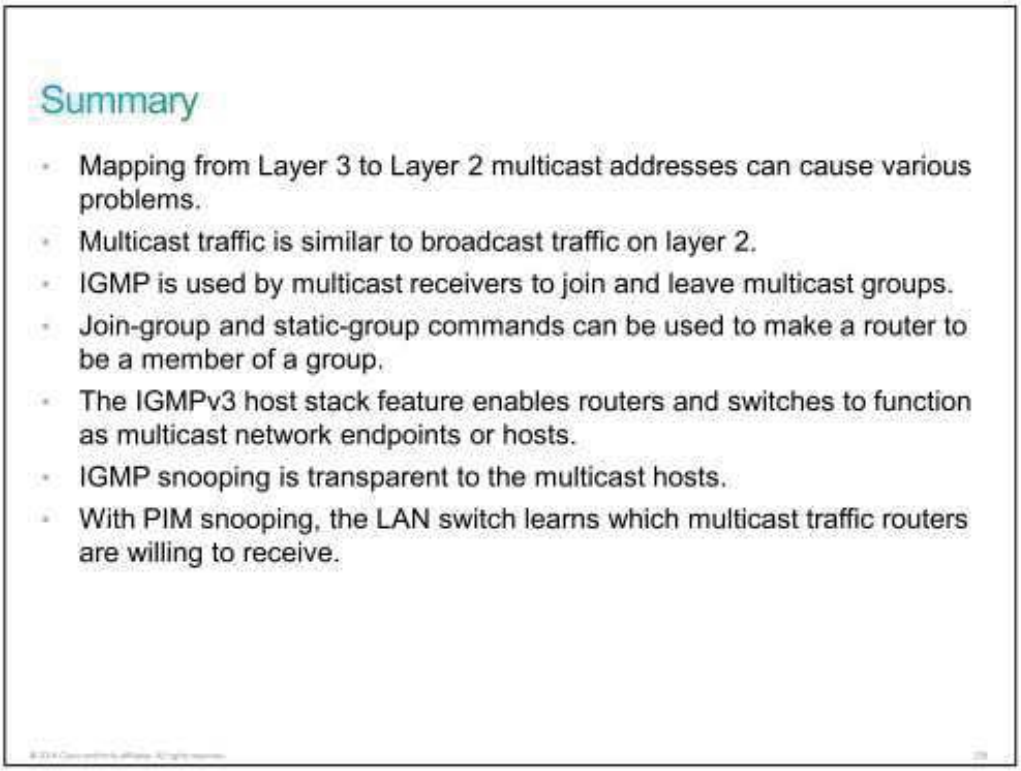
Note	To use PIM snooping, you must first enable IGMP snooping.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Summary

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



https://t.me/learningnets

https://t.me/learningnets

Lesson 4: Populating the Mroute Table

Overview

This lesson explains the purpose of the multicast route (mroute) table and the routing protocols that are supported for mroute table population. The MP-BGP feature adds multicast capabilities to BGP. Specifically, it enables a multicast routing policy throughout the Internet and connects multicast topologies within and between BGP autonomous systems. This lesson explains the basic concepts of MP-BGP and its use in the IP multicast environment.

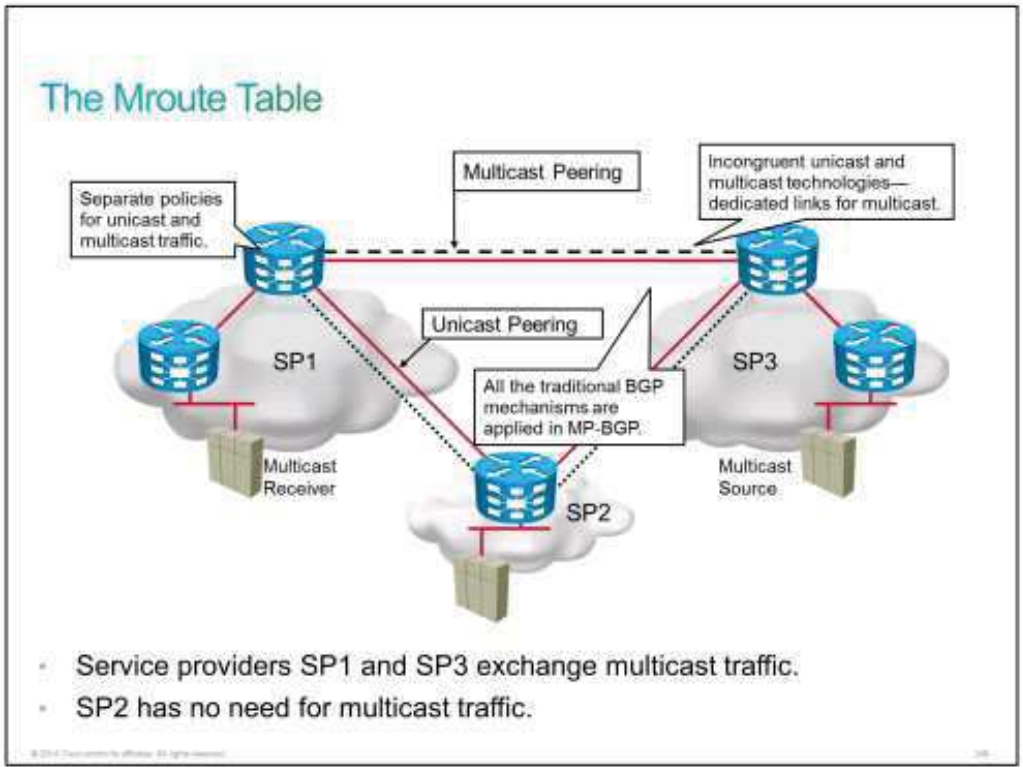
Upon completing this lesson, you will be able to:

- Show an example where the multicast and unicast routing topology is incongruent and explain the mroute table
- Describe MP-BGP for IP multicast
- Describe MP-BGP capability negotiation
- Describe MP-BGP multicast configuration
- Describe MP-BGP multicast verification

https://t.me/learningnets

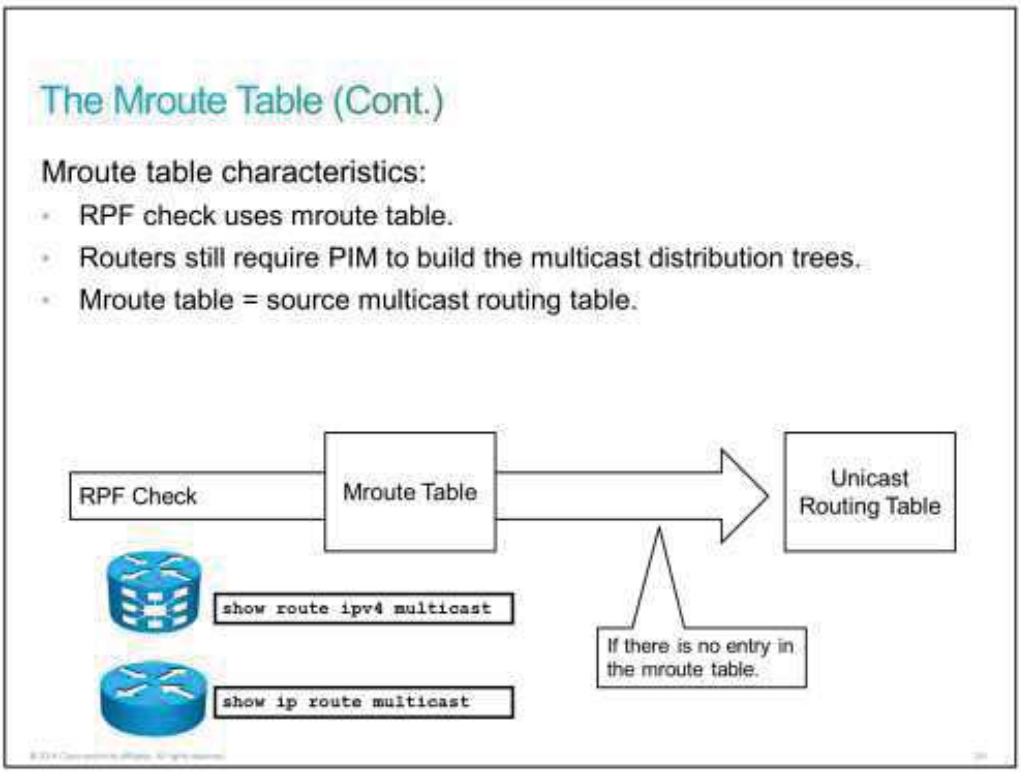
The Mroute Table

In a situation with incongruent unicast and multicast topologies, the use of MP-BGP is necessary. Even in a situation where the topologies overlap, MP-BGP is recommended, in order to implement separate policies for unicast and multicast traffic.



This figure shows a service provider situation where three service providers have BGP peering with one another. SP2 currently has no plans to deploy IP multicast. However, SP1 and SP3 plan to exchange the multicast traffic and want to dedicate a separate link to this exchange. Therefore, MP-BGP is needed in order to provide the proper information for RPF checks and similar IP multicast-related operations.

https://t.me/learningnets



Mroute contains prefixes for IP multicast purposes. URIB is used for unicast.

The RPF check uses the mroute table. If there is no mroute table entry, the unicast routing table is used instead. Routers will still require PIM to build the multicast distribution trees.

The mroute table is used to reach the multicast source, and the mroute table may be referred as the source multicast routing table. The mroute table is shown with the Cisco IOS/IOS XE **show ip route multicast** command or the Cisco IOS XR **show route ipv4 multicast** command. Static multicast routes or MP-BGP can be used to populate the mroute table.

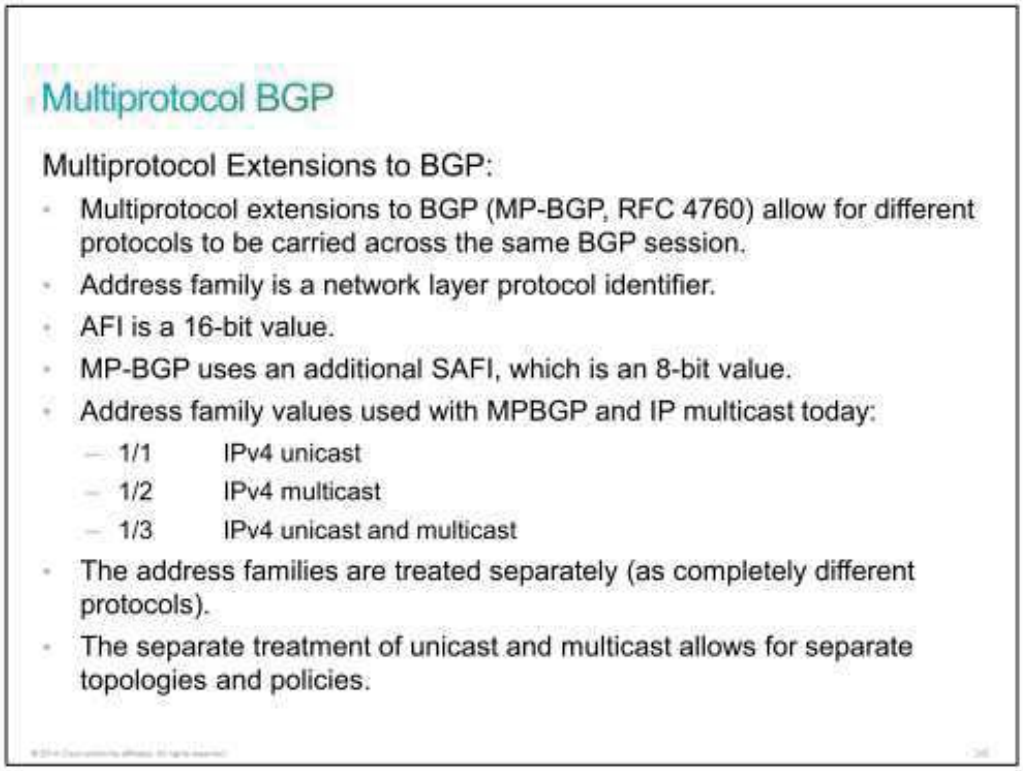
Historically, IP multicast has evolved in Cisco IOS without a multicast RPF routing table. Some of the functionalities that might be associated with a multicast RPF routing table have been emulated inside the multicast subsystem through a sequence of lookups in multiple tables: unicast routing table, MBGP, and a database of multicast-specific static routes. The results of the multiple lookups were combined to implement a preference strategy where the RPF checks can be performed against three different sources in the following order:

1. If static mroutes (which have a default administrative distance of 0) are configured, the static mroutes are first used for doing the RPF checks.
2. If MP-BGP is configured, then the MBGP (mroute) table is used for doing the RPF checks if there are no static mroutes.
3. Lastly, by default, if there are no static mroutes or MBGP tables, the unicast routing table is used for doing the RPF checks.

https://t.me/learningnets

Multiprotocol BGP

MP-BGP, as defined in RFC 4760, allows the BGP to carry different network layer protocols in its update messages. Since IP multicast is, in a way, a different network layer protocol (although it is still IP), different policies are applied to it.



The MP-BGP distinguishes between address families, which reflect the network layer protocol. Address families are identified by the AFI and support for sub-address families is provided with a SAFI.

In order to distinguish different types of traffic within a certain network layer protocol, an AFI/SAFI combination such as 1/2 is usually used. AFI = 1 (IPv4) and AFI = 2 (IPv6). Thus, a single BGP update message simultaneously carries various protocols.

From the IPv4 multicast perspective, the following three AFI/SAFI combinations are of interest:

- 1/1: IPv4 unicast information
- 1/2: IPv4 multicast information
- 1/3: IPv4 unicast and multicast information

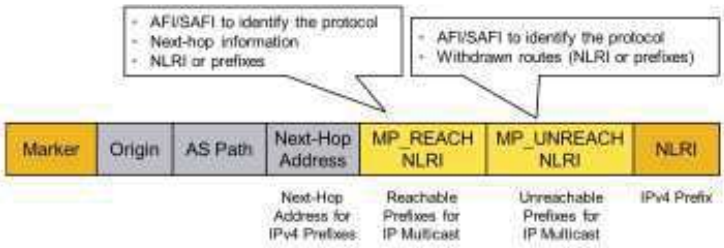
These different values ensure that those address and sub-address families are treated as completely separate protocols. In turn, this treatment provides for separate topologies and separate policies. The routing tables in BGP and RIBs are kept separately for different address families.

https://t.me/learningnets

Multiprotocol BGP (MP-BGP) (Cont.)

MP-BGP and IP Multicast:

- MP-BGP can carry routing information for any Layer 3 protocol.
- MP-BGP for IP multicast:
  - MBGP
  - The unicast routes carried in updates are used for multicast purposes—for example, RPF checks to multicast sources or RPs
- Two new path attributes in MP-BGP: MP_REACH_NLRI and MP_UNREACH_NLRI



MP-BGP carries any network layer protocol information. Because of its underlying BGP features, MP-BGP is selected as the most appropriate protocol to also carry IP multicast information.

When used for IP multicast, MP-BGP carries prefixes that are stored in a separate BGP MRIB. The information in BGP MRIB is used to perform RPF checks and other IP multicast-related operations, primarily between domains.

In MP-BGP, two new attributes are introduced: **MP_REACH_NLRI** (Multiprotocol Reachable NLRI) and **MP_UNREACH_NLRI** (Multiprotocol Unreachable NLRI). Both of these are optional nontransitive BGP attributes.

The MP_REACH_NLRI attribute is used to announce reachable prefixes and contains the following fields:

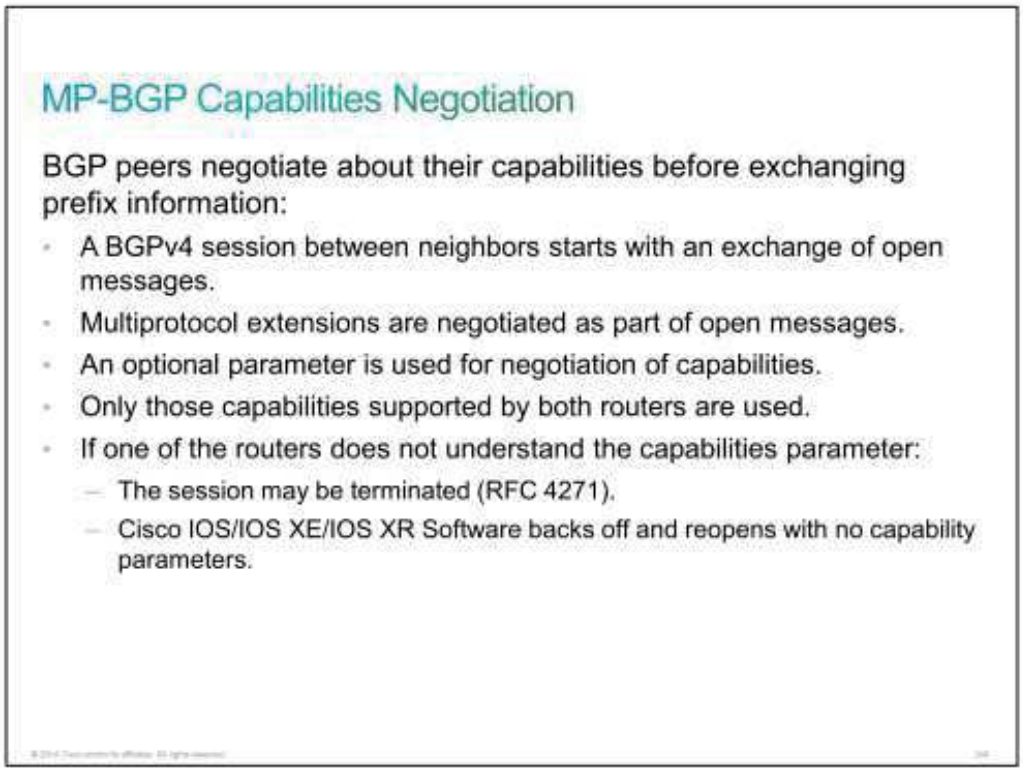
- **AFI/SAFI:** This field provides identification information for the network layer protocol (address family) and possible sub-address family within the protocol.
- **Next-hop address** (expressed in the same address family)**:** This field shows, for example, that for IPv4, the next hop is provided in IPv4 format, whereas for IPv6, the IPv6 address format is used.
- **NLRI:** This field shows the prefixes (routes) that are reachable via the same next-hop address.

The MP_UNREACH_NLRI information is used to revoke certain prefixes, and contains similar information to MP_REACH_NLRI. The only missing field is the next-hop address, since it is not needed for the withdrawn routes.

https://t.me/learningnets

MP-BGP Capability Negotiation

Since the introduction of MP-BGP, the multiprotocol extensions are negotiated between BGP neighbors during the open phase of a BGP session. During this phase, the neighbors agree on which address families they will use.



The capabilities parameter (type 2) is an optional parameter that is used to negotiate the capabilities between the peers. The capabilities are distinguished by the capability code, which is an 8-bit field. In an IP multicast environment, the capabilities negotiation uses the capability code value 1. The AFI/SAFI values are 1/1 for unicast, 1/2 for multicast, and 1/3 for unicast and multicast information.

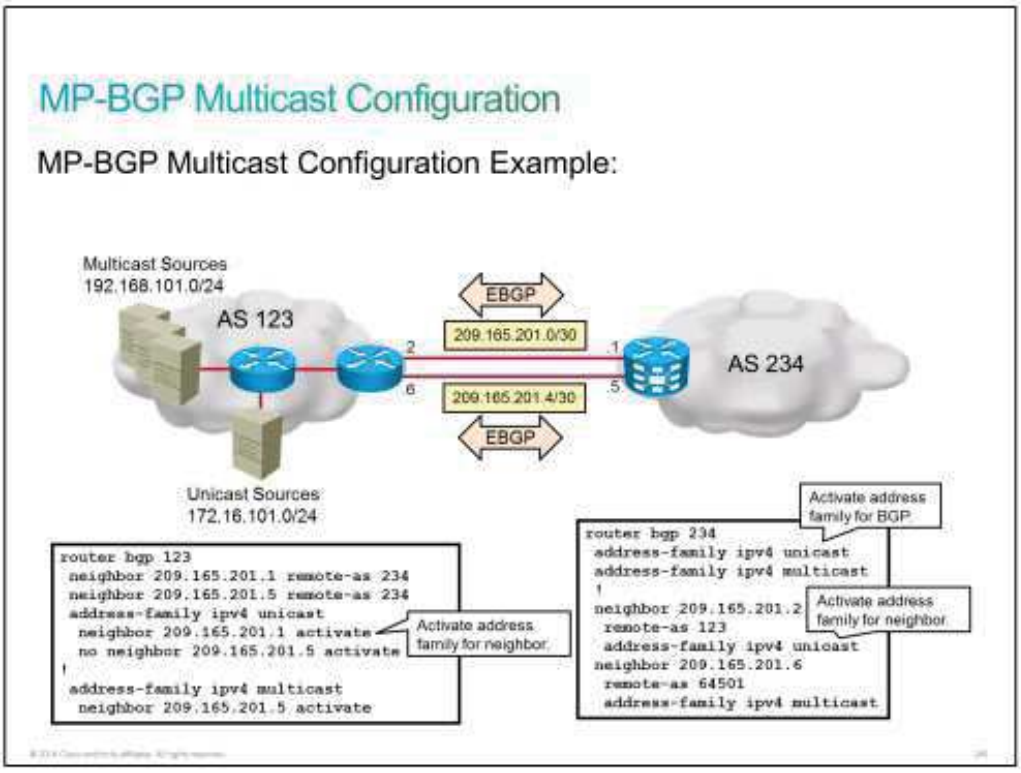
During the negotiation phase, both neighbors offer the capabilities that they are willing to support. The result of the negotiation is that only those capabilities that are supported by both neighbors are finally used. If no capabilities match, the notification message is generated, and peering is not established.

Some of the older implementations of BGP do not understand the capabilities parameter (type 2). If this parameter is received during the open phase and is not understood, the receiving peer may terminate the BGP session (according to RFC 4271). The Cisco IOS/IOS XE/IOS XR implementation smoothly manages the situation. If the capabilities parameter is not understood, the BGP reopens without this parameter.

https://t.me/learningnets

MP-BGP Multicast Configuration

Different BGP configurations are necessary to enable BGP multicast support on different routing Cisco platforms.

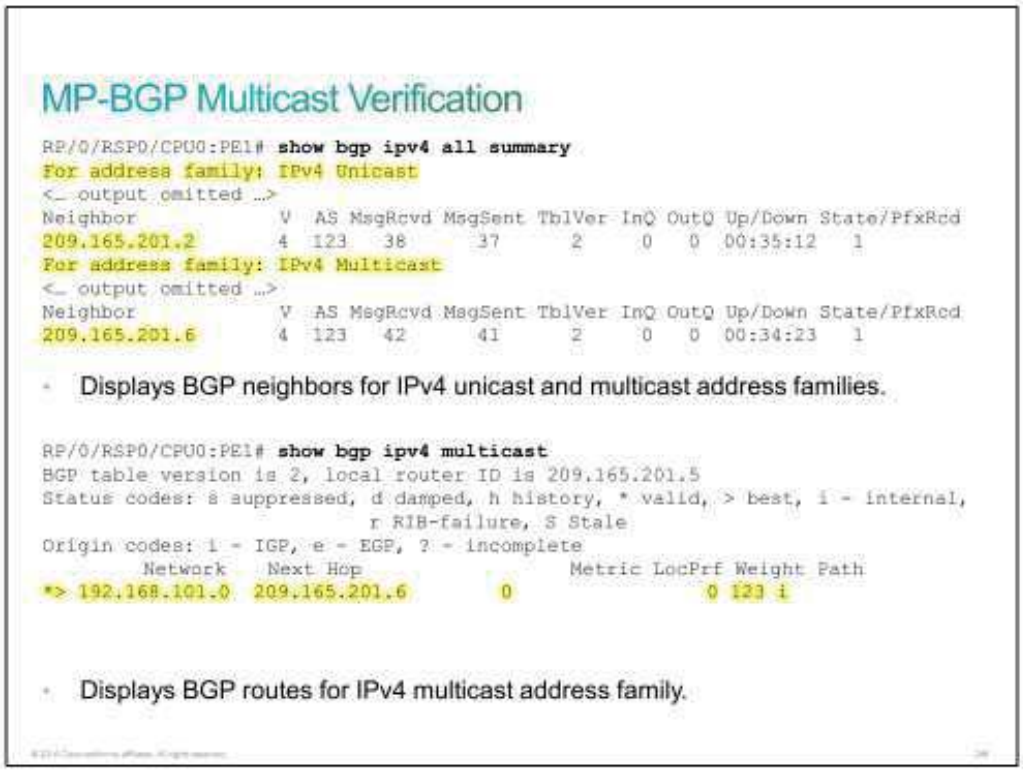


To enable BGP multicast support on the Cisco IOS XR Software platform, the address family for IPv4 multicast needs to be activated globally and then on the BGP neighbor level. On the Cisco IOS/IOS XE Software platform, the BGP neighbor needs to be activated in the BGP address family for IPv4 multicast. BGP neighbors are activated by default for IPv4 unicast address family on Cisco IOS and IOS XE Software routers.

https://t.me/learningnets

MP-BGP Multicast Verification

To verify MP-BGP multicast operation, use the **show bgp ipv4 all summary** command. This command will display BGP neighbors for IPv4 unicast and multicast.

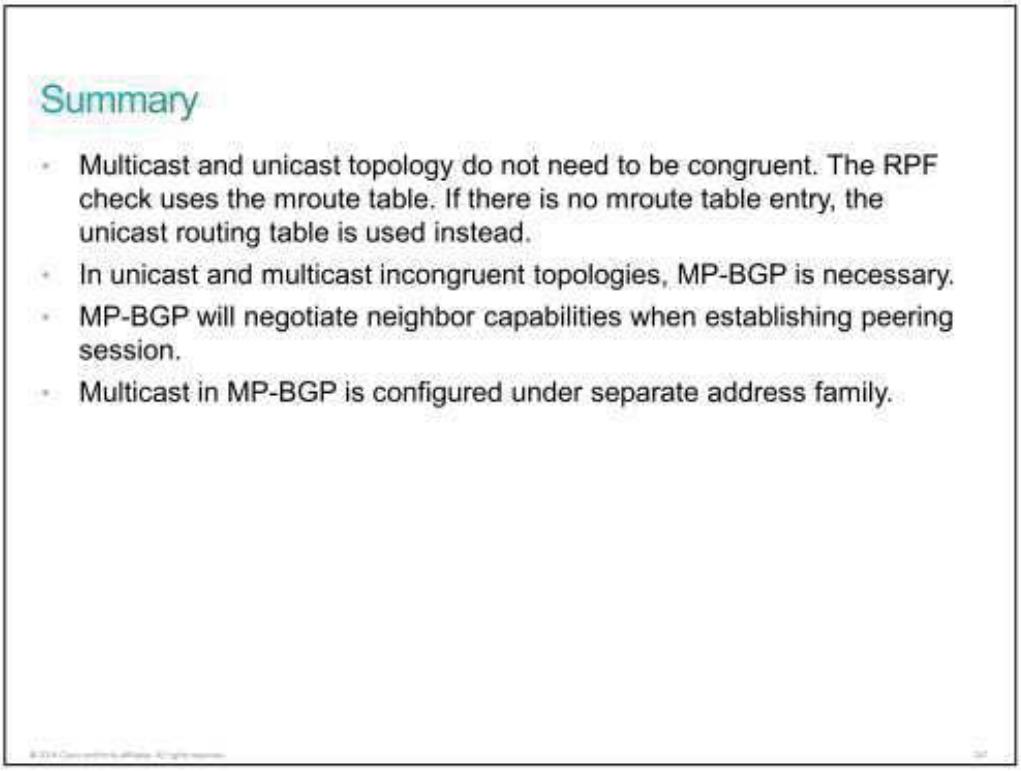


Use the **show bgp ipv4 multicast** command to display BGP routes that are used for multicast to determine RPF information.

https://t.me/learningnets

Summary

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

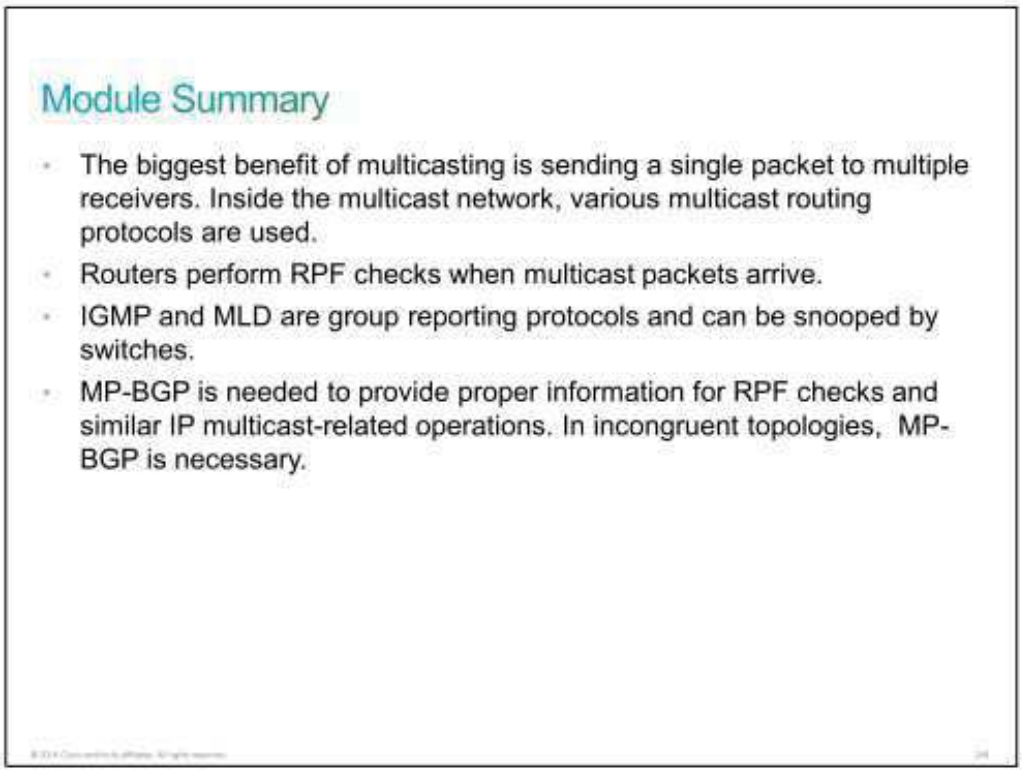


https://t.me/learningnets

https://t.me/learningnets

Lesson 5: Module Summary

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



https://t.me/learningnets

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. Which three advantages does multicast have over unicast? (Choose three.) (Source: Introducing IP Multicast)

   A. Enhanced efficiency
   B. Optimized performance
   C. Distributed applications management
   D. Best-effort delivery
   E. Congestion control
   F. Out-of-sequence delivery of packets

2. Which two multicast application models are the most common? (Choose two.) (Source: Introducing IP Multicast)

   A. One-to-many
   B. Many-to-many
   C. Many-to-one
   D. Few-to-many

3. Which multicast traffic is usually sent out with a TTL of 1? (Source: Introducing IP Multicast)

   A. Locally scoped traffic destined for addresses in the range of 224.0.0.0 through 224.0.0.255
   B. Globally scoped traffic destined for addresses in the range of 224.0.0.0 through 224.0.0.255
   C. Globally scoped traffic destined for addresses in the range of 224.0.1.0 through 238.255.255.255
   D. Locally scoped traffic destined for addresses in the range of 224.0.1.0 through 238.255.255.255

https://t.me/learningnets

4. How do multicast applications learn about available sessions? (Source: Introducing IP Multicast)

A. Through announcements about a well-known (predefined) group
B. Through a directory server
C. Through a web page
D. Through email
E. All of the above

5. Which protocol runs between the source and the first-hop router? (Source: Introducing IP Multicast)

A. IGMP
B. MP-BGP
C. MSDP
D. PIM
E. None of the above

6. How does an RPF check work? (Source: Introducing IP Multicast)

A. The routing table for unicast traffic is checked against the source address in the multicast datagram.
B. The routing table for unicast traffic is checked against the source address in the unicast datagram.
C. The routing table for multicast traffic is checked against the source address in the multicast datagram.
D. The routing table for multicast traffic is checked against the source address in the unicast datagram.

7. Why does multicast routing use RPF information? (Source: Defining Multicast Distribution Trees and Forwarding)

A. To prevent forwarding loops
B. To prevent forwarding loops
C. To prevent switching loops
D. All of the above
E. None of the above

8. In the sparse mode multicast-enabled network, which router knows about all sources in the network? (Source: Defining Multicast Distribution Trees and Forwarding)

A. The first-hop router
B. The rendezvous point
C. The last-hop router
D. All routers

9. Which two statements are true about (*,G) entries in the multicast routing table? (Choose two.) (Source: Defining Multicast Distribution Trees and Forwarding)

A. (*,G) is used for a particular source sending to a particular group.
B. (*,G) is used for any source sending to a particular group.
C. Traffic is forwarded via the shortest path from the source.
D. Traffic is forwarded via a meeting point for a particular group.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

10. Which two multicast routing protocols are sparse mode protocols? (Choose two.) (Source: Defining Multicast Distribution Trees and Forwarding)

A. PIM-DM
B. PIM-SM
C. DVMRP
D. MOSPF
E. CBT

11. Which two statements about interdomain multicast routing protocols are true? (Choose two.) (Source: Defining Multicast Distribution Trees and Forwarding)

A. MP-BGP is responsible for RPF information.
B. MSDP is responsible for RPF information.
C. MSDP distributes information about the existence of active sources.
D. MP-BGP distributes information about the existence of active sources.

12. Which IP multicast address is used by the PIMv2 protocol? (Source: Defining Multicast on the LAN)

A. 224.0.0.1
B. 224.0.0.2
C. 224.0.0.13
D. 224.0.1.39
E. 224.0.1.40

13. Which protocol does the host use to tell routers about group membership? (Source: Defining Multicast on the LAN)

A. ARP
B. RARP
C. IGMP
D. DNS

14. Which Cisco IOS XR Software command displays the version of IGMP that is currently active on an interface? (Source: Defining Multicast on the LAN)

A. **show igmp interface**
B. **show interface**
C. **show interface brief**
D. **show igmp interface**

15. Which IP multicast address is used by IGMPv3? (Source: Defining Multicast on the LAN)

A. Global multicast group address 224.0.0.13
B. Global multicast group address 224.0.0.22
C. Link-local multicast group address 224.0.0.13
D. Link-local multicast group address 224.0.0.22

https://t.me/learningnets

16. Which command is used to configure the frequency at which Cisco IOS Software sends IGMP host query messages? (Source: Defining Multicast on the LAN)

A. **ip igmp query frequency**
B. **ip igmp interval**
C. **ip igmp query-interval**
D. **igmp query-interval**

17. Which Cisco IOS XR Software command correctly enables the IGMPv3 host stack? (Source: Defining Multicast on the LAN)

A. **Join-group 232.2.2.2 1.1.1.1**
B. **Source 232.2.2.2 join-group 1.1.1.1**
C. **Join-group 1.1.1.1 source 232.2.2.2**
D. **Join-group 232.2.2.2 / source 1.1.1.1**

18. In the incongruent unicast and multicast topologies, the use of MP-BGP is not necessary. (Source: Populating the Mroute Table)

A. True
B. False

19. Which two attributes were introduced in the new version of MP-BGP? (Choose two.) (Source: Populating the Mroute Table)

A. MP_REACH_NLRI
B. BGP_REACH_NLRI
C. MP_UNREACH_NLRI
D. BGP_UNREACH_NLRI

20. Which Cisco IOS/IOS XE command series is used to enable the IP multicast capability in MP-BGP? (Source: Populating the Mroute Table)

A. **address-family ipv4 multicast** and **neighbor** *ip-address* **activate**
B. **address-family multicast** and **neighbor** *ip-address* **activate**
C. **address-family ipv4 multicast activate** and **neighbor** *ip-address* **multicast**
D. **address-family multicast activate** and **neighbor** *ip-address* **multicast**
E. **address-family multicast activate** and **neighbor** *ip-address* **activate**

21. What does AFI/SAFI value 1/2 stand for? (Source: Populating the Mroute Table)

A. IP unicast BGP information
B. IP unicast and multicast BGP information
C. IP multicast BGP information
D. None of the above

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

22. Which Cisco IOS/IOS XE command lists the capabilities that are advertised and received? (Source: Populating the Mroute Table)

A. **show ip bgp neighbor**
B. **show ip bgp interface**
C. **show ip bgp**
D. **debug ip bgp**

https://t.me/learningnets

https://t.me/learningnets

Module Self-Check Answers

Answer Key

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

https://t.me/learningnets

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

Glossary

ACK

acknowledgment. Notification sent from one network device to another to acknowledge that some event occurred (for example, the receipt of a message).

ACL

access control list. A list kept by routers to control access to or from the router for a number of services (for example, to prevent packets with a certain IP address from leaving a particular interface on the router).

AFI

Address Family Identifier.

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.

ASIC

application-specific integrated circuit.

ASM

Any Source Multicast.

Auto-RP

Auto-Rendezvous Point

BDR

backup designated router.

BFD

Bidirectional Forwarding Detection

BGP

Border Gateway Protocol. Interdomain routing protocol that replaces EGP. BGP exchanges reachability information with other BGP systems. It is defined by RFC 1163.

CAM

content-addressable memory.

CBT

Core Based Tree

https://t.me/learningnets

CE

customer edge. Identifies the network devices, connected to a provider network, that are under the administrative control of the customer.

CLI

command-line interface. An interface that allows the user to interact with the operating system by entering commands and optional arguments. The UNIX operating system and DOS provide CLIs.

CoPP

Control Plane Policing.

CPU

central processing unit. The hardware within a computer system or smartphone that carries out the instructions of a computer program by performing the basic arithmetical, logical, and input-output operations of the system.

DDoS

distributed denial of service

DoS

denial of service. An intentional or unintentional attack on a device that makes the resource unavailable to perform its normal function.

DR

designated router.

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.

DVMRP

Distance Vector Multicast Routing Protocol

EBGP

Exterior Border Gateway Protocol.

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.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

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.

FIB

forwarding information base

FTTx

fiber to the x.

GARP

Gratuitous Address Resolution Protocol.

Gigabit Ethernet

Standard for a high-speed Ethernet, approved by the IEEE (Institute of Electrical and Electronics Engineers) 802.3z standards committee in 1996.

GMRP

Generic Attribute Registration Protocol (GARP) Multicast Registration Protocol

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.

IBGP

Interior Border Gateway Protocol.

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.

IEEE

Institute of Electrical and Electronics Engineers. Professional organization whose activities include the development of communications and network standards. IEEE LAN standards are the predominant LAN standards today.

IEEE 802.3

IEEE LAN protocol that specifies an implementation of the physical layer and the MAC sublayer of the data link layer. IEEE 802.3 uses CSMA/CD access at a variety of speeds over a variety of physical media. Extensions to the IEEE 802.3 standard specify implementations for Fast Ethernet. Physical variations of the original IEEE 802.3 specification include10Base2,10Base5,10BaseF,10BaseT, and10Broad36. Physical variations for Fast Ethernetinclude100BaseT,100BaseT4, and100BaseX.

https://t.me/learningnets

IETF

Internet Engineering Task Force. Task force consisting of over 80 working groups responsi 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 men adjacent multicast router.

IGMPv1

Internet Group Management Protocol version 1

IGMPv2

Internet Group Management Protocol version 2

IGMPv3

Internet Group Management Protocol version 3

IGP

interior gateway protocol. Internet protocol used to exchange routing information within an system. Examples of common Internet IGPs include IGRP, OSPF, and RIP.

IP

Internet Protocol. Network layer protocol in the TCP/IP stack offering a connectionless inte service. IP provides features for addressing, type-of-service specification, fragmentation an 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 E) and is written as 4 octets separated by periods (dotted decimal format). Each address con network number, an optional subnetwork number, and a host number. The network and sub numbers together are used for routing, and the host number is used to address an individual network or subnetwork. A subnet mask is used to extract network and subnetwork informat address. CIDR provides a new way of representing IP addresses and subnet masks. Also ca address.

IPC

interprocess communication.

IPTV

Internet Protocol television (IPTV) is a system through which television services are delive Internet protocol suite.

IPv4

IP version 4. Internet Protocol version 4 is the fourth version in the development of the IP a version of the protocol to be widely deployed. Along with IPv6, IPv4 is at the core of stand internetworking methods of the Internet. IPv4 is still used to route most traffic across the In connectionless protocol for use on packet-switched link layer networks (for example, Ether on a best-effort delivery model in that it does not guarantee delivery and does not assure pr or avoidance of duplicate delivery.

https://t.me/learningnets

IETF

Internet Engineering Task Force. Task force consisting of over 80 working groups responsi 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 men adjacent multicast router.

IGMPv1

Internet Group Management Protocol version 1

IGMPv2

Internet Group Management Protocol version 2

IGMPv3

Internet Group Management Protocol version 3

IGP

interior gateway protocol. Internet protocol used to exchange routing information within an system. Examples of common Internet IGPs include IGRP, OSPF, and RIP.

IP

Internet Protocol. Network layer protocol in the TCP/IP stack offering a connectionless inte service. IP provides features for addressing, type-of-service specification, fragmentation an 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 E) and is written as 4 octets separated by periods (dotted decimal format). Each address con network number, an optional subnetwork number, and a host number. The network and sub numbers together are used for routing, and the host number is used to address an individual network or subnetwork. A subnet mask is used to extract network and subnetwork informat address. CIDR provides a new way of representing IP addresses and subnet masks. Also ca address.

IPC

interprocess communication.

IPTV

Internet Protocol television (IPTV) is a system through which television services are delive Internet protocol suite.

IPv4

IP version 4. Internet Protocol version 4 is the fourth version in the development of the IP a version of the protocol to be widely deployed. Along with IPv6, IPv4 is at the core of stand internetworking methods of the Internet. IPv4 is still used to route most traffic across the In connectionless protocol for use on packet-switched link layer networks (for example, Ether on a best-effort delivery model in that it does not guarantee delivery and does not assure pr or avoidance of duplicate delivery.

https://t.me/learningnets

IETF

Internet Engineering Task Force. Task force consisting of over 80 working groups responsi 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 men adjacent multicast router.

IGMPv1

Internet Group Management Protocol version 1

IGMPv2

Internet Group Management Protocol version 2

IGMPv3

Internet Group Management Protocol version 3

IGP

interior gateway protocol. Internet protocol used to exchange routing information within an system. Examples of common Internet IGPs include IGRP, OSPF, and RIP.

IP

Internet Protocol. Network layer protocol in the TCP/IP stack offering a connectionless inte service. IP provides features for addressing, type-of-service specification, fragmentation an 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 E) and is written as 4 octets separated by periods (dotted decimal format). Each address con network number, an optional subnetwork number, and a host number. The network and sub numbers together are used for routing, and the host number is used to address an individual network or subnetwork. A subnet mask is used to extract network and subnetwork informat address. CIDR provides a new way of representing IP addresses and subnet masks. Also ca address.

IPC

interprocess communication.

IPTV

Internet Protocol television (IPTV) is a system through which television services are delive Internet protocol suite.

IPv4

IP version 4. Internet Protocol version 4 is the fourth version in the development of the IP a version of the protocol to be widely deployed. Along with IPv6, IPv4 is at the core of stand internetworking methods of the Internet. IPv4 is still used to route most traffic across the In connectionless protocol for use on packet-switched link layer networks (for example, Ether on a best-effort delivery model in that it does not guarantee delivery and does not assure pr or avoidance of duplicate delivery.

https://t.me/learningnets

IETF

Internet Engineering Task Force. Task force consisting of over 80 working groups responsi 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 men adjacent multicast router.

IGMPv1

Internet Group Management Protocol version 1

IGMPv2

Internet Group Management Protocol version 2

IGMPv3

Internet Group Management Protocol version 3

IGP

interior gateway protocol. Internet protocol used to exchange routing information within an system. Examples of common Internet IGPs include IGRP, OSPF, and RIP.

IP

Internet Protocol. Network layer protocol in the TCP/IP stack offering a connectionless inte service. IP provides features for addressing, type-of-service specification, fragmentation an 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 E) and is written as 4 octets separated by periods (dotted decimal format). Each address con network number, an optional subnetwork number, and a host number. The network and sub numbers together are used for routing, and the host number is used to address an individual network or subnetwork. A subnet mask is used to extract network and subnetwork informat address. CIDR provides a new way of representing IP addresses and subnet masks. Also ca address.

IPC

interprocess communication.

IPTV

Internet Protocol television (IPTV) is a system through which television services are delive Internet protocol suite.

IPv4

IP version 4. Internet Protocol version 4 is the fourth version in the development of the IP a version of the protocol to be widely deployed. Along with IPv6, IPv4 is at the core of stand internetworking methods of the Internet. IPv4 is still used to route most traffic across the In connectionless protocol for use on packet-switched link layer networks (for example, Ether on a best-effort delivery model in that it does not guarantee delivery and does not assure pr or avoidance of duplicate delivery.

https://t.me/learningnets

IPv6

IP version 6. Replacement for the current version of IP (version 4). IPv6 includes supp
packet header, which can be used to identify flows. Formerly called IPng (next generat

IS-IS

Intermediate System-to-Intermediate System. OSI link-state hierarchical routing proto
Phase V routing, whereby ISs (routers) exchange routing information based on a singl
network topology.

ISP

Internet service provider. Company that provides Internet access to other companies ar

IXP

Internet exchange point

LAN

local-area network. High-speed, low-error data network covering a relatively small ge
few thousand meters). LANs connect workstations, peripherals, terminals, and other d
building or other geographically limited area. LAN standards specify cabling and sign
and data link layers of the OSI model. Ethernet, FDDI, and Token Ring are widely use

LIR

local Internet registry

LPTS

Local Packet Transport Services.

MAC

Media Access Control. Lower of the two sublayers of the data link layer defined by th
sublayer handles access to shared media, such as whether token passing or contention

MAC address

Standardized data link layer address that is required for every port or device that conn
devices in the network use these addresses to locate specific ports in the network and t
routing tables and data structures. MAC addresses are 6 bytes long and are controlled
known as a hardware address, MAC layer address, and physical address.

MASC

Multicast Address Set Claim

MBGP

Multicast Border Gateway Protocol.

MD5

Message Digest 5. A one-way hashing algorithm that produces a 128-bit hash. Both M
Algorithm (SHA) are variations on MD4 and are designed to strengthen the security of
algorithm. Cisco uses hashes for authentication within the IPsec framework. Also used
authentication in SNMP v.2. MD5 verifies the integrity of the communication, authent
checks for timeliness.

https://t.me/learningnets

IPv6

IP version 6. Replacement for the current version of IP (version 4). IPv6 includes supp
packet header, which can be used to identify flows. Formerly called IPng (next generat

IS-IS

Intermediate System-to-Intermediate System. OSI link-state hierarchical routing proto
Phase V routing, whereby ISs (routers) exchange routing information based on a singl
network topology.

ISP

Internet service provider. Company that provides Internet access to other companies ar

IXP

Internet exchange point

LAN

local-area network. High-speed, low-error data network covering a relatively small ge
few thousand meters). LANs connect workstations, peripherals, terminals, and other d
building or other geographically limited area. LAN standards specify cabling and sign
and data link layers of the OSI model. Ethernet, FDDI, and Token Ring are widely use

LIR

local Internet registry

LPTS

Local Packet Transport Services.

MAC

Media Access Control. Lower of the two sublayers of the data link layer defined by th
sublayer handles access to shared media, such as whether token passing or contention

MAC address

Standardized data link layer address that is required for every port or device that conn
devices in the network use these addresses to locate specific ports in the network and t
routing tables and data structures. MAC addresses are 6 bytes long and are controlled
known as a hardware address, MAC layer address, and physical address.

MASC

Multicast Address Set Claim

MBGP

Multicast Border Gateway Protocol.

MD5

Message Digest 5. A one-way hashing algorithm that produces a 128-bit hash. Both M
ariations on MD4 and are designed to strengthen the security of
ashes for authentication within the IPsec framework. Also used
v.2. MD5 verifies the integrity of the communication, authent

https://t.me/learningnets

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).

IS-IS

Intermediate System-to-Intermediate System. OSI link-state hierarchical routing protocol based on DECnet Phase V routing, whereby ISs (routers) exchange routing information based on a single metric to determine network topology.

ISP

Internet service provider. Company that provides Internet access to other companies and individuals.

IXP

Internet exchange point

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.

LIR

local Internet registry

LPTS

Local Packet Transport Services.

MAC

Media Access Control. Lower of the two sublayers of the data link layer defined by the IEEE. The MAC sublayer handles access to shared media, such as whether token passing or contention will be used.

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.

MASC

Multicast Address Set Claim

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.

https://t.me/learningnets

MED

multi-exit discriminator.

MLD

Multicast Listener Discovery

MOSPF

Multicast Open Shortest Path First

MP-BGP

Multiprotocol Border Gateway Protocol.

MPLS

Multiprotocol Label Switching. Switching method that forwards IP traffic using a label. This label instructs the routers and the switches in the network where to forward the packets based on pre-established IP routing information.

MQC

modular QoS CLI. Command line structure that allows modular configuration of QoS configuration elements to provide independence between classification and policy.

MRIB

Multicast Routing Information Base

MSS

maximum segment size. A TCP protocol parameter that specifies the maximum amount of data that a device can receive in a single TCP segment.

MTU

maximum transmission unit. Maximum packet size, in bytes, that a particular interface can handle.

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."

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.

NGN

next-generation network.

NLRI

Network Layer Reachability Information.

https://t.me/learningnets

NSF

Nonstop Forwarding

NSR

Nonstop Routing.

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

OUI

Organizational Unique Identifier. Three octets that are assigned by the IEEE in a block of 48-bit LAN addresses.

P router

provider router.

PA address

provider-assigned address.

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.

PI address

provider-independent address.

PIC

Prefix-Independent Convergence.

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.

https://t.me/learningnets

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.

PMTU

path maximum transmission unit.

P-network

provider network

QoS

quality of service. Measure of performance for a transmission system that reflects its transmission quality and service availability.

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.

RGMP

Router-Port Group Management Protocol

RIB

routing information base

RIP

Routing Information Protocol. A distance-vector routing protocol that uses hop count as a routing metric.

RIPv2

Routing Information Protocol version 2

RIR

regional Internet registry

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.

https://t.me/learningnets

RP

route processor

RPF

Reverse Path Forwarding

RPL

Routing Policy Language

RTBH

Remotely triggered black hole

SAFI

Subsequent Address Family Identifier

SAP

Session Announcement Protocol.

SDR

Session Directory

SPD

Selective Packet Discard.

SPT

shortest path tree

SSM

Source Specific Multicast

SSO

single sign-on.

SSO

Stateful Switchover.

SYN

synchronization.

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.

TCP/IP

Transmission Control Protocol/Internet Protocol. Common name for the suite of protocols developed by the U.S. DoD in the 1970s to support the construction of worldwide internetworks. TCP and IP are the two best-known protocols in the suite.

https://t.me/learningnets

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.

URIB

Unicast Routing Information Base

URL

uniform resource locator. Type of formatted identifier that describes the access method and the location of an information resource object on the Internet. [RFC 1738]

uRPF

Unicast Reverse Path Forwarding.

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.

VoD

video on demand. System using video compression to supply video programs to viewers when requested via ISDN or cable.

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.

VPNv4

virtual private network version 4.

VRF

A VPN routing/forwarding instance. A VRF consists of an IP routing table, a derived forwarding table, a set of interfaces that use the forwarding table, and a set of rules and routing protocols that determine what goes into the forwarding table. In general, a VRF includes the routing information that defines a customer VPN site that is attached to a PE router.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets