

Maintaining Cisco Service Provider Routing Protocols

Volume 3

Version 1.0

Student Guide



Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
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.

 CCDE, CCENT, CCSI, Cisco Eos, Cisco Explorer, Cisco HealthPresence, Cisco IronPort, the Cisco logo, Cisco Nurse Connect, Cisco Pulse, Cisco SensorBase, Cisco StackPower, Cisco StadiumVision, Cisco TelePresence, Cisco TrustSec, Cisco Unified Computing System, Cisco WebEx, DCE, Flip Channels, Flip for Good, Flip Mino, Flipshare (Design), Flip Ultra, Flip Video, Flip Video (Design), Instant Broadband, and Welcome to the Human Network are trademarks; Changing the Way We Work, Live, Play, and Learn, Cisco Capital, Cisco Capital (Design), Cisco Financial (Stylized), Cisco Store, Flip Gift Card, and One Million Acts of Green are service marks; and Access Registrar, Aironet, AllTouch, AsyncOS, Bringing the Meeting To You, Catalyst, CCDA, CCDP, CCIE, CCIP, CCNA, CCNP, CCSP, CCVP, Cisco, the Cisco Certified Internetwork Expert logo, Cisco IOS, Cisco Lumin, Cisco Nexus, Cisco Press, Cisco Systems, Cisco Systems Capital, the Cisco Systems logo, Cisco Unity, Collaboration Without Limitation, Continuum, EtherFast, EtherSwitch, Event Center, Explorer, Follow Me Browsing, GainMaker, iLUNIX, IOS, iPhone, IronPort, the IronPort logo, Laser Link, LightStream, Linksys, MeetingPlace, MeetingPlace Chime Sound, MGX, Networkers, Networking Academy, PCNow, PIX, PowerKEY, PowerPanels, PowerTV, PowerTV (Design), PowerVu, Prisma, ProConnect, ROSA, SenderBase, SMARTnet, Spectrum Expert, StackWise, WebEx, and the WebEx logo are registered trademarks of Cisco and/or its affiliates in the United States and certain other countries.

All other trademarks mentioned in this document or website 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. (1002R)

DISCLAIMER WARRANTY: THIS CONTENT IS BEING PROVIDED “AS IS.” 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.

Table of Contents

Volume 3

<i>Performance Management for Routing Protocols</i>	3-1
Overview	3-1
Module Objectives	3-1
Auditing Internal Routing	3-3
Overview	3-3
Objectives	3-3
Monitoring Internal Routing Resources	3-4
Monitoring CPU Utilization by IGP	3-7
Monitoring Memory Utilization by IGP	3-15
Monitoring Link Utilization by IGP	3-23
Auditing IGP Configuration	3-28
Summary	3-38
Lesson Self-Check	3-39
Lesson Self-Check Answer Key	3-40
Auditing External Routing	3-41
Overview	3-41
Objectives	3-41
Monitoring External Routing Resources	3-42
Monitoring CPU Utilization by BGP	3-43
Monitoring BGP Memory Utilization	3-61
Monitoring BGP Link Utilization	3-68
Auditing BGP Configuration	3-83
Summary	3-90
Lesson Self-Check	3-91
Lesson Self-Check Answer Key	3-93
Testing Internal Routing	3-95
Overview	3-95
Objectives	3-95
Testing IGP	3-96
Testing OSPF	3-103
Testing IS-IS	3-108
Summary	3-117
Lesson Self-Check	3-118
Lesson Self-Check Answer Key	3-119
Testing External Routing	3-121
Overview	3-121
Objectives	3-121
Testing BGP	3-122
Testing IBGP	3-131
Testing EBG	3-134
Summary	3-142
Lesson Self-Check	3-143
Lesson Self-Check Answer Key	3-144
Module Summary	3-145

<i>Fault Management for Routing Protocols</i>	<i>4-1</i>
Overview	4-1
Module Objectives	4-1
<i>Troubleshooting Routing Issues in the Network</i>	<i>4-3</i>
Overview	4-3
Objectives	4-3
Typical Routing Failures in Service Provider Environments	4-4
Troubleshooting Techniques	4-15
Troubleshooting Typical Failures	4-20
Summary	4-32
Lesson Self-Check	4-33
Lesson Self-Check Answer Key	4-34
<i>Troubleshooting External Routing Issues</i>	<i>4-35</i>
Overview	4-35
Objectives	4-35
Typical External Routing Failures in Service Provider Environments	4-36
Troubleshooting Typical Failures	4-44
Summary	4-56
Lesson Self-Check	4-57
Lesson Self-Check Answer Key	4-58
Module Summary	4-59

Performance Management for Routing Protocols

Overview

Each network device has finite processing, memory, and link resources. In some cases, routing protocols can consume too many resources. In order to make routing protocols more efficient, you should understand and be able to identify performance issues for Border Gateway Protocol (BGP), Open Shortest Path First (OSPF), or Intermediate System-to-Intermediate System (IS-IS) based on the monitoring system. Achieving this efficiency includes proposing changes to improve the performance of internal routing protocols and BGP. This module describes convergence time definition and the main design goals when deploying BGP and internal routing protocols in the network. The module explains tuning options for OSPF, IS-IS, and BGP to improve routing protocol convergence.

Module Objectives

Upon completing this module, you will be able to gather and interpret performance statistics for routing protocols using various tools, such as show commands and Simple Network Management Protocol (SNMP)-based monitoring tools. This ability includes being able to meet these objectives:

- Identify performance issues for OSPF or IS-IS based on monitoring systems, and propose changes to improve the performance of internal routing protocols
- Identify performance issues for BGP based on monitoring systems, and propose changes to improve the performance of BGP
- Identify performance issues for OSPF or IS-IS based on planned and executed tests, and propose changes to improve the performance of internal routing protocols
- Identify performance issues for BGP based on planned and executed tests, and propose changes to improve the performance of BGP

Auditing Internal Routing

Overview

Each network device has a finite CPU, memory, and link resources. In some cases, internal routing protocols can consume too many resources. In order to make routing protocols more efficient, you should understand and be able to identify performance issues for Open Shortest Path First (OSPF) or Intermediate System-to-Intermediate System (IS-IS) based on the monitoring system. You should also be able to propose changes to improve the performance of internal routing protocols.

Objectives

Upon completing this lesson, you will be able to identify performance issues for OSPF or IS-IS. This ability includes being able to meet these objectives:

- Monitor performance issues using SNMP and CLI
- Identify high CPU utilization that is caused by routing protocols using monitoring systems and propose improvements
- Identify high memory utilization that is caused by routing protocols and propose improvements
- Identify high link utilization and improve performance by changing the routing paths
- Perform a manual audit of internal routing protocols and optimize operation of routing protocols

Monitoring Internal Routing Resources

This topic describes the characteristics and requirements for monitoring routing protocols in service provider environments.

Monitoring Internal Routing Resources

- A router consists of a CPU, RAM, storage, and interfaces.
- What constitutes normal behavior on your network?
- Performance should be measured continuously.
- IGP performance can be measured in different ways:
 - Optimality of routing paths
 - Resource utilization (CPU, memory, link)
- Optimization of one aspect can negatively affect other aspects of IGP performance.

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-3-3

Any network device is a specialized computer, which at minimum consists of a CPU, RAM, and storage, allowing it to boot and run the operating system and interfaces. This structure allows for the reception and transmission of network traffic. Once you decide that a problem that you are observing on a given device may be hardware-related, it is worth verifying the operation of these generic components. An essential troubleshooting technique is to compare what is happening on the service provider network to what is expected or to what is normal on the network.

The only way to know what is normal for your network is to measure the network's behavior continuously. Typical statistics to gather are the packet and byte counters on interfaces and device CPU and memory utilization. Performance of interior gateway protocols (IGPs) can be measured in two different ways. On the one hand, you can measure resource utilization, usually CPU and memory. On the other hand, you can measure packet and byte counters to estimate if routing paths for traffic are desired.

When optimizing the performance of IGPs, take special care because optimization of one aspect can negatively affect other aspects of IGP performance. For instance, introducing areas in OSPF can result in suboptimal paths, while using a single area in OSPF can result in greatly increased CPU utilization.

Monitoring Internal Routing Resources (Cont.)

- OSPF and IS-IS characteristics:
 - CPU utilization can be high in large networks.
 - The SPF algorithm's complexity is squarely proportional to the number of routers in an area and level.
 - Memory is typically not an issue (in contrast with BGP).
- Statistics can be read using show commands or SNMP protocol.
- Proactive monitoring of the network allows for a faster response to failures.
- Optimize OSPF and IS-IS using manual audit.

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0—3-4

Routing protocols may introduce significant overhead regarding CPU and memory utilization, especially in service provider networks. In contrast with Border Gateway Protocol (BGP), memory is typically not an issue with OSPF or IS-IS in service provider networks. On the other hand, CPU utilization can be high in large service provider networks. Shortest Path First (SPF) algorithm complexity is squarely proportional to the number of routers in an area or in a level.

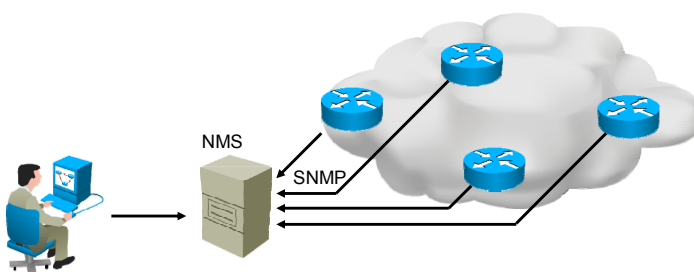
Statistics of interior gateway protocol (IGP) performance can be read using Cisco IOS and Cisco IOS XR commands or collected via the Simple Network Management Protocol (SNMP). The raw data is stored in a database and can then be analyzed or graphed. The network management system (NMS) will continuously check your network device's availability and status. There are various solutions that include the capability to collect statistics via SNMP and graph the results. A well-known example of this type of software is the Open Source Multi Router Traffic Grapher (MRTG) and other products that are based on it.

Proactively monitoring the network allows you to detect abnormal network behavior faster. You can detect possible problems as soon as they occur and that may allow you to diagnose and resolve these problems before they become apparent to the customers. You can analyze and represent collected data by graphing and reporting the performance statistics.

Even if the CPU, memory, and link utilization are not critical, you can optimize OSPF and IS-IS by changing network type on the interfaces and enabling Cisco Nonstop Forwarding (Cisco NSF).

Monitoring Internal Routing Resources (Cont.)

- NMS polls the SNMP agent on the router to obtain statistics.
- Analyzing and representing the results:
 - Graphing
 - Reporting
- Thresholds can be set to trigger a notification process when exceeded.



© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-3.5

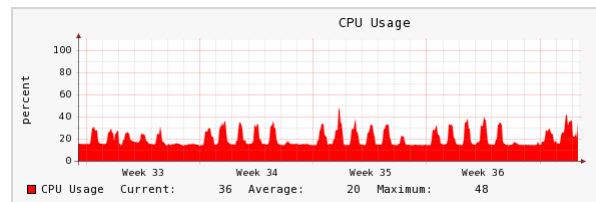
Routers and other network devices keep statistics about the information of their processes and interfaces locally. The SNMP runs a special process, which is called an agent, on a device. Using the SNMP, this agent can be queried to obtain the values of statistics or parameters. By periodically querying or “polling” the SNMP agent on a device, statistics can be gathered and collected over time by an NMS. This data can then be processed and analyzed in various ways. Averages, minimums, or maximums can be calculated; the data can be graphed or thresholds can be set to trigger a notification process when they are exceeded. NMS polls devices periodically to obtain the values of the MIB objects that it is set up to collect.

Monitoring CPU Utilization by IGP

This topic describes the methods for monitoring CPU utilization that is caused by IGPs.

CPU Utilization

- Routers have a main CPU that runs software processes.
- CPU should be measured continuously.
- Define the normal value for CPU utilization.
- Notification should be sent when the threshold is exceeded.
- Nondistributed platforms should not have CPU utilization greater than ~40%.
- Distributed platforms can have large CPU utilization (user data is not affected by higher CPU utilization).



Routers have a main CPU that executes the processes that constitute the Cisco IOS or Cisco IOS XR Software. Processes are scheduled to share the available CPU cycles and take turns executing their code. Because the CPU is a primary resource, it should be measured continuously. Gather CPU statistics at NMS and graph the statistics. Observe CPU utilization for a long period and define which values are still acceptable. You can set threshold to these values, meaning that when CPU utilization exceeds this threshold, notification will be sent.

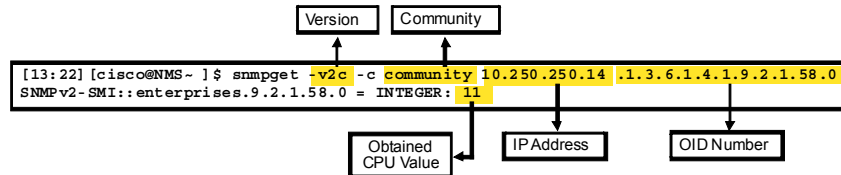
The figure shows a graphed value for the five-minute CPU utilization of a router (taken using the open source tool called Cacti).

You should also consider the differences in the architecture of routers:

- Nondistribute platforms, such as a Cisco 7200 Series Routers, will use the main CPU to process the control, the management, and the data plane. If any one of those planes is causing high CPU utilization, it will also be reflected on other planes. For example, if the BGP process is consuming too much CPU, it may increase the latency of user data packets.
- Distributed platforms, such as the Cisco 7600 Series Routers, are designed to separate the data plane from the control and the management plane. The same scenario, in which the BGP process causes a near 100 percent CPU utilization, will not affect the user data. It is being independently processed by the intelligent line card. Conversely, high CPU utilization—because of interrupts on the route processor—can indicate that the route processor is processing the user packets instead of the line card. This situation can indicate a misconfiguration that results in the line card not being capable of supporting a certain feature and, therefore, is sending packets to the route processor.

CPU Utilization (Cont.)

- MIB is a collection of information organized hierarchically.
- OIDs uniquely identify managed objects in MIB.
- OID: **1.3.6.1.4.1.9.2.1.58.0**:
 - 5-minute exponentially moving average of the CPU busy percentage



© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0—3-8

The SNMP graphing tool periodically polls the SNMP agent, in this case a router, and graphs the obtained values.

MIB is a collection of information that is organized hierarchically. The information is then accessed using a protocol such as SNMP. Object IDs (OID) or Object Identifiers uniquely identify managed objects in an MIB hierarchy. This organization can be depicted as a tree, in which different organizations assign the levels. Top-level MIB OIDs belong to different standard organizations. Vendors define private branches including managed objects for their own products.

OIDs belonging to Cisco are .iso(1).org(3).dod(6).internet(1).private(4).enterprises(1).cisco(9). Output in the figure shows an example of the SNMP application that is called **snmpget** issued on a network management station. Using the **snmpget** application, you can manually obtain a value for a five-minute exponentially moving average of the CPU busy percentage, as the output in the figure shows. You have to specify SNMP version, correct community, IP address of the network device you want to query, and the OID number.

CPU Threshold Notification

- Sends SNMP trap message and a log message when preconfigured CPU thresholds are crossed:
 - Rising threshold
 - Falling threshold

```
process cpu threshold type process rising 80 interval 5 falling 70 interval 5
```

- Sends a rising threshold notification when process CPU utilization exceeds 80 percent for a period of 5 seconds or longer
- Sends a falling threshold notification when process CPU utilization falls below 70 percent for a period of 5 seconds or longer
- Thresholds and alerts can also be implemented with most monitoring servers.

© 2010 Cisco Systems, Inc. All rights reserved.

MS PRP v1.0-3-9

To monitor CPU utilization, you can periodically poll the MIB. However, scheduled polling of the MIB may delay the actions necessary to maximize network availability.

The other option is to enable CPU Threshold Notification, which sends SNMP trap notification and log messages when thresholds are exceeded. Two CPU utilization thresholds are supported:

- **Rising threshold:** This threshold sends a CPU threshold notification when the CPU resource utilization exceeds the configured percentage for the configured time.
- **Falling threshold:** This threshold sends a CPU threshold notification when the CPU resource utilization falls below the configured percentage level for the configured time.

The example in the figure shows how to set a rising CPU threshold notification for the total CPU utilization. When the total CPU utilization exceeds 80 percent for 5 seconds or longer, a rising threshold notification is sent. When the total CPU utilization falls below 70 percent for 5 seconds or longer, a falling threshold notification is sent.

The following are examples of CPU threshold log messages:

Total CPU utilization rises above the rising threshold:

```
*Sep 16 10:04:34.743: %SYS-1-CPURISINGTHRESHOLD: Threshold: Total CPU  
Utilization(Total/Intr): 6%/99%, Top 3 processes(Pid/Util): 176/3%,  
5/0%, 131/0%
```

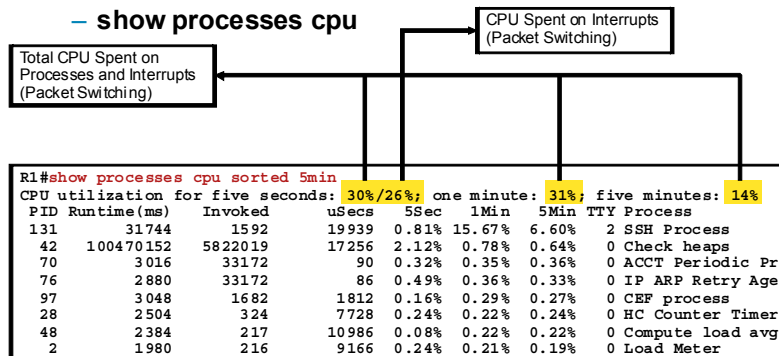
Total CPU utilization falls below the falling threshold:

```
*Sep 16 10:04:59.731: %SYS-1-CPUFALLINGTHRESHOLD: Threshold: Total CPU  
Utilization(Total/Intr) 1%/0%
```

Monitoring CPU Utilization by IGP

- Overview of all processes running in Cisco IOS CLI:

– show processes cpu



- High interrupt-based CPU utilization on distributed platforms typically indicates software processing of packets.

© 2010 Cisco Systems, Inc. All rights reserved.

MS PRP v1.0-3-10

When you are notified that the threshold of CPU utilization that you set is exceeded, you can take a closer look at the problem using command-line interface (CLI) commands. The **show processes cpu** command in the Cisco IOS Software gives you an overview of all processes that are currently running on the router:

- The total amount of CPU time that they have consumed over their lifetime
- Their CPU usage over the last five seconds
- One- and five-minute weighted average of their CPU usage

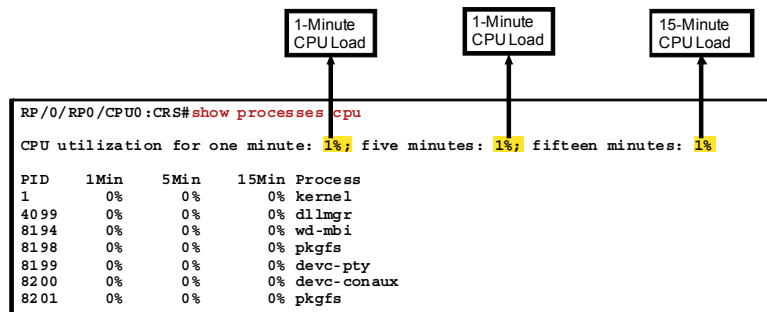
At the top, you will also see the total percentages of available CPU cycles spent. From this information, you can see if the total CPU usage is high or low, and which processes may be causing the CPU load. By default the processes are sorted by process ID, but they can be sorted based on the five-second, one-minute and five-minute averages. In the sample output in the figure, processes are sorted based on five-minute average CPU utilization.

In the example, you can see that over the past minute 31 percent of the available CPU has been used. The “SSH Process” was responsible for roughly half of these CPU cycles (15.67 percent). However, the next process in the sorted list is the “Check heaps” process, which has only consumed 0.78 percent of the total available CPU time over the last minute. The list quickly drops off after that process. The remaining 15 percent of the CPU cycles were spent on packet switching because the CPU that is used to run the operating system processes is also responsible for packet switching.

Once you determined that the CPU load is abnormally high and you decide to investigate further, you will generally have to resort to platform-specific commands to gather more information.

Monitoring CPU Utilization by IGPs (Cont.)

- Overview of all processes running in Cisco IOS XR CLI:
 - show processes cpu**



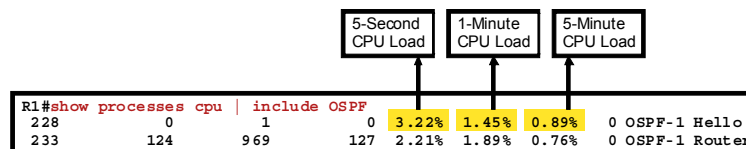
© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-3.11

Cisco IOS XR Software has a slightly different output when issuing the **show processes cpu** command, showing CPU utilization only for 1-, 5- and 15-minute weighted averages of CPU usage.

Monitoring CPU Utilization by IGPs (Cont.)

- CPU utilization caused by IGP on Cisco IOS Software:
 - show processes cpu | include OSPF**
 - show processes cpu | include IS-IS**



© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-3.12

The Cisco IOS Software enables you to check the CPU utilization that OSPF or IS-IS causes by appending a pipe (|) character to the **show processes cpu** command that is followed by the **include** character. Be careful, because the string that follows the keyword **include** is case-sensitive.

Monitoring CPU Utilization by IGP (Cont.)

- CPU utilization caused by IGP on Cisco IOS XR Software:
 - **show processes cpu | include ospf**
 - **show processes cpu | include is-is**

	15-Minute CPU Load	10-Minute CPU Load	15-Minute CPU Load	
RP/0/RP0/CPU0:CRS#				show processes cpu include ospf
295064	0%	0%	0%	ospf_policy_reg_agent
295065	0%	0%	0%	ospfv3_policy_reg_agent
332024	0%	0%	0%	ospf
332025	0%	0%	0%	ospf
332026	0%	0%	0%	ospf_uv
4391170	0%	0%	0%	ospf

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-3.13

The Cisco IOS XR Software enables you to check the CPU utilization that OSPF or IS-IS causes by appending a pipe (|) character to the **show processes cpu** command, followed by the **include** character. Be careful, because the string that follows the keyword **include** is case-sensitive.

Monitoring CPU Utilization by IGP

Solution

- In case of **OSPF**, you can propose the following changes:
 - Enable OSPF with incremental SPF Minor Change
 - Enable the OSPF SPF throttling feature Minor Change
 - Use multiple areas and summarize on ABRs Major Change
 - Use special area types (stub, totally stubby) Major Change
- In case of **IS-IS**, you can propose the following changes:
 - Use multiple levels/areas with summarization Major Change
 - Customize IS-IS throttling feature Minor Change
 - Use IS-IS with incremental SPF Minor Change

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0–314

If you use OSPF as IGP and you experience high CPU utilization that OSPF causes, consider doing the following:

- **Enable the OSPF incremental Shortest Path First (SPF) feature:** OSPF uses an SPF algorithm to compute the shortest path tree. When topology changes occur in an area, the entire tree is recomputed. In many cases, the entire tree need not be recomputed because most of the tree remains unchanged. Incremental SPF allows the system to recompute only the affected part of the tree. This results in faster OSPF convergence and saves CPU resources.
- **Enable the OSPF SPF throttling feature:** The OSPF SPF throttling feature makes it possible to configure SPF scheduling in millisecond intervals and to potentially delay SPF calculations during network instability. SPF is scheduled to calculate the shortest path tree when there is a change in topology. The interval at which the SPF calculations occur is chosen dynamically and is based on the frequency of topology changes in the network. The chosen interval is within the boundary of the user-specified value ranges. If network topology is unstable, SPF throttling calculates SPF scheduling intervals to be longer until topology becomes stable. This scheduling saves CPU resources because SPF recalculation is not done for every topology change.
- **Use multiple areas and summarize networks on Area Border Routers (ABRs):** Cisco recommends having no more than 50 routers in one OSPF area. If you have more than 50 routers in an area and you experience high CPU utilization, you should consider creating additional areas. The advantage of multiarea design is containment of link-state advertisement (LSA) flooding in an area and thus, SPF recalculation not happening in a neighboring area. However, LSA Type 3 is sent for every advertised network to a neighboring area by default. Proper summarization on ABRs is needed to prevent SPF recalculation because of flapping routes in other areas. Use of multiple areas results in lower CPU utilization, but requires redesigning a network.

- **Use special area types, such as stub and totally stubby areas:** Additional LSAs are contained within an area (LSA Type 3 in case of stub, and LSA Types 4 and 5 in case of totally stubby). Only the default route is sent from Area 0 to neighboring areas. Flapping of external and summary routes does not, therefore, result in SPF recalculation.

IS-IS is more scalable than OSPF and is also more efficient in using CPU resources. IS-IS groups updates and sends them as one LSA. The result is fewer LSAs to process. Because OSPF uses networks as nodes in a shortest path tree, each change in topology results in SPF recalculation. IS-IS, on the other hand, uses routers as nodes in a shortest path tree, and only changes on transit interfaces (interface over which an adjacency is established) result in SPF recalculation.

However, if you experience high CPU utilization that IS-IS causes, consider doing the following:

- **Use multiple areas by assigning different network entity title (NET) prefixes to routers:** Configure routers in one area as Level 1 routers, which results in maintaining a link-state database (LSDB) calculating SPF only for that area. Configure routers connecting to other areas as Level 1–2 routers.
- **Customize IS-IS SPF throttling feature:** IS-IS throttles three main events: link-state protocol data unit (PDU) generation, SPF computation, and partial route calculation (PRC). Throttling slows down the frequency of these events during times of network instability, which results in saving CPU resources.

When detecting a scalability limitation because of high CPU utilization that OSPF or IS-IS causes, you can propose optimization techniques that may have different impacts on the running network:

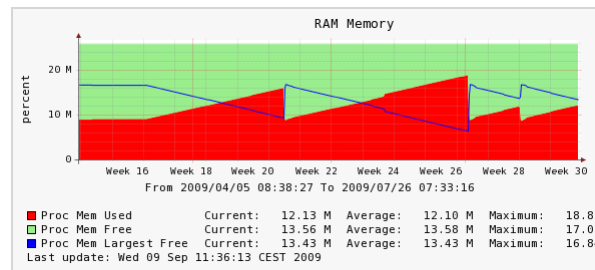
- Enabling incremental SPF does not negatively affect the network (low risk and low impact change). It is still recommended that you consider the feature for the entire network. The change should be tested before being deployed networkwide.
- Introducing areas and summarization is a major change for the network and should be approached with great care. Most service providers today use Multiprotocol Label Switching (MPLS) and many also use MPLS Traffic Engineering (MPLS TE). This feature and others must be considered when proposing changes that will not only affect routing but other dependent features.

Monitoring Memory Utilization by IGP

This topic describes the methods for monitoring memory utilization by IGP.

Memory Utilization

- Memory is a finite resource.
- Not having sufficient free memory can cause memory allocation problems.
- Memory should be measured continuously.
- Notification should be sent when threshold is exceeded.



© 2010 Cisco Systems, Inc. All rights reserved.

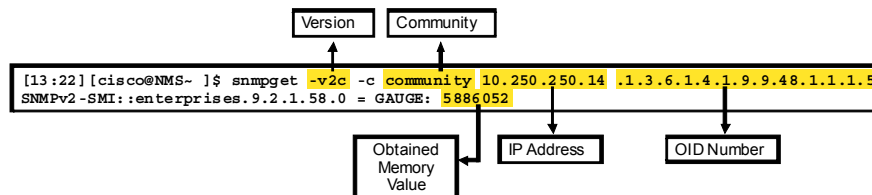
MSRP v1.0-3.16

Like available CPU cycles, memory is a finite resource shared by the various processes that together form the Cisco IOS or IOS XR Software. Memory is divided into different pools and is used for different purposes. The processor pool contains memory that scheduled processes use. The I/O pool is used to temporarily buffer packets during packet switching. Processes allocate and release memory as needed from the processor pool. Generally, there should be more than enough free memory for all the processes to share. However, when you run BGP on your router and you need to carry a complete Internet routing table, you may need more memory than the default memory included with the router.

Not having enough free memory to satisfy the request of a process may result in unpredictable disruptions or failures. Therefore, you should continuously measure the amount of free memory. This measurement can be done using SNMP and setting the threshold. When the used memory exceeds this threshold, notification should be sent.

Memory Utilization (Cont.)

- OID: 1.3.6.1.4.1.9.9.48.1.1.1.5:
 - Indicates the number of bytes from the memory pool that are currently in use by applications on the managed device.
- OID: 1.3.6.1.4.1.9.9.48.1.1.1.6:
 - Indicates the number of bytes from the memory pool that are currently unused on the managed device.



© 2010 Cisco Systems, Inc. All rights reserved.

MS PRP v1.0-3-17

Basically, what the SNMP graphing tool does is periodically poll the SNMP agent, in this case the router, and graph the obtained values. Using the **snmpget** application, you can manually obtain a value for the number of bytes that are currently in use by applications on the managed device from the memory pool. You can also manually obtain a value for the number of bytes that are currently unused on the managed device from the memory pool. The sum of these two numbers is the total amount of memory in the pool. You have to specify the SNMP version, the correct community, the IP address of the network device you want to query, and the OID number.

Memory Threshold Notification

- Sends a log message when preconfigured memory thresholds are crossed:
 - Notification indicates free memory has fallen below a configured threshold.

```
memory free low-watermark processor 20000
```
 - Sends a log message when free processor memory falls below 20,000 KB.
- Most monitoring systems can also be configured to track memory via SNMP and send alerts when thresholds are exceeded.
- Monitoring memory on line cards should also be implemented because there is typically less memory available (such as TCAM limitations).

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-318

The Memory Threshold Notification feature enables the router to send a notification when a free memory falls below a configured threshold.

Notifications are log messages that the router issues. When you specify a memory threshold using the **memory free low-watermark** command, for example, the router issues a notification when available free memory falls below the specified threshold and again once it rises to five percent above it. The following are examples of memory threshold log messages:

Available Free Memory Less Than the Specified Threshold:

```
*Sep 16 10:21:19.227: %SYS-4-FREEMEMLOW: Free Memory has dropped below
low watermark
Pool: Processor Free: 66814056 Threshold: 204800000
```

Available Free Memory Recovered to More Than the Specified Threshold:

```
*Sep 16 10:25:51.195: %SYS-5-FREEMEMRECOVER: Free Memory has recovered
above low watermark
Pool: Processor Free: 113815892 Threshold: 51200000
```

In addition to monitoring the memory on route processors, it is important to monitor the capabilities of line cards, for example:

- Cisco 7600 Series Routers can hold more than a million routes in their route processor memory. However, they have limited memory for routing entries in the Ternary Content Addressable Memory (TCAM), which is typically one million IPv4 routes.
- The Cisco Carrier Routing System (Cisco CRS-1) series routers can also hold many millions of routes in the route processor but are limited to two million routes on the Modular Services Card (MSC).

The memory limitations of line cards may become more apparent with the introduction and growth of IPv6 in the Internet.

Monitoring Memory Utilization by IGPs

- Overview of memory consumption in Cisco IOS CLI:
 - show processes memory**
 - Shows memory in bytes, after the Cisco IOS Software is loaded.

```

R0#show processes memory
Processor Pool Total: 37544320 Used: 24968952 Free: 12575368
I/O Pool Total: 16777216 Used: 5135920 Free: 11641296

PID TTY Allocated Freed Holding Getbufs Retbufs Process
0 0 43004120 15623216 24823160 0 0 *Init*
0 0 12052 105348 12052 0 0 *Sched*
0 0 13128 343444 716 180836 180836 *Dead*
1 0 108956 13968 116160 0 0 Chunk Manager
  
```

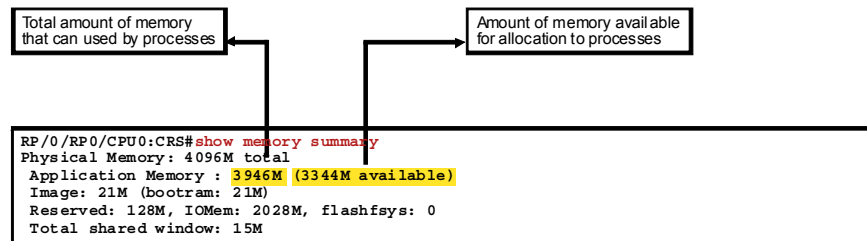
© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-3-19

When you are notified that the threshold of memory utilization that you set has been exceeded, you can take a closer look at the problem using CLI commands. The **show processes memory** command in the Cisco IOS Software gives you an overview of all processes that are currently running on the router and the amount of memory they use. At the top, you also see the total used and the free amount of processor pool and I/O pool memory in bytes, after the Cisco IOS Software has been loaded. From the printout, you can see if the amount of free memory is high or low and which processes may be using it.

Monitoring Memory Utilization by IGP's (Cont.)

- Overview of memory consumption in Cisco IOS XR CLI:
 - **show memory summary**



The Cisco IOS XR Software enables you to verify memory utilization using the **show memory summary** command. The printout shows the total amount of physical memory, total amount of memory that processes can use, and the amount of available memory for allocation to processes.

Monitoring Memory Utilization by IGPs (Cont.)

- Memory utilization caused by IGP on Cisco IOS software:
 - **show processes memory | include OSPF**
 - **show processes memory | include IS-IS**

```
R0#show processes memory
Processor Pool Total: 37544320 Used: 24968952 Free: 12575368
I/O Pool Total: 16777216 Used: 5135920 Free: 11641296

PID TTY Allocated Freed Holding Getbufs Retbufs Process
0 0 43004120 15623216 24823160 0 0 *Init*
<...output omitted...>

R1#show processes memory | include OSPF
153 0 1208 0 11380 0 0 OSPF-1 Hello
236 0 19280 760 18468 0 0 OSPF-1 Router
```

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-321

The Cisco IOS Software enables you to check the memory utilization that OSPF or IS-IS causes by appending a pipe (|) character to the **show processes memory** command, which is followed by the **include** character. Be careful, because the string that follows keyword **include** is case-sensitive.

In the sample printout, you can see the amount of memory in bytes allocated to OSPF processes.

Monitoring Memory Utilization by IGP (Cont.)

- CPU utilization caused by IGP on Cisco IOS XR Software:
 - **show processes memory | include ospf**
 - **show processes memory | include is-is**

Size of Dynamically Allocated Memory



```
RP/0/RP0/CPU0:CRS#show processes memory
JID   Text      Data      Stack    Dynamic  Process
60    65536     32768     61440    38117376 eth_server
168   4096      0         24576    13246464 fgid_server
<...output omitted...>
```



```
RP/0/RP0/CPU0:CRS#show processes memory | include ospf
398   4096      0         61440    3506176  ospf
299   4096      0         61440    3022848  ospf
395   4096      0         65536    2719744  ospf
394   4096      0         65536    2719744  ospf
303   4096      0         16384    225280   ospfv3_policy_reg_agent
300   4096      0         16384    225280   ospf_policy_reg_agent
301   4096      0         12288    163840   ospf_uv
```

The Cisco IOS XR Software enables you to check the memory utilization that OSPF or IS-IS causes by appending a pipe (|) character to the **show processes memory** command, which is followed by the **include** character. Be careful, because the string that follows keyword **include** is case-sensitive.

In the sample printout, you can see the amount of memory in bytes dynamically allocated to OSPF processes.

Monitoring Memory Utilization by IGP

Solution

- In case of OSPF:
 - Use multiple areas and summarize on ABRs
 - Use special area types (stub, totally stubby)
- In case of IS-IS, use multiple areas and summarize on Level 1–2 routers.
- OSPF and IS-IS are typically not carrying excessive amounts of routing information.
- BGP is usually the cause of memory exhaustion.
- Upgrade router with more memory:
 - On route processor
 - On line card (for example, replace 3C with 3CXL modules on Cisco 7600 Series Routers)

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0–323

If you experience high memory utilization when using OSPF, consider doing the following:

- **Use multiple areas and summarize on ABRs:** Route summarization is important. It reduces OSPF LSA flooding and link-state database and routing table sizes, which reduce memory and CPU utilization on the routers. However, this solution often leads to redesigning and readdressing a network. This repeated work is one of the reasons why it is important, from the beginning, to design network and addressing to support multiple areas and summarization on ABRs.
- **Use special area types, such as stub and totally stubby areas:** Instead of having external and summary routes on routers, only a default route can be present on routers in stub and totally stubby areas. Only summary routes are sent into a stub area, and external routes are suppressed. Instead, a default route is sent into a stub area by an ABR. In totally stubby areas, summary and external routes are suppressed and the default route is sent to a totally stubby area instead. This way, memory can be preserved because of a smaller link-state database and routing table.

If you use IS-IS as IGP and you experience high memory utilization, you can do the following:

- **Use multiple areas and summarize on Level 1–2 routers:** Configure routers in one area as Level 1 routers, which results in maintaining an LSDB and calculating SPF only for one area. Configure routers connecting to other areas as Level 1–2 routers. Level 2 routes are not advertised to Level 1 routers by Level 1–2 routers. The default route is sent to Level 1 routers instead. This configuration reduces link-state database and routing table size on Level 1 routers. To reduce link-state database and routing table size on Level 1–2 and Level 2 routers, configure summarization on Level 1–2 routers.

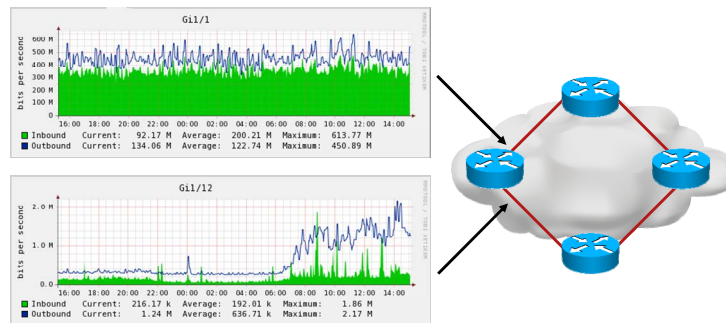
Sometimes networks cannot be further optimized. In this case, the only option is to upgrade the router with additional memory.

Monitoring Link Utilization by IGP

This topic describes how to monitor link utilization and how to improve network performance by changing routing paths.

Link Utilization

- Some links may be overutilized, some links may be underutilized.
- Link utilization should be measured continuously.
- Use SNMP, CLI, or both to determine link utilization.
- Modify path selection to balance traffic.



© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-3-25

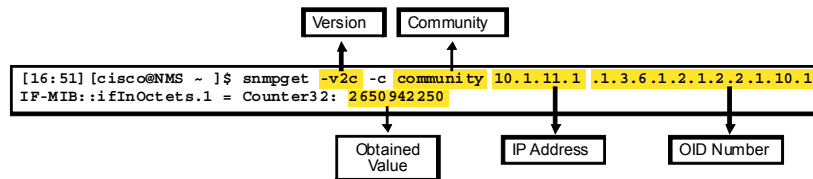
It is very unlikely that the use of IGP will diminish free bandwidth on links, as can be the case with CPU and memory resources. On the other hand, you can measure link utilization to determine if routing paths for traffic are desired.

In service provider networks, some links can be over-utilized and others can be under-utilized. To determine utilization, continuously monitor link utilization on all critical links using SNMP. When you determine high utilization on a specific link, you can further investigate utilization using CLI.

The figure shows a sample network, in which two paths are available from one destination to another. Continuous monitoring and visualization of link utilization using SNMP shows that one link is highly utilized (around 300 Mb/s) and another is under-utilized (less than 1 Mb/s). In this case, routing paths should be changed to equally distribute traffic over both links.

Link Utilization (Cont.)

- OID: 1.3.6.1.2.1.2.2.1.16:
 - Indicates the total number of octets transmitted out of the interface, including framing characters.
- OID: 1.3.6.1.2.1.2.2.1.10:
 - Indicates the total number of octets received on the interface, including framing characters.



© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-326

Basically, what the SNMP graphing tool does is periodically poll the SNMP agent, in this case a router, and graph obtained values. Using the **snmpget** application, you can manually obtain the total number of bytes received on the interface or total number of bytes transmitted out of the interface. To obtain these numbers you have to specify the SNMP version, the correct community, the IP address of the network device you want to query, and the OID number.

To calculate the data rate on a link, periodically obtain the value from a router, subtract a previously obtained value from the newly obtained value, and divide the difference with the time interval. If you divide the calculated number with the available bandwidth on an interface (specified by the **bandwidth** command), the result is a percentage of link utilization.

Monitoring Link Utilization

- Link utilization in Cisco IOS CLI:

- **show interface**

```
R1#show interfaces GigabitEthernet1/0
GigabitEthernet1/0 is up, line protocol is up
<...output omitted...>
Last input 00:00:04, output 00:00:07, output hang never
Last clearing of "show interface" counters never
Input queue: 0/75/0/32 (size/max/drops/flushes); Total output drops: 0
Queueing strategy: fifo
Output queue: 0/40 (size/max)
5 minute input rate 1000 bits/sec, 1 packets/sec
5 minute output rate 1000 bits/sec, 2 packets/sec
 930 packets input, 105731 bytes, 0 no buffer
Received 28 broadcasts, 0 runts, 0 giants, 0 throttles
 0 input errors, 0 CRC, 0 frame, 0 overrun, 0 ignored
 0 watchdog, 0 multicast, 0 pause input
 0 input packets with dribble condition detected
1034 packets output, 118159 bytes, 0 underruns
 0 output errors, 0 collisions, 1 interface resets
 0 babbles, 0 late collision, 0 deferred
 0 lost carrier, 0 no carrier, 0 pause output
 0 output buffer failures, 0 output buffers swapped out
```

When you notice that the link utilization is not equally distributed over links, take a closer look at the problem using CLI commands. The **show interfaces** command in the Cisco IOS Software lists a number of key statistics that should be checked:

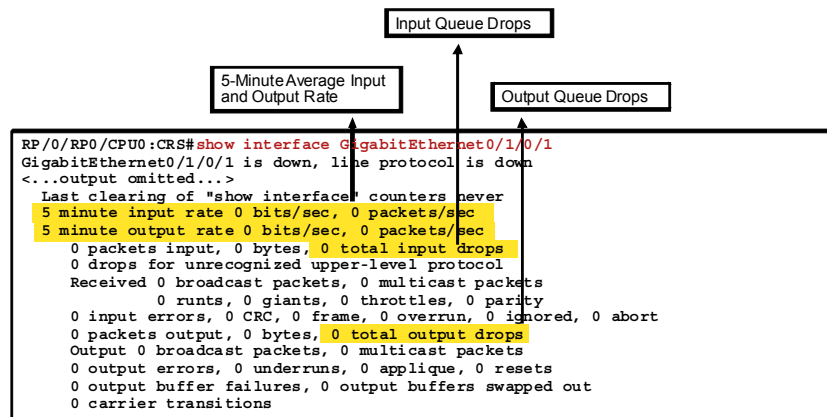
- **Input queue drops:** Input queue drops signify that, at some point, more traffic was delivered to the router than it could process. This drop does not necessarily indicate a problem, as it could be normal during traffic peaks. However it could be an indication that the CPU is not able to process packets in time. If this number is constantly high, it is worth trying to spot at which moments these counters are increasing and how these increases relate to CPU usage.
- **Output queue drops:** Indicates that packets were dropped due to congestion on the interface. Seeing output drops is normal for any point where the aggregate input traffic is higher than the output traffic. During traffic peaks, packets will be dropped if traffic is delivered to the interface faster than it can be sent out. However, even if this behavior is considered to be normal, it leads to packet drops and queuing delays. Constantly seeing output drops indicates that quality of service (QoS) should be implemented; otherwise, delay- and drop-sensitive applications may suffer from performance issues.
- **Five-minute input rate:** Indicates the average number of bits and packets that are received on the interface per second in the last five minutes. By comparing this number to the bandwidth of the interface (configured by the **bandwidth** command), you can calculate link utilization in the inbound direction.
- **Five-minute output rate:** Indicates the average number of bits and packets that are sent by the interface per second in the last five minutes. By comparing this number to the bandwidth of the interface (configured by the **bandwidth** command), you can calculate link utilization in the outbound direction.

If you experience high link utilization, you should consider changing routing patterns to divert some traffic over other links.

Sometimes it is recommended to reset the interface counters using the **clear counters** command to ensure that you are observing recent data.

Monitoring Link Utilization (Cont.)

- Link utilization in Cisco IOS XR CLI:
 - **show interface**



© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-328

To verify link utilization in Cisco IOS XR Software, use the **show interfaces** command. The output format differs slightly from the one in Cisco IOS Software, but it shows the same statistics.

Link Utilization

Solution

Propose changes to optimize link utilization in the network:

- IGP metric modification:
 - OSPF: Change interface cost or reference bandwidth Minor Change
 - IS-IS: Change interface metric Minor Change
- Equal-cost load balancing Minor Change
- MPLS traffic engineering Major Change

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0–329

If you experience high link utilization over one link and you use OSPF as the internal routing protocol, you can change routing paths by changing OSPF cost on specific interfaces or by changing reference bandwidth. For example, you could change the metric in a way that traffic going to some destinations would go over one path, and traffic going to other destinations would go over another path.

In the case of IS-IS, which does not differentiate between different interfaces by default, it is even more desirable to change the IS-IS metric on each interface. To influence path selection that IS-IS does, it might also be necessary to use a wide-style metric, which supports more granular path selection.

If load balancing is desired in the network, you can also change the metric of different paths to achieve equal-cost paths. In this case, load balancing can occur.

In some cases, it is not possible to change routing patterns by changing the metric on interfaces. In this case, the only solution is MPLS TE, in which you can specify the exact path that should be taken by packets going to a specific destination.

Auditing IGP Configuration

This topic describes how to further optimize performance of internal routing protocols.

Auditing IGP Configuration

Optimize IGP configuration:

- Change network type on broadcast point-to-point links to point-to-point Minor Change
- Enable adjacency change logging Minor Change
- Enable Cisco Nonstop Forwarding feature Major Change

© 2010 Cisco Systems, Inc. All rights reserved. MSPRP v1.0-331

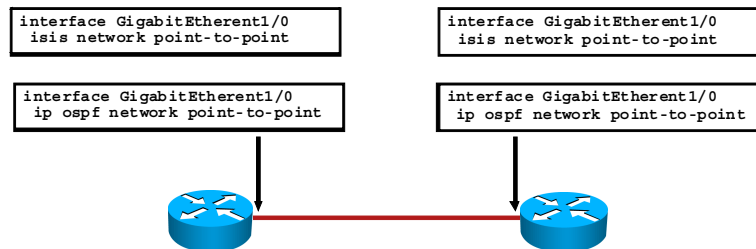
Even if you do not experience high CPU or the memory utilization that IGP's cause, perform an IGP configuration audit to verify that the configuration follows the best practices.

Be sure to do the following:

- Change the network type to **point-to-point** on broadcast point-to-point links to improve network convergence times.
- Enable adjacency change logging to see if there are any flapping neighbors.
- Enable Cisco NSF on devices that use two route processors to minimize the amount of time a network is unavailable to its users following a route processor switchover.

Changing Network Type

- Default network type on broadcast media is broadcast.
- Change network type on broadcast point-to-point links to point-to-point.
- This provides performance improvements to the network convergence.



© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-332

Point-to-point and broadcast are the two predominant circuit types that link-state routing protocols such as IS-IS and OSPF use. They are treated differently with respect to establishing neighbor adjacencies, flooding link-state information, representing the topology, and calculating the SPF and protocol packets. The most important differences are that broadcast circuits utilize the concept of a designated router and are represented topologically as virtual nodes in the network topology graph.

Even though a broadcast circuit is meant to process more than two devices, there are cases in which only two routers are connected over either the physical or logical LAN segment. In such cases, the link-state routing protocols will treat the media as a broadcast circuit by default. It makes more sense to treat the connection between the two routers as a point-to-point circuit.

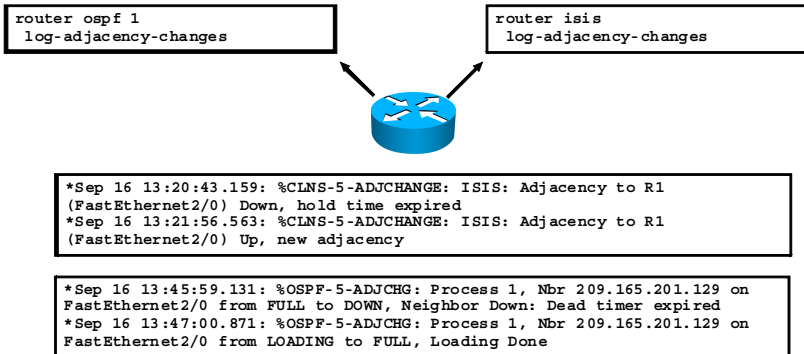
Being able to treat a broadcast media as a point-to-point circuit provides the benefit of reduction in the amount of information routing protocols must carry and manage. The designated router, in the case of OSPF, and the Designated Intermediate System (DIS), in the case of IS-IS election, can be omitted. This omission also simplifies SPF calculation.

To change the OSPF network type to point-to-point, use the **ip ospf network point-to-point** command in interface configuration mode. Make sure that you change the OSPF network type on both routers that are connected to the same segment.

To change the IS-IS network type to point-to-point, use the **isis network point-to-point** command in the interface configuration mode. Make sure that you change the IS-IS network type on both routers that are connected to the same segment.

Enabling Adjacency Logging

- Neighbor going up and down results in LSA flooding and SPF recalculations.
- Neighbor state logging should be enabled.
- Logging messages are sent to the configured logging output.



© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-333

A flapping link in a network results in neighbor adjacencies going up and down. Consequently, there are LSAs flooded all over the network that triggers SPF recalculation on all routers. It is important to log any messages regarding neighbor state changes to see any changes immediately and to take the appropriate actions.

When using OSPF or IS-IS as IGP, make sure that you enable adjacency logging to see any changes in neighbor state. Logging messages are sent to the configured logging output, such as console, syslog server, or system buffer.

To configure the router to send a log message when there is a neighbor change, use the **log-adjacency-changes** command in router configuration mode. In the sample printout, you can see two examples where IS-IS and OSPF adjacencies went down because of an expired hold timer and back up one minute later, which could indicate a flapping link.

Cisco Nonstop Forwarding

- Cisco NSF is applicable in platforms with dual route processors.
- Cisco NSF allows forwarding of data packets while the routing protocol information is being restored.
- Cisco NSF works together with stateful switchover (SSO).
- Cisco NSF is supported by:
 - Routing protocols (OSPF, IS-IS, EIGRP, BGP)
 - Forwarding operation (Cisco Express Forwarding)
- Device has to be **Cisco NSF-capable**

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0–334

Cisco NSF works with the Stateful Switchover (SSO) to minimize the amount of time a network is unavailable to its users following a switchover. The main objective of NSF is to continue forwarding IP packets following a route processor switchover in platforms with dual route processors.

Usually, when a networking device restarts, 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 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 Cisco 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 route processor during a switchover.

Cisco NSF is supported by the BGP, Enhanced Interior Gateway Routing Protocol (EIGRP), OSPF, and IS-IS protocols for routing and by Cisco Express Forwarding for forwarding. Of the routing protocols, BGP, EIGRP, OSPF, and IS-IS have been enhanced with Cisco NSF-capability and awareness. This means 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 Cisco NSF-compatible software. A device is said to be Cisco NSF-capable if it has been configured to support Cisco NSF; therefore, it would rebuild routing information from Cisco NSF-aware or Cisco NSF-capable neighbors.

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

Cisco NSF Overview

- One RP is active, one is standby.
- Standby RP synchronizes FIB and adjacency table with the active RP.
- Upon switchover, new active RP uses the old FIB and adjacency table to forward packets while the routing protocol converges.
- Routing protocol (OSPF, IS-IS) has to:
 - Relearn neighbors without causing a reset of neighbor relationship
 - Synchronize link-state database
- As the routing protocol starts to repopulate RIB, it updates Cisco Express Forwarding.
- Neighboring device has to be **NSF-aware**.

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0–335

Cisco NSF always runs together with SSO. In specific Cisco networking devices that support dual route processors, SSO establishes one of the route processors as the active processor while the other route processor is designated as the standby processor. SSO then synchronizes FIB and the adjacency table between them. A switchover from the active to the standby processor occurs when the active route processor fails, is removed from the networking device, or is manually taken down for maintenance. During the switchover, the new active route processor 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 run only on the active route processor, and they receive routing updates from their neighbor routers. Routing protocols do not run on the standby route processor. If NSF is not configured, a neighboring device would tear down the adjacency during a switchover, resulting in a routing flap that would spread across the network.

Using Cisco NSF, the routing protocols on a Cisco NSF-capable device requests that the Cisco NSF-aware neighbor devices send state 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 it will send link-state information to synchronize the link-state database on a Cisco NSF-capable device instead.

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

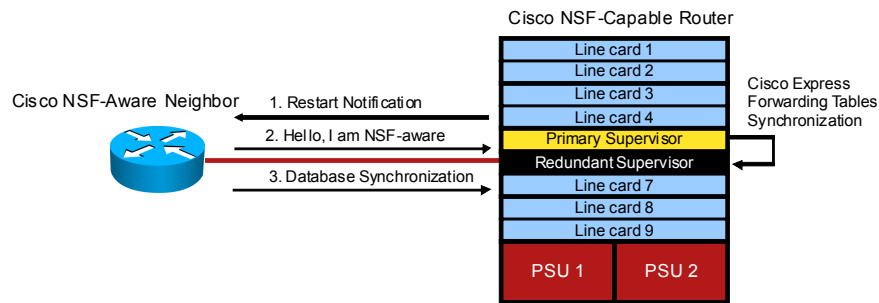
Cisco NSF Overview (Cont.)

Cisco NSF-aware neighbors:

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

Cisco NSF-capable router:

- Data is forwarded in hardware based on pre-switchover Cisco Express Forwarding information while routing protocols reconverge
- Rebuilds Layer 3 routing protocol database from neighbor



© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-336

When a switchover is triggered, a packet continues based on last-known FIB and adjacency entries while the standby route processor takes over. A routing protocol on Cisco NSF-capable routers notifies the Cisco NSF-aware neighboring router that the adjacency is being reacquired and that the Cisco NSF-aware router should not reinitialize the neighbor relationship. Upon receiving a restart notification, protocol-specific procedures occur to allow adjacencies to be maintained. Neighbors that are not Cisco NSF-aware ignore the restart indication and bring down the adjacency.

Then the synchronization of protocol-specific databases occurs. The routing protocol processes rebuild their database using database information from Cisco NSF-aware neighbors. After the routing databases are synchronized and the RIB is repopulated with new routing entries, corresponding Cisco Express Forwarding tables are updated.

Cisco NSF Configuration for OSPF

- Cisco NSF configuration for OSPF on Cisco IOS Software:



- Cisco NSF configuration for OSPF on Cisco IOS XR Software:



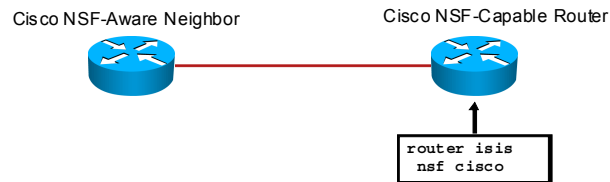
To configure OSPF for Cisco NSF support on a Cisco NSF-capable Cisco IOS router, use the **nsf cisco** command in router OSPF configuration mode. This command is entered on only Cisco NSF-capable routers. Cisco NSF awareness is enabled by default when a supporting version of Cisco IOS Software is installed on a router that supports Cisco NSF capability or Cisco NSF awareness.

The configuration of OSPF for Cisco NSF on a Cisco IOS XR router is the same as on the Cisco IOS router.

Note In a multivendor network environment, use the **nsf ietf** command instead.

Cisco NSF Configuration for IS-IS

- Cisco NSF configuration for IS-IS on Cisco IOS Software:



- Cisco NSF configuration for IS-IS on Cisco IOS XR Software:



© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0—338

To configure IS-IS Cisco NSF support on a Cisco NSF-capable Cisco IOS router, use the **nsf cisco** command in `router isis` configuration mode. This command is entered only on Cisco NSF-capable routers. Cisco NSF awareness is enabled by default when a supporting version of Cisco IOS Software is installed on a router that supports Cisco NSF capability or Cisco NSF awareness.

The configuration of IS-IS for Cisco NSF on a Cisco IOS XR router is the same as on the Cisco IOS router.

Note In a multivendor network environment, use the **nsf ietf** command instead.

Cisco Nonstop Forwarding Benefits

- Improved network availability
- Overall network stability
- Neighboring routers do not detect link flapping
- Routing flaps prevented
- No loss of user sessions

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0–339

The Cisco NSF feature has several benefits, including the following:

- **Improved network availability:** Cisco NSF continues forwarding network traffic and application state information so that user session information is maintained after a switchover.
- **Overall network stability:** Network stability may be improved with the reduction in the number of route flaps that had been created when routers in the network failed and lost their routing tables.
- **Neighboring routers do not detect link flapping:** Because the interfaces remain up across a switchover, neighboring routers do not detect a link flap (that is, the link does not go down and come back up).
- **Prevents routing flaps:** Because SSO continues forwarding network traffic in the event of a switchover, routing flaps are avoided.
- **No loss of user sessions:** User sessions that are established prior to the switchover are maintained.

Summary

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

Summary

- Each network device has finite hardware resources.
- Measure CPU, memory, and link resources using SNMP continuously.
- Take actions to optimize IGP operations in cases of high CPU and memory utilization caused by IGPs.
- Change routing paths in cases of high or unequal link utilization.
- Perform a manual audit and make sure that you follow the best implementation practices when implementing OSPF or IS-IS as IGP.

Lesson Self-Check

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

- Q1) How can you continuously measure CPU, memory and link utilization on Cisco routers? (Source: Monitor Internal Routing Resources)
- A) by periodically polling the SNMP agent on a router
 - B) using CLI commands
 - C) using log messages
 - D) by setting thresholds to trigger notifications
- Q2) Which three measures can be taken when OSPF uses too much CPU resources? (Choose three.) (Source: Monitor CPU Utilization by IGP)
- A) use multiple areas
 - B) enable OSPF Incremental SPF feature
 - C) use special area types
 - D) disable OSPF SPF Throttling feature
- Q3) Which two are ways that you can change routing paths to equally distribute load over links when using OSPF or IS-IS as IGP? (Choose two.) (Source: Monitoring Link Utilization by IGP)
- A) modify metric
 - B) configure unequal-cost load balancing
 - C) use multiple areas
 - D) configure MPLS TE
- Q4) Which two are ways to further optimize IGP operations? (Choose two.) (Source: Auditing IGP Configuration)
- A) by enabling adjacency logging
 - B) by changing network type on broadcast point-to-point links to broadcast
 - C) by configuring Cisco NSF on devices with one route processor
 - D) by configuring Cisco NSF on devices with two route processors
- Q5) Which two statements describe what happens on a Cisco NSF-capable router during the switchover? (Choose two.) (Source: Auditing IGP Configuration)
- A) Cisco NSF-capable router helps the NSF-aware router to restart.
 - B) Data is forwarded in hardware that is based on Cisco Express Forwarding information.
 - C) Cisco NSF-capable router rebuilds routing database from NSF-aware neighbor
 - D) Cisco NSF-capable router sends routing table to NSF-aware router to synchronize routing tables.

Lesson Self-Check Answer Key

- Q1) A
- Q2) A, B, C
- Q3) A, D
- Q4) A, D
- Q5) B, C

Auditing External Routing

Overview

Each network device has a finite CPU, memory, and link resources. In some cases, external routing protocols can consume too many resources. In order to make routing protocols more efficient, you should understand and be able to identify performance issues for Border Gateway Protocol (BGP) based on the monitoring system. You should also be able to propose changes to improve BGP performance.

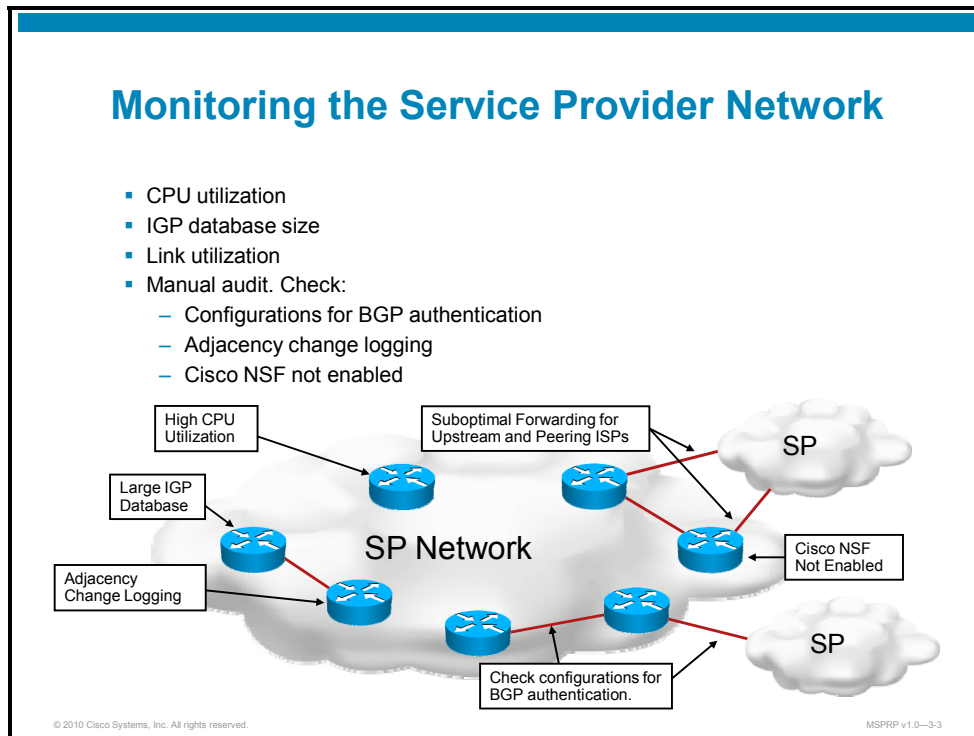
Objectives

Upon completing this lesson, you will be able to identify performance issues in the BGP based on monitoring systems and propose changes to improve BGP performance. This ability includes being able to meet these objectives:

- Describe the characteristics and requirements for monitoring routing protocols in service provider environments
- Describe the methods for monitoring CPU utilization by BGP
- Describe the methods for monitoring BGP memory utilization
- Describe the methods for monitoring BGP link utilization
- Perform a manual audit of BGP and optimize BGP operation

Monitoring External Routing Resources

This topic describes the characteristics and requirements for monitoring routing protocols in service provider environments.



In the service provider network, certain parameters must be monitored:

- CPU utilization
- Interior routing protocol (IGP) database size and memory utilization
- Link utilization
- Manual audit

Monitoring CPU Utilization by BGP

This topic describes the methods for monitoring CPU utilization by BGP.

Symptoms of High CPU Utilization

- High percentages in the **show processes cpu** command output
- Slow performance
- Services on the router fail to respond, for instance:
 - Slow response in Telnet or unable to connect Telnet to the router.
 - Slow response on the console.
 - Slow or no response to ping.
 - Router does not send routing updates to other routers.

© 2010 Cisco Systems, Inc. All rights reserved.

MS PRP v1.0—3-5

The list describes common symptoms of high CPU utilization. If you notice any of these symptoms, follow the troubleshooting steps in this topic to address the problem.

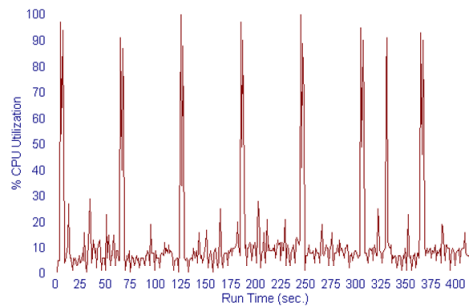
- High percentages in the **show processes cpu** command output

If you have the output of a **show processes cpu** command from your Cisco device, you can use Output Interpreter to display potential issues and fixes. To use Output Interpreter, you must be a registered customer, be logged in, and have JavaScript enabled.
- Slow performance
- Services on the router fail to respond, for instance:
 - Slow response in Telnet or unable to connect Telnet to the router.
 - Slow response on the console.
 - Slow or no response to ping.
 - Router does not send routing updates to other routers.

Monitoring CPU Utilization

Monitoring methods:

- SNMP
- CLI



```
router#show processes cpu
CPU utilization for five seconds: 8%/4%; one minute: 6%; five minutes: 5%
PID Runtime(uS)   Invoked  uSecs   5Sec   1Min   5Min  TTY Process
1      384        32789    11      0.00%  0.00%  0.00%  0 Load Meter
2     2752        1179    2334    0.73%  1.06%  0.29%  0 Exec
3    318592       5273   60419    0.00%  0.15%  0.17%  0 Check heaps
4         4         1    4000    0.00%  0.00%  0.00%  0 Pool Manager
5     6472        6568    985    0.00%  0.00%  0.00%  0 ARP Input
6    10892       9461   1151    0.00%  0.00%  0.00%  0 IP Input
7    67388      53244   1265    0.16%  0.04%  0.02%  0 CDP Protocol
8   145520     166455    874    0.40%  0.29%  0.29%  0 IP Background
```

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-3.6

To monitor CPU utilization, use the following methods:

- Simple Network Management Protocol (SNMP)
- Command-line interface (CLI)

Monitoring CPU Utilization (Cont.)

Devices with single CPU:

- **cpmCPUTotal5secRev (.1.3.6.1.4.1.9.9.109.1.1.1.1.6):** The overall CPU busy percentage in the last five-second period
- **cpmCPUTotal1minRev (.1.3.6.1.4.1.9.9.109.1.1.1.1.7):** The overall CPU busy percentage in the last one-minute period
- **cpmCPUTotal5minRev (.1.3.6.1.4.1.9.9.109.1.1.1.1.8):** The overall CPU busy percentage in the last five-minute period

Devices with multiple CPU:

- **cpmCPUTotal5minRev (.1.3.6.1.4.1.9.9.109.1.1.1.1.5):** The overall CPU busy percentage in the last five-minute period

© 2010 Cisco Systems, Inc. All rights reserved.

MS PRP v1.0-3-7

Critical router functions, like routing protocol processing and process packet switching, are processed in memory and share the CPU. Thus, if CPU utilization is very high, it is possible that a routing update cannot be processed or a process-switching packet can be dropped. From the CISCO-PROCESS-MIB, the cpmCPUTotal5minRev MIB object value reports the percentage of the processor in use over a five-minute average.

The cpmCPUTotal5minRev MIB object provides a more accurate view of the router's performance over time than the MIB objects cpmCPUTotal1minRev and cpmCPUTotal5secRev. These MIB objects are not accurate, because they look at CPU at one-minute and five-second intervals, respectively.

- **cpmCPUTotal5secRev (.1.3.6.1.4.1.9.9.109.1.1.1.1.6):** The overall CPU busy percentage in the last five-second period. This object uses the value range to (0..100).
- **cpmCPUTotal1minRev (.1.3.6.1.4.1.9.9.109.1.1.1.1.7):** The overall CPU busy percentage in the last one-minute period. This object uses the value range to (0..100).
- **cpmCPUTotal5minRev (.1.3.6.1.4.1.9.9.109.1.1.1.1.8):** The overall CPU busy percentage in the last five-minute period. This object uses the value range to (0..100).

If your Cisco IOS device has several CPUs, you must use CISCO-PROCESS MIB and its object cpmCPUTotal5minRev. This structure shows CPU statistics for different physical entities in the router, like different CPU chips, groups of CPUs, or CPUs in different modules and cards.

Proposed Changes for CPU Utilization

External routing = BGP

- BGP uses a number of processes.
- Although BGP was designed to scale, it can cause excessive CPU utilization:
 - Overtuning to speed up BGP convergence
 - Periodic CPU spikes because of BGP scanning
 - Excessive number of updates (flapping in the Internet)
- Available BGP mechanisms to optimize BGP:
 - Route flap dampening
 - Increasing scan timers
 - Increasing BGP advertisement interval
 - Increasing BGP keepalive intervals

Major Change

Major Change

Major Change

Major Change

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-3-8

You can propose changes in the BGP design and implementation to improve CPU utilization.

- Route Flap Dampening can minimize the impact of frequently flapping routes by dampening them for a time.
- Scan timers can be increased to make sure that the scan process completes before the next run of the scan is scheduled.
- BGP advertisement interval can be increased to make sure that multiple changes are packed into a single update. The default BGP advertisement intervals are designed to ensure stability (5 seconds for internal BGP sessions and 30 seconds for external BGP sessions).
- BGP keepalives can be set to high enough values to ensure that the BGP sessions can be maintained. The default BGP keepalive timer is designed for stability (60 seconds) but is often reduced to lower values to improve failure detection. Short keepalive timers, in combination with many BGP sessions, can cause an increase in CPU utilization.

Although none of the listed mechanisms normally causes any disruption in routing and traffic flow, a misconfiguration using timers that are too aggressive can exhibit negative effects on the network. Some negative effects could also become apparent at some later date. For example, dampening routes from upstream service providers, because of a few flaps during maintenance that is performed by an upstream service provider, can cause Internet destinations to be unnecessarily unreachable for a longer period.

BGP Route Dampening

- Designed to reduce router processing load caused by unstable routes
- Prevents sustained routing oscillations without affecting other well-behaved routes
- Defined in RFC 2439: *BGP Route Flap Dampening*
- Designed to help minimize the number of BGP updates
- Other update reduction tools:
 - Batching BGP updates
 - Per-neighbor update timers

© 2010 Cisco Systems, Inc. All rights reserved.

MS PRP v1.0—3-9

BGP is the only routing protocol that is designed for large internetworks with the specific intention of carrying many prefixes. Several mechanisms that are built into BGP 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 5-second timer for internal neighbors and a 30-second timer 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 repeatedly (available, unavailable, available, unavailable, and so on) 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 assists in reducing 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 that is called route flap dampening was created to reduce route update processing requirements by suppressing unstable routes.

BGP Route Dampening (Cont.)

- Minimizes the amount of BGP update processing in the Internet by suppressing unstable (flapping) routes
- Does not suppress routes that occasionally flap
- Suppresses routes that are likely to flap in the future based on the history of their behavior:
 - Flap = Remove route
 - Suppress = Do not use a route after it reappears

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0—3-10

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 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.

If a route flaps once or twice, it is typically not considered a flap from an administrative perspective. If the flapping happens more often, however, there is probably something 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. It does not, therefore, propagate it to other BGP neighbors or use it for data forwarding.

Route Dampening Operation

- Each time an EBGp route flaps, it gets 1000 penalty points (IGBP routes are not dampened).
- The penalty placed on a route decays according to the exponential decay algorithm.
- When the penalty exceeds the **suppress limit**, the route is dampened (no longer used or propagated to other neighbors).
- A dampened route is propagated when the penalty drops below the **reuse limit**.

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-311

A BGP router with route dampening enabled keeps track of 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 by using a decaying algorithm. If a route flaps several times, it 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 reduces penalty points that are applied to a flapping route. 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.

Route Dampening Operation (Cont.)

- The flap history is forgotten when the penalty drops below half of the **reuse limit**.
- A route is never dampened for more time than the **maximum suppress limit**.
- An unreachable route with a flap history is put in the **history state**; it stays in the BGP table but only to maintain the flap history.
- A penalty is applied on the individual path in the BGP table, not on the IP prefix.

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0—3-12

A router stops tracking penalty points when they are lower than 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 a BGP neighbor has withdrawn can still be seen in the BGP table and is marked with an “h,” which means history state.

A penalty is always applied to a path and not a prefix. If a path is flapping, it does not mean that the destination is flapping.

Configuring BGP Route Dampening

```
Router(config-router)#  
RP/0/RP0/CPU0:CRS(config-bgp-af)#  
bgp dampening [half-life reuse suppress max-suppress-time]  
[route-map map-name]
```

- Configures BGP route dampening.
- BGP dampening parameters:
 - **Half-life**: Decay time in which the penalty is halved
 - **Suppress**: Value when the route starts dampening
 - **Reuse**: Value when the dampened route is reused
 - **Max-suppress-time**: Maximum time to suppress the route
 - **Route map**: Name of route map controlling dampening

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0–313

To enable route dampening, use the **bgp dampening** command. Optionally, you can change the default settings of the route dampening parameters.

Route flap dampening requires 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).
- **Max-suppress-time**: No route can be suppressed longer than the max-suppress-time minutes (default is 1 hour; maximum is 255 minutes).

You can specify the four route flap dampening parameters directly with the **bgp dampening** command. Alternatively, you can create a route map that specifies different dampening parameters for different sets of routes, and then you can apply the route map with the **bgp dampening route-map** command.

Configuring BGP Route Dampening (Cont.)

Most Internet service providers use default values:

- A flapping route is dampened after three successive flaps.
- A route stays suppressed for approximately 30 minutes.
- **Net result:** The route is lost for 30 minutes if a BGP session with a neighbor is cleared three times in succession.
- Default dampening parameter values are:
 - **Half-life:** 15 minutes
 - **Suppress:** 2000
 - **Reuse:** 750
 - **Max-suppress-time:** 60 minutes (4x half-life)
 - **Per-flap penalty:** 1000 (nonconfigurable)

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0–3-14

This sample calculation shows how long a route that flaps three times is suppressed with the default values of the Cisco IOS route dampening parameters. Each time that a route flaps, it accumulates another 1000 points. After the third flap, the route has almost 3000 points. Remember that the penalty is gradually decreased by using a decaying algorithm, causing a reduction in the number of points that the route accumulates. It takes 15 minutes for the penalty to drop below 1500 (provided there are no further flaps) and another 15 minutes to drop below the reuse limit of 750.

Many service providers change the default parameters to allow a maximum suppress time of several hours.

Note	Using the clear ip bgp * command is regarded as a flap to neighboring autonomous systems. Using this command several times may cause neighboring autonomous systems to suppress prefixes for some time, even if there is nothing wrong with the route.
-------------	---

Note	Using the clear ip bgp * [soft] [in out] command is not regarded as a flap. This command should be used instead of clear ip bgp * for clearing the neighbor relationships.
-------------	--

You can change all default values if you specify them in the optional parameters of the **bgp dampening** command. The per-flap penalty is the only value that is not configurable.

Configuring BGP Route Dampening (Cont.)

```
Router(config-route-map)#  
RP/0/RP0/CPU0:CRS(config-rpl-if)#
```

```
set dampening half-life reuse suppress max-suppress-time
```

- This command sets the BGP dampening parameters for individual routes matched by a route map or policy map entry.
- Apply this route map to the **bgp dampening** command instead of specifying individual parameters.
- Applications:
 - Dampens routes toward root DNS servers (or other servers) less aggressively
 - Dampens smaller prefixes more aggressively
 - Selective dampening based on BGP neighbor and route map match criteria

© 2010 Cisco Systems, Inc. All rights reserved.MSRP v1.0-315

Many service providers prefer to implement selective dampening. Larger prefixes are usually less likely to flap and should not be penalized as aggressively as the smaller prefixes that populate most of the BGP table.

You can use a route map in combination with a prefix list to match on prefix length and to set different route dampening parameters for larger prefixes than for smaller ones. A practical service provider policy is to use a route map to exclude root Domain Name System (DNS) servers from dampening altogether.

You can then attach the route map to the BGP route dampening process with the **bgp dampening route-map** command.

Releasing Dampened Routes

Router#

```
clear ip bgp ip-address flap-statistics [{regex regexp} |  
{filter-list list-name} | {ip-address network-mask}]
```

RP/0/RP0/CPU0:CRS#

```
clear bgp ipv4 unicast ip-address flap-statistics [{regex  
regexp} | {filter-list list-name} | {ip-address network-  
mask}]
```

- Clears the flap statistics, but does not release dampened routes

Router#

```
clear ip bgp dampening [ip-address network-mask]
```

RP/0/RP0/CPU0:CRS#

```
clear bgp ipv4 unicast dampening [ip-address network-mask]
```

- Releases all the dampened routes or just the specified network
- Flap statistics or dampened routes also cleared when the BGP session with the neighbor is lost

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-3-16

Two timers are calculated for every route when it flaps.

- **Time to forget:** The time that it takes before all flap history is deleted. Using the **clear ip bgp flap-statistics** Cisco IOS command or **clear bgp ipv4 unicast flap-statistics** Cisco IOS XR command deletes all penalty information, but it does not release the dampened paths.
- **Time to reuse:** The time that it takes before a route can be considered for best-path processing. Using the **clear ip bgp dampening** Cisco IOS command and **clear bgp ipv4 unicast dampening** Cisco IOS XR command resets this timer for specified networks or for all networks so that they are no longer suppressed. The flap statistics, however, are still kept. The next flap will cause the previously dampened paths to be suppressed again.

Monitoring Route Dampening

Router#

```
show ip bgp dampened-paths
```

RP/0/RP0/CPU0:CRS#

```
show bgp ipv4 unicast dampened-paths
```

- Displays the dampened routes

Router#

```
show ip bgp flap-statistics [{regex regex} | {filter-list access-list} | {ip-address mask [longer-prefix]}]
```

RP/0/RP0/CPU0:CRS#

```
show bgp ipv4 unicast flap-statistics [{regex regex} | {filter-list access-list} | {ip-address mask [longer-prefix]}]
```

- Displays flap statistics for all routes with dampening history
- Can match routes against regular expressions, AS path access lists, a specific route, or more specific routes

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-317

The penalty that is placed on a network is decayed until the reuse limit is reached, when the route is once again advertised. Every time that a route flap occurs, the penalty is recalculated. In the router, the penalty is encoded as the time that it takes for the penalty to decay below the reuse limit. At half of the reuse limit, the dampening information for the route to the network is removed.

Use the **show ip bgp dampened-paths** Cisco IOS command and **show bgp ipv4 unicast dampened-paths** Cisco IOS XR command to list all routes that are currently suppressed because of dampening.

Use the **show ip bgp flap-statistics** Cisco IOS command and **show bgp ipv4 unicast flap-statistics** Cisco IOS XR command to list all routes that have a penalty that is still higher than the time-to-forget limit. You can also use this command in combination with regular expressions and filter lists.

Monitoring Route Dampening (Cont.)

Router#

```
debug ip bgp dampening
```

RP/0/RP0/CPU0:CRS#

```
debug bgp dampening ipv4 unicast
```

- Displays the BGP dampening events

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0—3-18

Use the **debug ip bgp dampening** Cisco IOS command and the **debug bgp dampening ipv4 unicast** Cisco IOS XR command to display BGP dampening events as they occur in real time.

BGP Scan Time

```
Router(config-router) #  
RP/0/RP0/CPU0:CRS(config-bgp) #  
bgp scan-time scanner-interval
```

- This command changes the default value of BGP scanner process runs (default = 60 seconds).
- The BGP scanner walks the BGP table and confirms the reachability of next hops.
- The BGP scanner process is also responsible for advanced features such as conditional advertisement check and performing route dampening.

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-319

Network administrators use the **bgp scan-time** command to configure the time interval for repetitions of the BGP scanner process.

The BGP scanner process walks (scans) the BGP table and confirms the reachability of next hops. A change of this status triggers a new BGP route selection for the network. The router propagates changes to established BGP neighbors. Increasing the BGP scanner process frequency will make the router find a changed status more quickly, but it will also consume more CPU resources.

The BGP scanner process is also responsible for some advanced BGP features. It checks the conditional advertisement to determine whether BGP should advertise conditional prefixes or perform route dampening.

Monitoring BGP Scan Time

```
Router#show ip bgp summary
BGP router identifier 172.16.1.1, local AS number 100
BGP table version is 199, main routing table version 199
37 network entries using 2850 bytes of memory
59 path entries using 5713 bytes of memory
18 BGP path attribute entries using 936 bytes of memory
2 multipath network entries and 4 multipath paths
10 BGP AS-PATH entries using 240 bytes of memory
7 BGP community entries using 168 bytes of memory
0 BGP route-map cache entries using 0 bytes of memory
0 BGP filter-list cache entries using 0 bytes of memory
36 received paths for inbound soft reconfiguration
BGP using 34249 total bytes of memory
Dampening enabled. 4 history paths, 0 dampened paths
BGP activity 37/2849 prefixes, 60/1 paths, scan interval 15 secs
Neighbor      V      AS MsgRcvd MsgSent  TblVer  InQ  OutQ  Up/Down State/PfxRcd
10.100.1.1     4      200      26      22      199   0     0 00:14:23 23
10.200.1.1     4      300      21      51      199   0     0 00:13:40 0
```

- Scan interval is defined per BGP router process and address family
- Checked with **show ip bgp summary** Cisco IOS command or **show bgp ipv4 unicast summary** Cisco IOS XR command

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0—320

You can check the configured BGP scan interval using the **show ip bgp summary** Cisco IOS command and the **show bgp ipv4 unicast summary** Cisco IOS XR command.

The configured BGP scan interval will apply to the entire BGP routing protocol process.

BGP Advertisement Interval

```
router(config-router)#
```

```
neighbor {ip-address | peer-group-name} advertisement-  
interval seconds
```

```
RP/0/RP0/CPU0:CRS(config-bgp-nbr)#
```

```
advertisement-interval seconds
```

- This command changes the default time interval when sending BGP routing updates for a specific neighbor:
 - If lowered, can improve convergence
 - If the value is set too low, can consume considerable resources in a jittery network
- Default values:
 - 30 seconds for EBGp neighbors
 - 5 seconds for IBGP neighbors

© 2010 Cisco Systems, Inc. All rights reserved.MSRP v1.0-321

With the help of the **neighbor advertisement-interval** Cisco IOS command and the **advertisement-interval** Cisco IOS XR command, network administrators can modify the default advertisement interval for a specific BGP peer. BGP advertisements are rate-limited by the router that is using the advertisement interval timer (defined as the parameter `MinRouteAdvertisementInterval` in RFC 1771). When a BGP-speaking router sends a route update to a neighbor for a specific destination, it is not allowed to send another update to the neighbor about the same destination until a period equal to the advertisement interval has elapsed. In that way, the advertisement interval timer acts as a form of rate limiting on a per-destination basis, even though the value of the advertisement interval is configured for each neighbor.

It is important to note that while the router is waiting for the advertisement interval timer to expire, the router can still receive and process route updates from BGP neighbors. Using the **neighbor advertisement-interval** Cisco IOS command and **advertisement-interval** Cisco IOS XR command does not rate-limit BGP route selection (inbound updates and subsequent processing). However, these commands do rate-limit outgoing route advertisements.

When faster propagation of successive BGP updates (which are batched and rate-limited) is required, the network administrator can lower the default value of the advertisement interval, thus improving convergence.

In routers that process large BGP tables and less stable networks, lowering the advertisement interval could potentially lead to consuming large portions of router resources.

Monitoring the BGP Advertisement Interval

```
Router#show ip bgp neighbors 10.108.50.2
BGP neighbor is 10.108.50.2, remote AS 1, internal link
  BGP version 4, remote router ID 192.168.252.252
  BGP state = Established, up for 00:24:25
  < text omitted >
  Default minimum time between advertisement runs is 5 seconds
  < text omitted >
```

- Defined per BGP neighbor and address family
- Manually configured minimum value stated under address family output of **show ip bgp neighbors** Cisco IOS command or **show bgp ipv4 unicast neighbors** Cisco IOS XR command

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0—322

You can examine the currently configured BGP advertisement interval with the **show ip bgp neighbors** Cisco IOS command and the **show bgp ipv4 unicast neighbors** Cisco IOS XR command. The advertisement interval is defined for a specific neighbor in a specific BGP address family. Actual values of the advertisement interval are therefore stated under the specific address-family portion of the neighbor output. The default timer for the Internal Border Gateway Protocol (IBGP) is 30 seconds.

Monitoring BGP Memory Utilization

This topic describes the methods for monitoring BGP memory utilization.

Monitoring Memory Utilization

Monitoring methods:

- SNMP:
 - Traps by router
 - Polling by monitoring server
- CLI

```
Router#show processes memory
Total: 5611448, Used: 2307548, Free: 3303900
  PID  TTY  Allocated    Freed   Holding  Getbufs  Retbufs  Process
    0   0    199592     1236   1907220         0         0  *Init*
    0   0         400     76928        400         0         0  *Sched*
    0   0   5431176   3340052   140760    349780         0  *Dead*
    1   0        256        256       1724         0         0  Load Meter
    2   0        264         0       5032         0         0  Exec
    3   0         0         0       2724         0         0  Check heaps
    4   0    97932         0       2852    32760         0  Pool Manager
    5   0        256        256       2724         0         0  Timers
    6   0         92         0       2816         0         0  CXBus hot stall
    7   0         0         0       2724         0         0  IPC Zone Manager
```

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-324

To monitor memory utilization, use the following methods:

- SNMP
- CLI

Monitoring Memory Utilization (Cont.)

SNMP Notification Support:

- CISCO-ENHANCED-MEMPOOL-MIB: Describes SNMP objects that enable users to remotely monitor the memory pool statistics of all physical entities.

Enabling memory pool—SNMP notification support example:

```
Router(config)#  
snmp-server enable traps memory bufferpeak  
snmp-server host myhost.cisco.com informs version 3 public memory
```

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0—325

The CISCO-ENHANCED-MEMPOOL-MIB module describes SNMP objects that enable users to remotely monitor the memory pool statistics of all physical entities, such as line cards and route processors, in a managed device. This monitoring is particularly useful for high-end routers that may have many line cards. Beginning in Cisco IOS Release 12.3(4)T, the MIB has been enhanced to provide buffer pool and buffer cache statistics.

In addition to the statistics provided by the MIB, SNMP notifications (traps or informs) can be configured to be sent when the maximum number of memory buffers changes (when a new buffer peak is reached).

In the example, the **snmp-server enable traps memory bufferpeak** command enables only buffer peak notifications (traps or informs). The **snmp-server host myhost.cisco.com informs version 3 public memory** command enables buffer peak notifications to be sent to the specified host.

Proposed Changes to Optimize Memory Utilization

Propose changes to mitigate memory issues:

- Filter prefixes longer than /24
- Limit the number of routes received from a neighbor
- Prevent customers from sending fragmented prefixes

Major Change

Major Change

Major Change

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-326

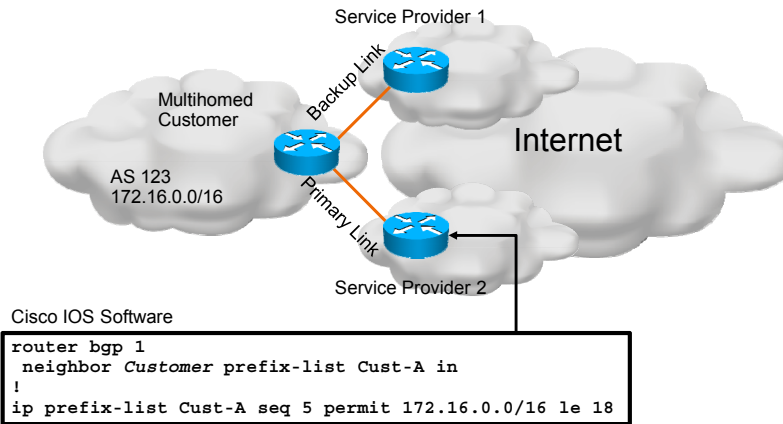
There are changes that can mitigate memory utilization.

- Filter prefixes longer than /24. Service providers employ this common filtering technique to limit the size of the Internet routing table.
- Limiting the number of routes that are received from a neighbor.
- Prevent customers from sending fragmented prefixes. Customers running BGP should only announce their whole prefix and should not send multiple smaller prefixes out of their contiguous address range. Optionally, you can let them send a few subnets to implement load balancing.

Any filtering or limiting of routing updates can cause partial or complete failure for one or multiple customers. Carefully study and design the optimization before deploying it.

Filtering Customer Prefix

Requirement: The customer will announce prefixes only from assigned address space (172.16.0.0/16), with subnet masks no longer than /18.



In this example, a multihomed customer has a provider-independent Class B network (RFC 1918 addressing is used in the example: 172.16.0.0/16). The customer may subnet this address space but may not announce subnets smaller than a subnet mask of 255.255.192.0. Larger subnets are accepted. If the customer has subnetted the network into smaller subnets, it must summarize the routing information about those subnets into at least /18 prefixes before announcing them to the service provider.

The primary ISP implements a prefix list named Cust-A to filter the incoming information from the multihomed customer. The prefix list permits all routes that have 172.16 in the first 16 b and have a subnet mask of 18 b or less that are received from the customer. Any other routes from the customer are denied and silently ignored.

Limiting the Number of Routes Received from a Neighbor

Definition of problem:

- All other filtering mechanisms specify only what you are willing to accept, but not how much.
- A misconfigured BGP neighbor can send a huge number of prefixes that can exhaust the memory of a router or overload the CPU (several Internet-wide incidents have already occurred).
- A tool is needed to establish a hard limit on the number of prefixes received from a neighbor.

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-328

Incoming filters and route maps indicate the BGP attribute values that a route should have to be accepted. Route filters 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 filter is one that matches the contents of the AS path attribute.

An ISP with a multihomed customer may use filters to ensure that the routes that are received from the customer originate within the AS of the customer. Using an AS path access list is one method of achieving this goal.

An improperly configured filter in a customer router may accidentally cause the customer to receive many Internet routes. Even worse, a faulty configuration may cause the customer to advertise prefixes as though the routes originated inside the customer AS. This situation would result in a BGP table in the ISP 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 ISP network would select those routes as the best (based on the AS path length). Much of the provider traffic that is intended for the Internet would be directed to the customer network.

An AS path filter in the ISP router would not prevent this accident. The routes that the customer sends have the anticipated AS path value. A prefix list that distinctly identifies and permits each of the network numbers that the customer may advertise would have prevented the accident. But such a prefix list is difficult to maintain.

Configuring the BGP Maximum-Prefix Function

Router(config-router)#

```
neighbor ip-address maximum-prefix maximum [threshold]
[warning-only] [restart restart-interval]
```

RP/0/RP0/CPU0:CRS(config-bgp-nbr-af)#

```
maximum-prefix maximum-prefix maximum [threshold]
[warning-only]
```

- This command **controls** how many prefixes can be received from a neighbor.
- The optional **threshold** parameter specifies the percentage when a warning message is logged (default is 75%).
- The optional **warning-only** keyword specifies the action on exceeding the maximum number (default is to drop the neighbor relationship).
- The optional **restart** keyword instructs the router to try to re-establish the session after the specified interval in minutes.

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0—329

A scalable solution for limiting the number of routes (prefixes) that are received from a BGP neighbor is to use a tool that limits the number of routes that are received from a specific neighbor.

To control how many prefixes that a BGP router can receive from a neighbor, use the **neighbor maximum-prefix** Cisco IOS configuration command and the **maximum-prefix** Cisco IOS XR configuration command.

Monitoring the BGP Maximum-Prefix Function

Router#

```
show ip bgp neighbors [address]
```

RP/0/RP0/CPU0:CRS#

```
show ip bgp neighbors [address]
```

- For neighbors with the maximum-prefix function configured, displays the maximum number of prefixes and the warning threshold.
- For neighbors exceeding the maximum number of prefixes, displays the reason that the BGP session is idle.

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0—330

Network administrators use the **show ip bgp neighbors** Cisco IOS command and the **show bgp ipv4 unicast neighbors** Cisco IOS XR command to monitor the status of BGP neighbors. Among other things, the command displays information about how many prefixes a BGP router has received from a neighbor and if any limits have been configured.

Monitoring BGP Link Utilization

Describe the methods for monitoring link utilization that is caused by BGP routing policy.

Monitoring Link Utilization

- BGP routing policies are responsible for route selection for external networks.
- Load on multiple upstream links can be disproportionately distributed.
- Analyze traffic load on edge links:
 - Inbound traffic is controlled by outbound BGP policies
 - Outgoing traffic is controlled by inbound BGP policies
- Propose changes to routing policies to distribute the load more evenly.

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-332

BGP routing policies are responsible for selecting routes for external networks. The load on multiple upstream links can be disproportionately distributed. You can analyze the traffic load on edge links.

- Inbound traffic is controlled by outbound BGP policies.
- Outgoing traffic is controlled by inbound BGP policies.

Propose changes to routing policies to distribute the load more evenly.

Proposed Changes for Link Utilization

- Link utilization on network edges can be optimized using routing policies in BGP:
 - Customers can implement their own policies (AS path prepending).
 - Service providers can provide services based on communities to help customers implement their policies.
 - Service providers can use local preference, AS path prepending, and upstream communities to implement edge routing policies.
- Propose routing policy changes to optimize link utilization:
 - Local preference Major Change
 - AS path prepending Major Change
 - Communities Major Change

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0–333

Proposed changes for link utilization are BGP policy modifications using the following:

- Local preference
- AS path prepending
- Communities

Monitoring Link Utilization

Monitoring methods:

- Monitoring system using SNMP
- CLI

```
RP/0/0/CPU0:PE1# show interfaces GigabitEthernet 0/4/0/0
GigabitEthernet0/4/0/0 is up, line protocol is up
  Interface state transitions: 2
  Hardware is GigabitEthernet/IEEE 802.3 interface(s), address is
0017.e0ae.6f90
  Internet address is Unknown
  MTU 1514 bytes, BW 1000000 Kbit
    reliability 255/255, txload 75/255, rxload 44/255
  Encapsulation ARPA,
  Full-duplex, 1000Mb/s, SXFD, link type is force-up
  output flow control is on, input flow control is off
  loopback not set,
  ARP type ARPA, ARP timeout 04:00:00
  Last clearing of "show interface" counters never
  30 second input rate 174348000 bits/sec, 148219 packets/sec
  30 second output rate 294322000 bits/sec, 158909 packets/sec
<rest omitted>
```

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-334

To monitor link utilization, use the following methods:

- SNMP
- CLI

BGP Local Preference to Control Outbound Traffic

- You can use local preference to ensure AS-wide route selection policy.
- Any BGP router can set local preference when processing incoming route updates, when doing redistribution, or when sending outgoing route updates.
- Local preference is used to select routes with equal weight.
- Local preference is stripped in outgoing EBGp updates, except in EBGp updates with confederation peers.

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-335

Local preference is similar to weight; because it is an attribute, you can set it once and then view it on neighboring routers without having to reset it. This attribute has a default value of 100, which the router will apply to locally originated routes and updates that come in from external neighbors. Updates that come from internal neighbors already have the local preference attribute.

Local preference is the second-strongest criterion in the route selection process. If there are two or more paths available for the same network, a router will first compare weight. If the weights are equal for all paths, the router will then compare the local preference attributes. The path with the highest local preference value will be preferred.

The local preference attribute is automatically stripped out of outgoing updates to External Border Gateway Protocol (EBGP) sessions. This practice means that you can use this attribute only within a single AS to influence the route selection process.

BGP Local Preference to Control Outbound Traffic (Cont.)

- Local preference is the second strongest BGP route selection parameter.
- Remember the BGP route selection rules:
 - Highest weight preferred (local to router)
 - Highest local preference preferred (global within AS)
 - Other BGP route selection rules
- To ensure consistent AS-wide route selection:
 - Do not change local preference within the AS.
 - Do not use BGP weights.

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0–336

Local preference is the second-strongest BGP route selection parameter. Remember the route selection rules:

1. Prefer the highest weight (local to router)
2. Prefer the highest local preference (global within AS)
3. Process all remaining BGP route selection rules

Because network administrators can use both weight and local preference to manipulate the route selection process, they must decide which one to use. If local preference is used, the weight should be the same for all paths.

Network administrators can use weight on an individual router to override local preference settings that are used in the rest of the AS.

In most cases, changing the default local preference on updates coming from external neighbors is enough. Network administrators should avoid changing the local preference attribute on internal sessions to prevent unnecessary complexity and unpredictable behavior.

Configuring Default Local Preference

```
Router(config-router) #  
RP/0/RP0/CPU0:CRS(config-bgp) #  
bgp default local-preference preference
```

- This command changes the default local preference value.
- The specified value is applied to all routes that do not have local preference set (EBGP routes).
- The default value of this parameter is 100, allowing you to specify more desirable or less desirable routers.

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-337

You can use the **bgp default local-preference** Cisco IOS and the Cisco IOS XR command in BGP configuration mode to change the default value of local preference. The new default value applies only to locally originated routes and those routes that are received from external neighbors.

Setting a value lower than the default of 100 will result in the router preferring internal paths to external paths (normally a router would prefer external routes).

Setting a value higher than 100 will result in external paths being preferred to all internal paths (also those paths with a shorter AS path).

Monitoring Local Preference

- Nondefault local preference is displayed in the **show ip bgp** and **show bgp ipv4 unicast** printout.
- Local preference is displayed in the **show ip bgp prefix** and **show bgp ipv4 unicast prefix** printout.
- Local preference is displayed in BGP update debugging only for inbound updates.

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0—338

Although local preference is not a mandatory attribute, it is applied to every route. When you are using the **show ip bgp** Cisco IOS command and the **show bgp ipv4 unicast** Cisco IOS XR command, a locally applied default value is not shown. All other values are displayed. Use the **show ip bgp prefix** Cisco IOS command and the **show bgp ipv4 unicast prefix** Cisco IOS XR command to also display the locally applied value.

The output that is displayed from **show** and **debug** commands will vary depending on the Cisco IOS Release.

AS Path Prepending

- The manual manipulation of AS path length is called **AS-path prepending**.
- The AS path should be extended with multiple copies of the **AS number of the sender**.
- AS-path prepending is used to:
 - Ensure proper return path selection
 - Distribute the return traffic load for multihomed customers

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0–339

You can manipulate AS paths by prepending AS numbers to existing AS paths. Normally, you perform AS path prepending on outgoing EBGp updates over the undesired return path. Because the AS paths sent out over the undesired link become longer than the AS paths sent out over the preferred path, the undesired link is now less likely to be used as the return path.

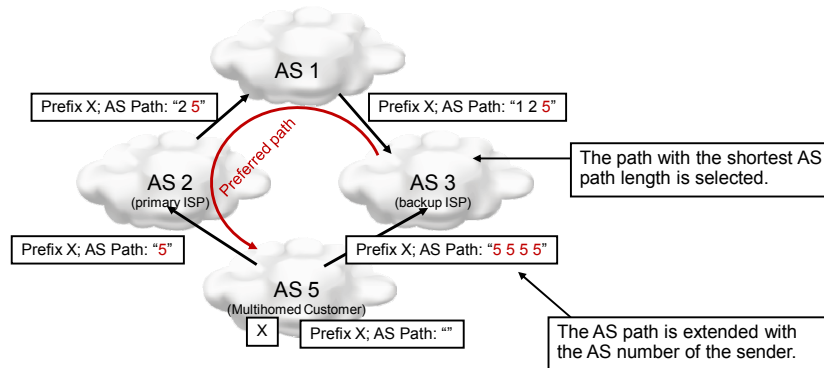
The length of the AS path is extended because additional copies of the AS number of the sender are prepended to (added to the beginning of) the AS path attribute. To avoid clashes with BGP loop prevention mechanisms, no other AS number, except that of the sending AS, should be prepended to the AS path attribute.

If another AS number is prepended in the AS path, the routers in the AS that has been prepended will reject the update because of BGP loop prevention mechanisms.

You can configure prepending on a router for all routing updates that you send to a neighbor or only on a subset of them.

AS Path Prepending (Cont.)

- Result: The return traffic flows over the desired return path



© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0—3-40

As long as the high-speed link between AS 2 and AS 5 is available, all traffic should flow toward AS 5 using the high-speed link.

To accomplish this goal, configure the router in AS 5 that sends EBGp updates to AS 3 by prepending the AS path with three copies of the AS number 5. AS 3 receives two alternative routes to reach network X. The first alternative route is the update that it received directly from AS 5 (that has a manipulated AS path with a length of four). The second alternative is the update that it received via AS 1. (That route was not manually manipulated and therefore contains an AS path length of three).

When AS 3 starts the route selection process to determine which route to use to reach network X, it checks the AS path length after the weight and local preference parameters. In this case, neither weight nor local preference has been configured, so the length of the AS path will be the deciding factor in the route selection process. Consequently, AS 3 prefers the shortest AS path and thus forwards packets toward network X via AS 1. The desired administrative policy has been met, and AS 5 will receive incoming traffic over the high-speed link.

If the forwarding path from AS 3 via AS 1 to AS 5 and network X is later broken, the BGP update to reach network X is revoked. Should such a network failure occur, AS 3 will have only one remaining path to reach network X. The route selection process now has only one choice, the route directly to AS 5 over the low-speed WAN link. The low-speed link will therefore serve as backup to the high-speed WAN link.

AS Path Prepending Design Considerations

- There is no exact mechanism to calculate the required prepended AS path length.
- If a **primary and backup** scenario is desired, use one of the listed options:
 - Use AS path prepending on the backup peer.
 - Use BGP communities to signal the backup peer to lower the local preference.
 - Use BGP communities to signal the backup peer to do prepending.
 - Use conditional advertising (longer convergence).
- If **load balancing** is desired:
 - Start with a short prepended AS path, monitor link use, and extend the prepended path length as needed.
 - Continuously monitor the link use and change the prepended AS path length if required.
 - Use conditional prepending if multiple address ranges are being advertised.

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0–341

How many copies of the sender AS number should you prepend to the AS path? The answer depends on the goals of the administrative policy. In general, it is difficult to determine the exact number of required AS numbers to prepend. The sending AS does not know what alternative paths are available to other autonomous systems.

There are two typical cases in which you can use AS path prepending for return path selection.

- **Establishing a primary and a backup link:** As an announced backup (prepended) route propagates through the Internet, all the routers along the way that receive the route need to store it together with its AS path attribute. If this information is long, it will consume extra memory in these routers. However, because routers forward only routes that are selected as best, an AS that receives multiple alternatives to a destination will select the route with the shortest AS path and forward only that route.

When both the primary and the secondary links are up, the neighboring AS will receive two routes to the same destination that differ only in the AS path length. The route with the shorter AS path will then be advertised through the Internet.

When the primary link fails, the route with the longer AS path is the only remaining route. As a result, the primary route is withdrawn, and the prepended route is advertised through the Internet. In this case, extra memory will be consumed in each Internet router because of the storage of the prepended (longer) AS path.

The longer the AS path that is announced to the EBGp neighbor on the other side of the backup link, the less likely it is that incoming traffic will be received from that neighbor. The network administrator can make a clever guess about how many copies of the AS number to prepend. After the prepending is implemented, the network administrator must examine the result. If the expected result is not achieved, the configuration can be changed and a few more copies of the AS number can be prepended.

After AS path prepending has generated the desired results, the network administrator may take the precaution of prepending a few more copies of the AS number to the AS path. This action protects the customer from packets being routed over the backup link at a possible later stage when the topology between remote autonomous systems has unexpectedly changed, yielding a longer AS path to reach the primary link.

- **Distributing the load of return traffic:** In a multihomed scenario, there is no way to exactly predetermine the volume of traffic that will be received over a particular link. The traffic load on different links will change depending on where the senders are located (which autonomous systems they belong to). The network topology and the way that different remote autonomous systems are interconnected may also change with time, changing the load distribution. Only constant monitoring and fine tuning will ensure that the desired results are achieved.

In a first attempt at load distribution, the network administrator can configure a router that is connected to an overused link to prepend only a few extra copies of the local AS number. After the network has been given time to converge, the network administrator must check the change in load distribution. Load monitoring must be done for a period long enough to be statistically significant (several days or more). If enough volume of traffic has not moved from the overused link to the underused link, the administrator must prepend more copies of the local AS number. The process of resending local routes and monitoring the results starts all over again.

AS Path Filtering Concerns

AS Path Prepending

- The incoming AS path filters of the service provider need to be modified to support AS path prepending.
- To support AS path prepending, service providers should implement regular expression variables to create a uniform AS path filter for all customers:
 - Example: `^([0-9]+)(\1)*$`

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0–342

Because the AS of the service provider will receive customer routes with prepended AS paths that have a length greater than one AS number, the provider must modify its incoming filters.

The service provider needs to create a new inbound regular expression filter, using regular expression variables and parentheses for recall.

What is needed is a filter that will allow any AS path containing one or multiple copies of the same AS number. The following is an example of such a filter:

■ `^([0-9]+)(\1)*$`

This filter matches any AS path beginning with any AS number and continues with no or multiple repetitions of that same AS number. (The variable “\1” repeats the value in the brackets.) The regular expression would therefore match AS paths 99 99 99, 2 2 2, or 100, but it would not match AS path 100 99.

Monitoring Communities

- Communities are displayed in the **show ip bgp prefix** and **show bgp ipv4 unicast prefix** printout.
- Communities are not displayed in debugging outputs.
- Routes in the BGP table that are tagged with a set of communities or routes matching a community-list can be displayed.

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0—343

Because a community is an attribute that can appear more than once in a single update, the **show ip bgp** command does not show it. You can view communities only if you use the **show ip bgp prefix** Cisco IOS command and the **show bgp ipv4 unicast prefix** Cisco IOS XR command.

Monitoring Communities (Cont.)

- Communities are displayed only in the **show ip bgp prefix** printout.

```
Router#show ip bgp 65.16.0.0
BGP routing table entry for 65.16.0.0/16, version 2
Paths: (1 available, best #1, table Default-IP-Routing-Table)
  Advertised to non peer-group peers:
    192.168.20.22
  99
    192.168.21.99 (metric 20) from 172.16.0.4 (172.16.0.4)
      Origin IGP, metric 0, localpref 100, valid, internal, best
      Community: 1:22 1:53 1:120 20:35
```

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0—344

This example shows the output of the **show ip bgp prefix** command. The update contains four community attributes (1:22, 1:53, 1:120, and 20:35).

Monitoring Communities (Cont.)

Router#

```
show ip bgp community
```

- Displays all routes in a BGP table that have at least one community attached

Router#

```
show ip bgp community as:nn [as:nn ...]
```

RP/0/RP0/CPU0:CRS#

```
show bgp ipv4 unicast community as:nn [as:nn ...]
```

- Displays all routes in a BGP table that have all the specified communities attached

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-345

To monitor communities, use the following Cisco IOS and Cisco IOS XR commands:

- The **show ip bgp community** Cisco IOS command
- The **show ip bgp community as:nn [as:nn ...]** Cisco IOS command
- The **show bgp ipv4 unicast community as:nn [as:nn ...]** Cisco IOS XR command

Monitoring Communities (Cont.)

Router#

```
show ip bgp community as:nn [as:nn ...] exact
```

RP/0/RP0/CPU0:CRS#

```
show bgp ipv4 unicast community as:nn [as:nn ...] exact-match
```

- Displays all routes in a BGP table that have exactly the specified communities attached.

Router#

```
show ip bgp community-list clist
```

- Displays all routes in a BGP table that match community-list "clist."

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-346

To monitor communities, use the following Cisco IOS and Cisco IOS XR commands:

- The **show ip bgp community** *as:nn [as:nn ...] exact* Cisco IOS command
- The **show bgp ipv4 unicast community** *as:nn [as:nn ...] exact-match* Cisco IOS XR command
- The **show ip bgp community-list** *clist* Cisco IOS command

Auditing BGP Configuration

This topic describes the methods for auditing BGP configuration.

Auditing BGP Configuration

- **Security:**
 - Authentication
 - BGP TTL security
- **Availability:**
 - Cisco Nonstop Forwarding (Cisco NSF)
 - Cisco Nonstop Routing (Cisco NSR)

© 2010 Cisco Systems, Inc. All rights reserved. MSRP v1.0—348

BGP performance can be impaired due to:

- Security breaches
- Network failures

When auditing BGP configuration, you should consider deploying the following methods:

- Message Digest 5 (MD5) authentication of BGP sessions
- Time-to-Live (TTL) security to prevent attacks from nonadjacent attackers
- Using stateful switchover (SSO) with Cisco Nonstop Forwarding (NSF) between adjacent routers that support Cisco NSF
- Using SSO with Cisco Nonstop Routing (Cisco NSR) with routers that do not support Cisco NSF

BGP Security

- BGP is used on network edges.
- MD5 authentication and TTL security should be used to prevent attacks against BGP.
- Both routers must be configured with the same password (MD5 shared secret):
 - Coordinate deployment or change of password with the customer and peer.
 - Changing BGP passwords requires sessions to be restarted.
 - Enabling TTL security requires sessions to be restarted.

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-349

Authentication between BGP neighbors can be negotiated between BGP-speaking routers using optional parameters in the Open message. If you are using Message Digest 5 (MD5) authentication, every TCP segment on the BGP session will be transmitted to the configured neighbor along with a checksum. The checksum is calculated together with a secret known by the two routers using the MD5 algorithm. The common secret is never transmitted on the network. If the receiver, which is using the same common secret, calculates the same checksum from the TCP segment, then the receiver can be almost certain that the information is transmitted from the correct source, and the information has not been altered.

Authentication of BGP sessions is a vital tool to avoid denial of service (DoS) attacks.

Additionally, BGP sessions can use TTL security when BGP peers send packets using TTL 255. The receiver only accepts packets when the TTL is not below the configured value. This setup prevents attacks from several hops away.

The BGP TTL-based security check is designed to protect BGP processes from CPU utilization-based attacks. The BGP protocol defines two types of sessions: "internal" BGP sessions that are established between peers within the same AS, and external BGP sessions that are established between peers in two different autonomous systems. Of particular interest are the external BGP sessions, which are the type of BGP sessions that would be established between an enterprise and its upstream service provider.

By default and per RFC 3682, when the external BGP is configured, the IP header TTL for all neighbor session packets is set to "1." This configuration was originally assumed to be useful because it prevented the establishment of an external BGP session beyond a single hop. However, the configuration does not take into account an attacker who could be located up to 255 hops away and could have the ability to send spoofed packets to BGP-speaking routers. For example, an attacker could send large amounts of TCP synchronization (SYN) packets at the BGP peer to overwhelm the BGP process. The BGP TTL security check leverages ISP external BGP peering sessions between routers that are next to each other (either between directly connected interfaces or possibly between loopbacks). Because TTL spoofing is considered nearly impossible, a mechanism that is based on an expected TTL value was developed to provide a simple yet robust defense from infrastructure attacks that are based on forged BGP packets.

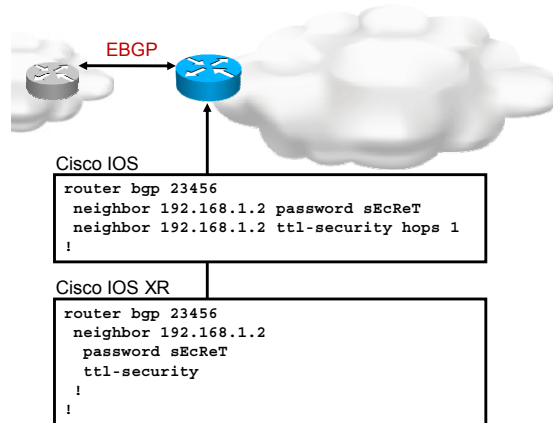
Note: This concept was originally defined in the document, *The BGP TTL Security Hack (BTSH)*. The concept was later extended beyond BGP in the document, *The Generalized TTL Security Mechanism (GTSM)* (RFC 3682).

When the BGP TTL security check is enabled, the initial value for external BGP is set to "255" (instead of "1"), and a minimum TTL value is enforced on all BGP packets that are associated with that external BGP session. Because the IP header TTL value is decremented by each hop (router interface) along its path to its final destination, using BTSH restricts possible attack origins to the directly connected routers.

Currently, Cisco IOS XR Software does not support GTSM for external BGP multi-hop scenarios; only directly connected or loopback-based external BGP peering sessions can use the GTSM feature. This feature is checked in hardware during packet ingress in line cards.

Implementing BGP Security

- Enables MD5 authentication on a specific BGP session.
- Password strings on both routers must match.



© 2010 Cisco Systems, Inc. All rights reserved.

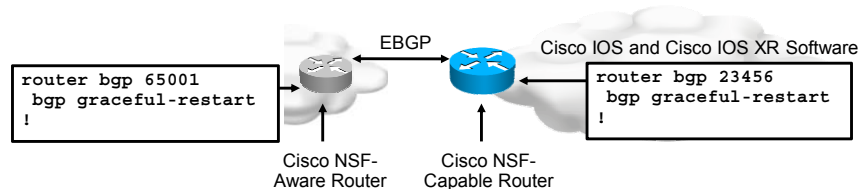
MSPRP v1.0—3-50

To enable BGP authentication, use the following:

- The **neighbor ip-address password *string*** Cisco IOS command
- The **password encrypted *string*** Cisco IOS XR command

NSF Awareness and Capability

- Distributed platforms with redundant RP support SSO:
 - Active and standby RP synchronize control plane databases.
 - A switchover without Cisco NSF requires sessions to be reset and all forwarding information to be refreshed (downtime).
 - A switchover with Cisco NSF:
 - Local forwarding information is retained on line cards.
 - Adjacent Cisco NSF-aware routers also keep the forwarding information while re-establishing sessions (BGP).



© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-3-51

Cisco NSF awareness allows a router to assist Cisco NSF-capable neighbors to continue forwarding packets during a stateful switchover (SSO) operation. The BGP Cisco NSF Awareness feature allows a Cisco NSF-aware router that is running BGP to forward packets along routes that are already known for a router that is performing an SSO operation. This capability allows the BGP peers of the failing router to retain the routing information that the failing router advertises and continue to use this information until the failed router has returned to normal operating behavior and is able to exchange routing information. The peering session is maintained throughout the entire Cisco NSF operation.

Cisco NSF works with the SSO feature in Cisco IOS Software; SSO is a prerequisite of Cisco NSF. Cisco NSF works with 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 (RP) switchover. Cisco NSF/SSO is configured in the core of your network, and Cisco NSF awareness is configured on IBGP peers in the core and on the edge of the network.

Prior to the introduction of BGP Cisco NSF with SSO in Cisco IOS Release 12.2(28)SB, BGP required that all neighboring devices participating in BGP Cisco NSF be configured to be either Cisco NSF-capable or Cisco NSF-aware. This requirement was met by configuring the devices to support the BGP graceful restart mechanism. Therefore, BGP Cisco NSF required that all neighboring devices be upgraded to a version of Cisco IOS Software that supports BGP graceful restart. However, in many Multiprotocol Label Switching Virtual Private Network (MPLS VPN) deployments, there are situations in which provider edge (PE) routers engage in EBGP peering sessions with customer edge (CE) routers that do not support BGP graceful restart. These routers cannot, therefore, be upgraded to a software version that supports BGP graceful restart in the same time frame as the provider routers.

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

Verifying BGP Cisco NSF Awareness

```
Router#show ip bgp neighbors
BGP neighbor is 192.168.2.2, remote AS YY, external link
  BGP version 4, remote router ID 192.168.2.2
  BGP state = Established, up for 00:01:18
  Last read 00:00:17, hold time is 180, keepalive interval is 60
  seconds
  Neighbor capabilities:
    Route refresh:advertised and received(new)
    Address family IPv4 Unicast:advertised and received
    Address family IPv4 Multicast:advertised and received
    Graceful Restart Capability:advertised and received
    Remote Restart timer is 120 seconds
    Address families preserved by peer:
      IPv4 Unicast, IPv4 Multicast
  Received 1539 messages, 0 notifications, 0 in queue
  Sent 1544 messages, 0 notifications, 0 in queue
  Default minimum time between advertisement runs is 30 seconds
```

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-352

To verify Cisco NSF for BGP, you must check that the graceful restart function is configured on the SSO-enabled networking device and on the neighbor devices. Verify that **bgp graceful-restart** appears in the BGP configuration of the SSO-enabled router by entering the **show running-config** command.

On peer devices (Cisco NSF-capable and Cisco NSF-aware), verify that the graceful restart function is shown as both advertised and received. Confirm the address families that have the graceful restart capability. If no address families are listed, then BGP Cisco NSF awareness is not enabled.

Summary

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

Summary

- In the service provider network, the following parameters must be monitored: CPU, memory, link utilization, and other manual audit parameters.
- Proposed changes for CPU utilization are route flap dampening, and increasing scan timers and the BGP advertisement interval.
- Proposed changes for memory utilization are to filter prefixes longer than /24 and to limit the number of routes received from a neighbor.
- Proposed changes for link utilization are BGP policy modifications using local preference, AS path prepending, and communities.
- To audit BGP configuration, consider these methods: BGP authentication and Cisco NSF setup verification.

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-353

Lesson Self-Check

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

- Q1) Which two things could happen to a BGP route that is penalized above the reuse limit but has an assigned penalty that is under the suppress limit? (Choose two.) (Source: Monitoring CPU Utilization by BGP)
- A) The route is suppressed from BGP updates if it is reachable.
 - B) The route is marked as a history entry in the BGP table.
 - C) The route is withdrawn from the IP routing table.
 - D) The route continues to be advertised.
- Q2) One of your BGP core routers is experiencing periodic slow responses to ping packets that are being directed to it from the network management console. The router has just been configured to receive complete Internet routes, and you suspect that the BGP router process is causing CPU utilization issues in the core router. Which two router commands should you use to confirm your suspicion? (Choose two.) (Source: Monitoring CPU Utilization by BGP)
- A) `show ip route`
 - B) `show ip bgp summary`
 - C) `show process cpu`
 - D) `show memory`
- Q3) What are three reasons to limit the number of BGP prefixes that are received from a neighbor? (Choose three.) (Source: Monitoring Memory Utilization by BGP)
- A) to prevent (DoS) attacks
 - B) to protect against incorrect router configuration on the neighbor side
 - C) to prevent redundant routing information from being loaded into the BGP table
 - D) to avoid overloading router memory and CPU resources
- Q4) Which Cisco IOS command is used to change the default value of local preference? (Source: Monitoring Link Utilization by BGP)
- A) **`set local-preference`**
 - B) **`bgp default local-preference`**
 - C) **`show ip bgp`**
 - D) **`show ip bgp prefix`**
- Q5) Which AS number selection is the best possible choice for a customer that is multihomed to multiple providers? (Source: Monitoring Link Utilization by BGP)
- A) a single public AS number
 - B) a single private AS number
 - C) two private AS numbers that are used with number translation
 - D) multiple private AS numbers, one used internally by the customer and the others used with AS-number translation for each provider

- Q6) Which three key pieces of information can you derive from the following router command output? (Choose three.) (Source: Monitoring Link Utilization by BGP)

```
AS387_Backup# show ip bgp 11.2.3.0
BGP routing table entry for 11.2.3.0/24, version 7
Paths: (2 available, best #1, not advertised to EBGp peer)
  Advertised to non peer-group peers:
    10.3.0.5
  Local
    0.0.0.0 from 0.0.0.0 (10.3.0.6)
      Origin incomplete, metric 0, localpref 100, weight
32768, valid,
      sourced, best
      Community: 387:31000 no-export
  Local
    10.3.0.2 (metric 128) from 10.3.0.5 (1.0.0.2)
      Origin incomplete, metric 0, localpref 100, valid,
internal
      Originator: 1.0.0.2, Cluster list: 10.3.0.5
      Community: 387:31000 no-export
```

- A) The primary link has come back up, so the backup router now sees two alternate routes.
- B) The primary link has not come back up, but the backup router still sees two alternate routes.
- C) The first route is the route that the router itself has redistributed into BGP using the floating static route. The AS locally sources this route, and it has been assigned a weight value of 32768.
- D) The second route is the one that IBGP has received from the primary edge router. The AS also sources this route, but no weight value is assigned.
- Q7) Which two of the following characteristics accurately describe the **show ip bgp neighbors** command? (Choose two.) (Source: Monitoring CPU Utilization by BGP)
- A) for neighbors with the maximum-prefix function configured, displays the maximum number of prefixes and the warning threshold
- B) for neighbors exceeding the maximum number of prefixes, displays the reason that the BGP session is idle
- C) for neighbors with unstable routes, displays the feasible successor for those routes
- D) for neighbors in confederations, displays the route reflector status of those neighbors

Lesson Self-Check Answer Key

- Q1) B, C
- Q2) B, C
- Q3) A, B, D
- Q4) B
- Q5) A
- Q6) A, C, D
- Q7) A, B

Testing Internal Routing

Overview

This lesson describes testing processes to measure convergence time, and it describes how interior gateway protocols (IGPs) calculate the alternative path. Convergence and the convergence process in the IGP-based network are described in detail.

This lesson also describes convergence time definition and the main design goals when deploying IGP in the network. Open Shortest Path First (OSPF) and Intermediate System-to-Intermediate System (IS-IS) tuning options are explained, in addition to Bidirectional Forwarding Detection (BFD) and Cisco Nonstop Forwarding (NSF) features.

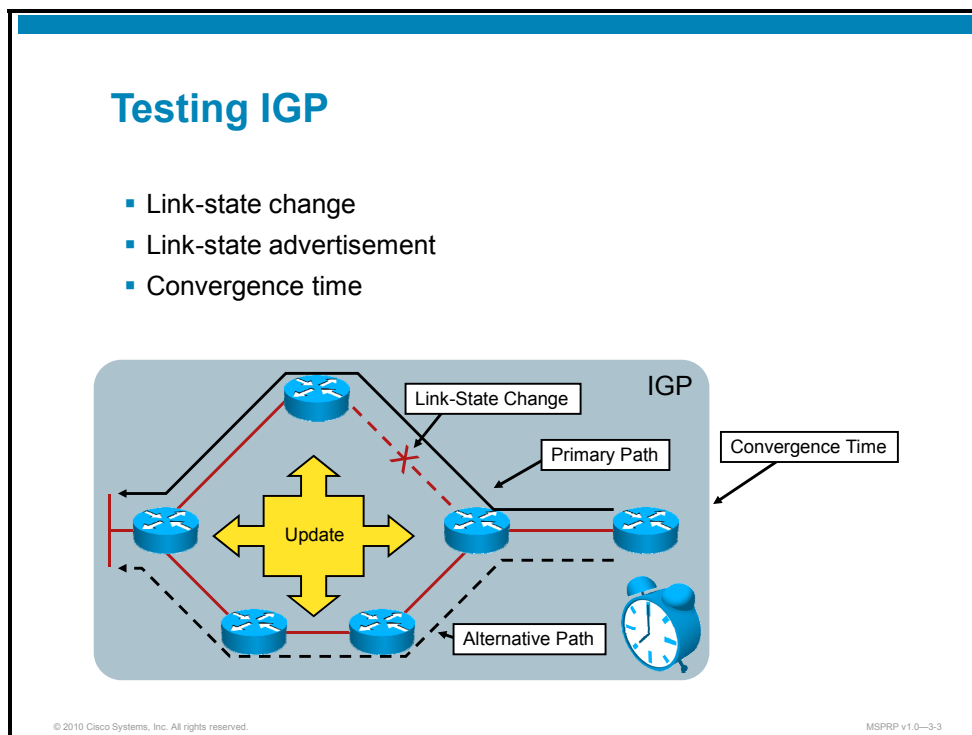
Objectives

Upon completing this lesson, you will be able to identify performance issues for OSPF and IS-IS based on planned and executed tests and propose changes to improve the performance of internal routing protocols. This ability includes being able to meet these objectives:

- Describe testing procedures for IGP performance
- Describe testing procedures for OSPF performance
- Describe testing procedures for IS-IS performance

Testing IGPs

This topic describes procedures for testing IGP performance.

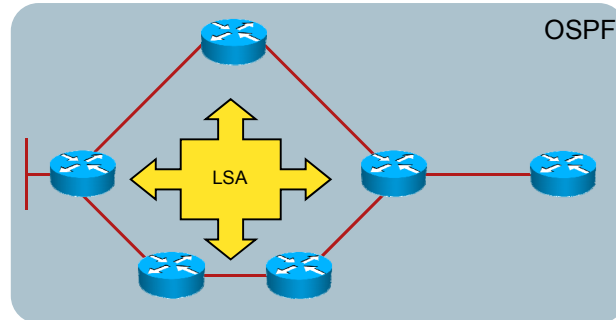


When testing IGP, the link state is changed by disabling the link on the primary path. Convergence time is measured until routers calculate the alternative path.

Link-State Advertisement in OSPF

The LSA types defined in OSPF are as follows:

- Type 1: Router LSA
- Type 2: Network LSA
- Type 3: Summary LSA
- Type 4: ASBR-Summary LSA
- Type 5: External LSA



© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0—3-4

Link-state advertisements (LSAs) are basic communication tools of the OSPF routing protocol for IP. They communicate the router's local routing topology to all other routers in the same OSPF area. OSPF is designed for scalability, so some LSAs are not flooded on all interfaces, but only on those interfaces that belong to the appropriate area. This allows detailed information to be localized while summary information is flooded to the rest of the network.

The first five LSA types defined in OSPF are as follows:

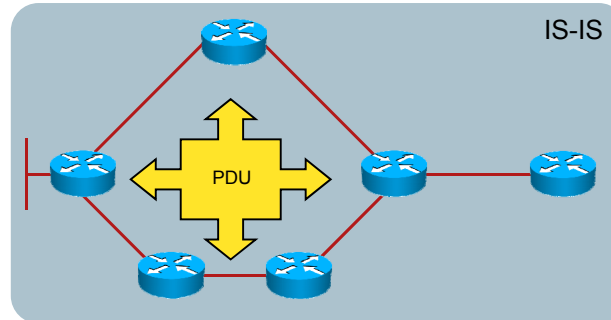
- **Type 1—Router LSA:** The router lists the links to other routers or networks in the same area, with the metric. Type 1 LSAs are flooded across their own area only. The link-state ID of the Type 1 LSA is the originating router ID.
- **Type 2—Network LSA:** The designated router (DR) on a broadcast segment lists which routers the segment joined. Type 2 LSAs are flooded across their own area only. The link-state ID of the Type 2 LSA is the IP interface address of the designated router.
- **Type 3—Summary LSA:** An Area Border Router (ABR) takes information that it has learned from one of its attached areas. It can summarize the information (but not by default) before sending it out to other areas it is connected to. This summarization helps provide scalability by removing detailed topology information for other areas; their routing information is summarized into just an address prefix and metric. The summarization process can also be configured to remove many detailed address prefixes and replace them with a single summary prefix, which also helps achieve scalability. The link-state ID is the destination network number for Type 3 LSAs.
- **Type 4—Autonomous System Boundary Router (ASBR) Summary LSA:** This is needed because Type 5 External LSAs are flooded to all areas, and detailed next-hop information may not be available in those other areas. An ABR solves this lack of information by flooding the information for the ASBR where the Type 5 originated. The link-state ID is the router ID of the described ASBR for Type 4 LSAs.

- **Type 5—External LSA:** These LSAs contain information that is imported into OSPF from other routing processes and are flooded to all areas (except stub areas). For "External Type 1" LSAs, routing decisions are made by adding the OSPF metric to get to the ASBR and the external metric from the ASBR to the rest of the network. For "External Type 2" LSAs, only the external metric is used. The link-state ID of the Type 5 LSA is the external network number.

Link-State Advertisement in IS-IS

IS-IS protocol data units

- Hello PDU (ESH, ISH, IIH)
- Link-state PDU (nonpseudonode and pseudonode)
- Partial sequence number PDU (PSNP)
- Complete sequence number PDU (CSNP)



IS-IS protocol data units (PDUs) are encapsulated directly into an Open Systems Interconnection (OSI) data-link frame. There is no Connectionless Network Protocol (CLNP) header and no IP header.

There are four types of IS-IS PDUs:

- **Hello PDU (end system hello, intermediate system hello):** Used to establish and maintain adjacencies
- **Link-state packet (LSP):** Used to distribute link-state information
- **Partial sequence number PDU (PSNP):** Used to acknowledge and request link-state information (LSPs)
- **Complete sequence number PDU (CSNP):** Used to distribute a router's complete link-state database

Convergence Time

- How fast a group of routers reach a converged state
- Main design goals: quickly converge, reliably converge
- Important performance indicator
- Network size

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-3-6

Convergence time is a measure of how fast a group of routers reach the state of convergence, and it is a primary design goal. Convergence time is an important performance indicator for routing protocols to implement a mechanism that allows all routers running this protocol to quickly and reliably converge. Of course, the size of the network also plays an important role; a larger network will converge more slowly than a smaller one.

Routing Information Protocol (RIP) is a routing protocol that converges very slowly; even a network of a few routers can take a couple of minutes to converge. When a new route is advertised, triggered updates can speed up RIP's convergence. To flush a route that previously existed will take longer due to the hold-down timers in use. OSPF is an example of a fast-converging routing protocol. A network of a few routers will converge in a matter of seconds.

Certain configuration and hardware conditions will prevent a network from ever converging. A flapping interface might cause conflicting information to propagate the network so that routers never agree on its current state. Under certain circumstances, it might even be desirable to withhold routing information from parts of the network by enforcing an unconverged network.

SPF Calculation

Routers find the best paths to destinations by applying Dijkstra's SPF algorithm to the link-state database.

- Every router in an area has the identical link-state database.
- Each router in the area places itself into the root of the tree that is built.
- The best path is calculated with respect to the lowest total cost of links to a specific destination.
- The best routes are put into the forwarding database (routing table).

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0—3-7

Edsger Dijkstra designed a mathematical algorithm for calculating the best paths through complex networks, and link-state routing protocols use Dijkstra's algorithm.

By assigning a cost to each link in the network, and by placing the specific node at the root of a tree, the branches of the tree can be calculated to determine the best path to each destination. The best paths are put in the forwarding database (routing table).

For OSPF, the default behavior is that the interface cost is calculated based on its configured bandwidth. An OSPF cost can also be manually defined for each interface, which overrides the default cost value.

OSPF and IS-IS Performance Characteristics

- To test and measure convergence, use ping or IP SLA.
- Tuning options to improve convergence:
 - Hello timers
 - Fast hellos
 - Bidirectional Forwarding Detection (BFD)
 - Cisco Nonstop Forwarding (Cisco NSF)

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-3-8

The routing protocol (OSPF or IS-IS) performance characteristics can be tested and convergence time can be measured by using ping or IP Service Level Agreements (IP SLAs).

These tuning options improve convergence:

- Hello timers
- Fast hellos
- Bidirectional Forwarding Detection (BFD)
- Cisco Nonstop Forwarding (Cisco NSF)

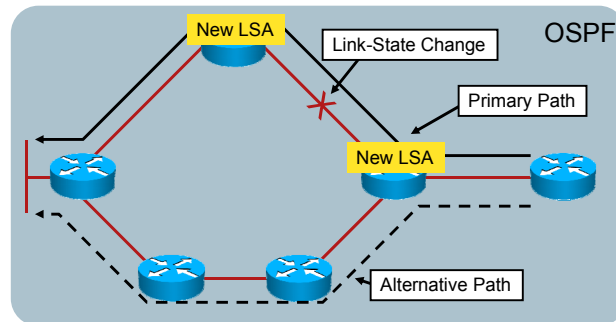
Testing OSPF

This topic describes procedures for testing OSPF performance.

OSPF Convergence Improvement

Tuning options to improve convergence:

- Hello timers
- Fast hellos
- BFD
- Cisco NSF



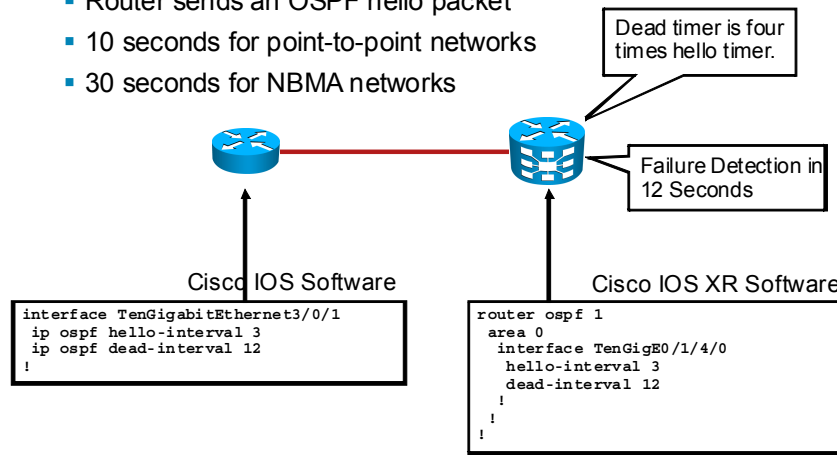
These tuning options improve convergence in an OSPF-based network:

- Adjust hello timers
- Adjust fast hellos
- Enable and adjust BFD parameters
- Enable and adjust Cisco NSF parameters

OSPF Hello Timers

Hello interval:

- Router sends an OSPF hello packet
- 10 seconds for point-to-point networks
- 30 seconds for NBMA networks



© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-311

These values are used for OSPF timers:

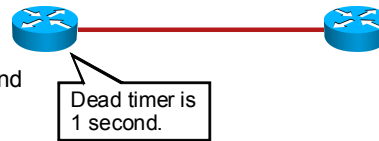
- **Hello:** Interval time in seconds that a router uses to send an OSPF hello packet. On broadcast and point-to-point links, the default is 10 seconds. On nonbroadcast multiaccess (NBMA) links, the default is 30 seconds.
- **Dead:** Time in seconds to wait before declaring a neighbor dead. By default, the dead timer interval is four times the hello timer interval.
- **Wait:** Timer interval that causes the interface to exit out of the wait period and select a DR on the network. This timer is always equal to the dead timer interval.
- **Retransmit:** Time to wait before retransmitting a database description (DBD) packet when it has not been acknowledged.
- **Hello Due In:** An OSPF timer that specifies when the next hello packet will be sent via the interface.

The example on the figure shows a configuration on the Cisco IOS and Cisco IOS XR-based routers for setting hello intervals to three seconds.

OSPF Fast Hellos

OSPF Fast Hello Packets:

- Less than one second
- Dead interval is set to one second
- Hello-multiplier value is set
- Faster convergence time
- Useful in LAN segments



Cisco IOS XR Software

```
interface TenGigabitEthernet3/0/1
ip ospf dead-interval minimal hello-multiplier 5
!
```

```
router_IOS#show ip ospf interface TenGigabitEthernet3/0/1
TenGigabitEthernet3/0/1 is up, line protocol is up
Internet Address 172.16.1.2/24, Area 0
Process ID 1, Router ID 172.17.0.2, Network Type BROADCAST, Cost:1
Transmit Delay is 1 sec, State DR, Priority 1
Designated Router (ID) 172.17.0.2, Interface address 172.16.1.2
Backup Designated router (ID) 172.16.0.1, Interface address 172.16.1.1
Timer intervals configured, Hello 200 msec, Dead 1, Wait 1, Retransmit 5
Hello due in 76 msec
< text omitted >
```

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-312

OSPF fast hello packets refer to hello packets being sent at intervals of less than one second. To understand fast hello packets, it is helpful to already understand the relationship between OSPF hello packets and the dead interval.

OSPF fast hello packets are achieved by using the **ip ospf dead-interval** command. The dead interval is set to one second. The hello-multiplier value is set to the number of hello packets you want sent during that one second, thus providing subsecond or "fast" hello packets.

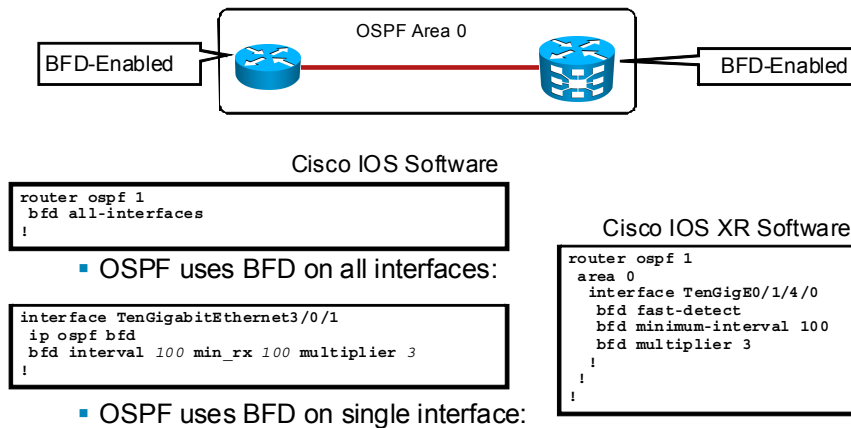
When fast hello packets are configured on the interface, the hello interval that is advertised in the hello packets that are sent over this interface is set to 0. The hello interval in the hello packets that were received over this interface is ignored.

The dead interval must be consistent on a segment, whether it is set to one second (for fast hello packets) or set to any other value. The hello multiplier need not be the same for the entire segment as long as at least one hello packet is sent within the dead interval.

The benefit of the OSPF Fast Hello Packets feature is that your OSPF network will experience faster convergence time than it would without fast hello packets. This feature allows you to detect lost neighbors within one second. It is especially useful in LAN segments, in which the OSI physical layer and the data-link layer might not detect neighbor loss.

In the example, OSPF Support for Fast Hello Packets is enabled by specifying the *minimal* keyword and the *hello-multiplier* keyword and value. Because the multiplier is set to five, five hello packets will be sent every second.

BFD for OSPF



After you have configured Bidirectional Forwarding Detection (BFD) on individual interfaces, you have to tell the routing protocols to use it. In most cases, there is no good reason that the routing protocols would not use BFD whenever possible. The only command that you have to use in these scenarios is the **bfd all-interfaces** Cisco IOS router configuration command.

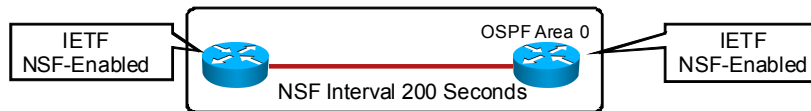
If you want to be more specific, you could enable BFD on individual interfaces with the specific command for that particular routing protocol. For OSPF, use the **ip ospf bfd** Cisco IOS interface configuration command.

You could also enable BFD on all interfaces with the **bfd all-interfaces** Cisco IOS router configuration command and disable it on specific interfaces. OSPF uses the **ip ospf bfd disable** Cisco IOS interface configuration command.

Disable BFD on an interface only if the interface is unreliable, but you want to retain routing adjacency across short failures. This course of action should not be necessary in most well-designed networks.

To enable BFD and adjust parameters, use the **bfd** Cisco IOS XR router commands.

Cisco NSF Specific to OSPF



Cisco IOS XR Software

```
router ospf 1
nsf ietf
nsf interval 200
```

RP/0/RP0/CPU0:CRS#show ospf 1

```
Routing Process "ospf 1" with ID 10.1.0.1
< text omitted >
Non-Stop Forwarding enabled
```

Cisco IOS Software

```
router ospf 1
nsf ietf interval 200
```

```
router_IOS#show ip ospf 1
Routing Process "ospf 1" with ID 172.16.2.2
< text omitted >
IETF Non-Stop Forwarding enabled
restart-interval limit: 200 sec
IETF NSF helper support enabled
Cisco NSF helper support enabled
Reference bandwidth unit is 100 mbps
...
```

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-314

OSPF Cisco NSF requires that all adjacent devices be Cisco NSF-aware. If a Cisco NSF-capable router discovers that it has non Cisco NSF-aware neighbors on a particular network segment, it disables Cisco NSF capabilities for that segment. Other network segments that are composed entirely of Cisco NSF-capable or Cisco NSF-aware routers continue to provide Cisco NSF capabilities.

The **nsf** Cisco IOS and Cisco IOS XR router command enables Cisco NSF operations for the OSPF process. The **nsf interval** Cisco IOS XR command sets the minimum time between Cisco NSF restart attempts.

The following example enables Internet Engineering Task Force (IETF) NSF for OSPF process 1 on the router. The restart interval has been changed from the 120-second default value to 200 seconds.

When the **show ip ospf** Cisco IOS command or the **show ospf** Cisco IOS XR command is entered for OSPF process 1, the output verifies that IETF NSF was configured on the router. A maximum route lifetime interval is set to 200 seconds. The command output also confirms that Cisco NSF helper mode is enabled by default for IETF NSF, as well as for Cisco NSF.

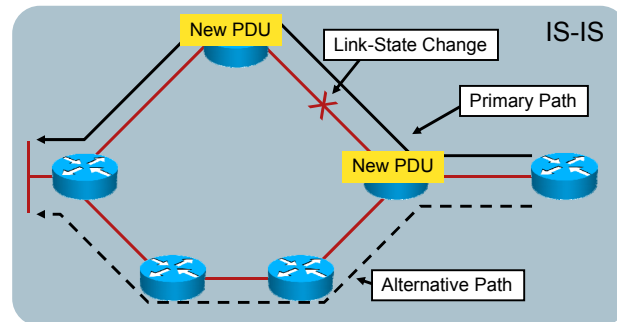
Testing IS-IS

This topic describes procedures for testing IS-IS performance.

IS-IS Convergence Improvement

Tuning options to improve convergence:

- Hello timers
- Fast hellos
- BFD
- Cisco NSF



© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0—3-16

To improve convergence in the IS-IS based network, you can use these tuning options:

- Adjust hello timers
- Adjust fast hellos
- Enable and adjust BFD parameters
- Enable and adjust Cisco NSF parameters

LSDB Synchronization

- SNP packets are used to ensure synchronization and reliability:
 - Contents are LSP descriptions
- PSNP is used:
 - For acknowledgment of LSPs on point-to-point links
 - To request missing pieces of LSDB
- CSNP is used:
 - Periodically by DIS on LAN to ensure LSDB accuracy
 - On point-to-point links when the link comes up

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-317

Sequence number PDUs (SNPs) are used to acknowledge the receipt of LSPs and to maintain link-state database (LSDB) synchronization. There are two types of SNP: CSNP and PSNP. The use of SNPs differs between point-to-point and broadcast media.

CSNPs and PSNPs share the same format. Each carries summarized LSP information. The main difference is that CSNPs contain summaries of all LSPs in the LSDB, while PSNPs contain only a subset of LSP entries.

Separate CSNPs and PSNPs are used for Level 1 and Level 2 adjacencies.

Adjacent IS-IS routers exchange CSNPs to compare their LSDB. In broadcast subnetworks, only the designated intermediate system (DIS) transmits CSNPs. All adjacent neighbors compare the LSP summaries that are received in the CSNP with the contents of their local LSDBs to determine if their LSDBs are synchronized. In other words, they determine if they have the same copies of LSPs as other routers for the appropriate levels and area of routing.

CSNPs are periodically multicast (every 10 seconds) by the DIS on a LAN to ensure LSDB accuracy.

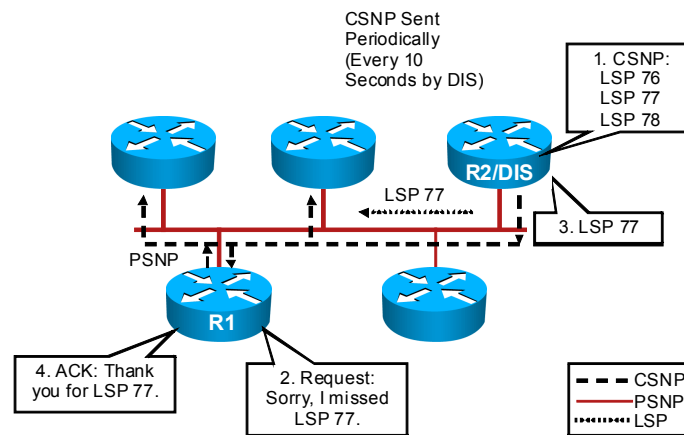
If there are too many LSPs to include in one CSNP, the LSPs are sent in ranges. The CSNP header indicates the starting and ending LSP ID in the range. If all LSPs fit in the CSNP, the range is set to default values.

Adjacent IS-IS routers use PSNPs to acknowledge the receipt of LSPs and to request transmission of missing or newer LSPs.

On point-to-point networks, CSNPs are sent when the link comes up to synchronize the LSDBs.

LSDB Synchronization

LAN



© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0—3-18

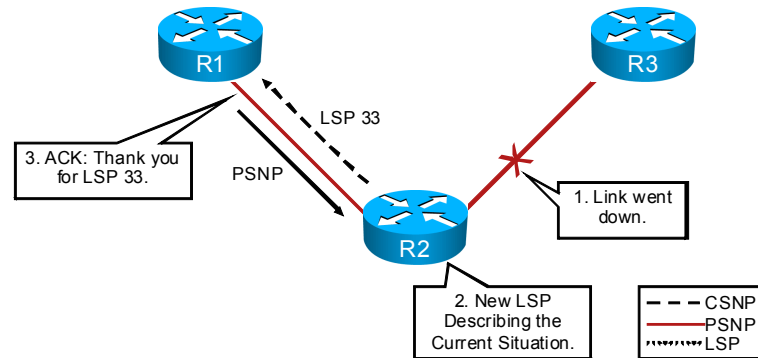
On a LAN, the DIS periodically (every 10 seconds) sends CSNPs that list the LSPs that it holds in its LSDB. This update is a multicast to all Level 1 or Level 2 IS-IS routers on the LAN.

In the example, router R1 compares this list of LSPs with its topology table and realizes that it is missing one LSP. Therefore, it sends a PSNP to the DIS (Router R2) to request the missing LSP.

The DIS reissues only that missing LSP (LSP 77), and router R1 acknowledges it with a PSNP.

LSDB Synchronization

Point-to-Point



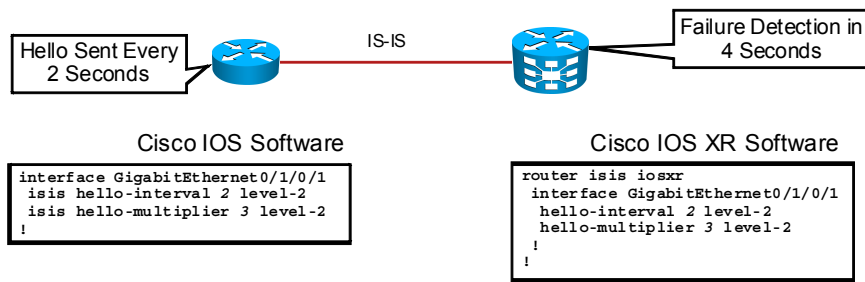
In contrast to broadcast links such as LAN links, CSNPs are not periodically sent on point-to-point links. A CSNP is sent only once, when the point-to-point link first becomes active. After that, LSPs are sent to describe topology changes, and they are acknowledged with a PSNP.

This figure shows what happens on a point-to-point link when a link failure is detected. The sequence is as follows:

- Step 1** A link fails.
- Step 2** Router R2 notices this failure and issues a new LSP noting the change.
- Step 3** Router R1 receives the LSP, stores it in its topology table, and sends a PSNP back to R2 to acknowledge receipt of the LSP.

Example: Fine-Tuning Hello Timers

- Changing hello and dead intervals can result in mismatched configuration between an adjacent pair of routers (lost adjacency):
 - Hello interval in seconds (default: 10)
 - Holdtime in seconds (default: 30; based on default multiplier of 3 and default hello time)



The hello interval is configured—per-interface—with the **isis hello-interval** Cisco IOS command and the **hello-interval** Cisco IOS XR command. Different hello intervals can be set for Level 1 and Level 2, if required.

The number of hellos to miss before declaring that the neighbor is down is specified with the **isis hello-multiplier** Cisco IOS command and the **hello-multiplier** Cisco IOS XR command (per interface and per level). The minimum value that can be set for the hello multiplier is three.

The default values are a hello interval of 10 seconds and a multiplier of 3 (so the default holdtime is 30 seconds).

The figure shows a Cisco IOS and Cisco IOS XR configuration example, in which the IS-IS hello timer is changed to two seconds for Level 2 adjacency, and the hello multiplier is changed to two.

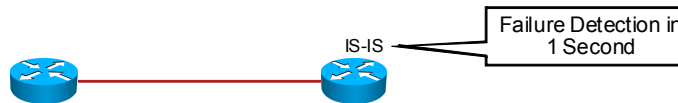
Fine-Tuning Hellos

Fast Hellos

```
router(config-if)#
```

```
isis hello-interval minimal [level-1 | level-2]  
isis hello-multiplier multiplier [level-1 | level-2]
```

- Minimum holding time has been reduced to one second; hello interval is one divided by the multiplier
- Advantage—Improved convergence time
- Disadvantages—Possible instability (adjacency flap), increased BW/buffer/CPU usage



© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0—321

Cisco IOS Software allows the hold interval to be reduced to one second. If the holdtime is set to one second, the hello time will be reduced. In this case, the hello interval is one divided by the multiplier.

This reduction could speed up neighbor loss detection but could lead to network instability. The low timeouts could lead to neighbors appearing to be lost when there is simply a short delay to their packets.

BFD for IS-IS

Cisco IOS XR Software

```
router isis 1
interface TenGigE0/1/4/0
  bfd fast-detect
  bfd minimum-interval 100
  bfd multiplier 3
```

Cisco IOS Software

```
router isis 1
  bfd all-interfaces
```

- IS-IS uses BFD on all interfaces:

```
interface TenGigabitEthernet3/0/1
ip isis bfd
bfd interval 100 min_rx 100 multiplier 3
```

- IS-IS uses BFD on specific interface only:



© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-322

BFD can be configured in two steps; the first step is setting the baseline parameters for all BFD sessions on an interface. The configuration occurs at the interface level, and the syntax is as follows:

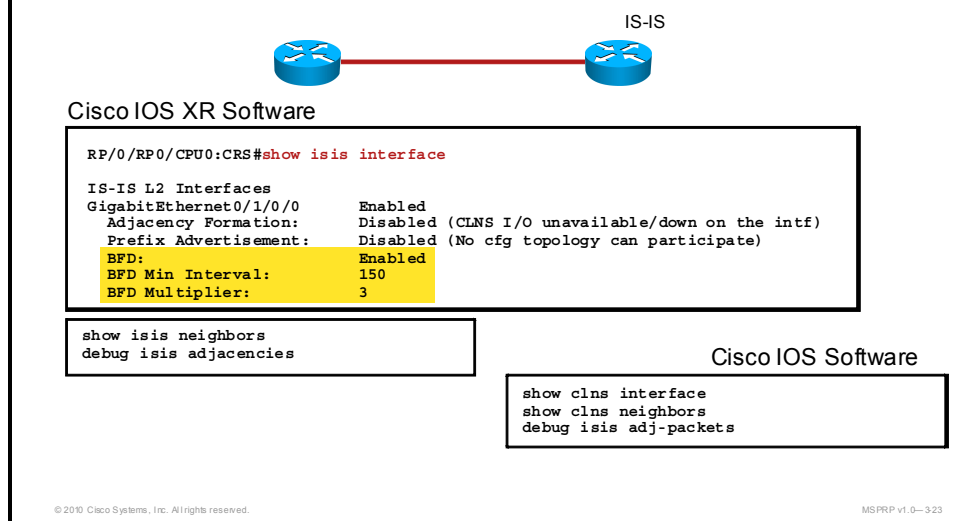
```
[no] bfd interval <50-999> min_rx <1-999> multiplier <3-50>
```

- **interval**: Determines how frequently (in milliseconds) BFD packets will be sent to BFD peers
- **min_rx**: Determines how frequently (in milliseconds) BFD packets will be expected to be received from BFD peers
- **multiplier**: The number of consecutive BFD packets that must be missed from a BFD peer before declaring that peer unavailable and informing the higher-layer protocols of the failure

Once the baseline parameters have been set, individual protocols must be informed that they will be using BFD for failure detection.

BFD for IS-IS

Monitoring



BFD is only useful if it is associated with a specific upper-layer protocol and can quickly inform that protocol about changes in the Layer 2 state. Therefore, a necessary part of troubleshooting BFD is being able to monitor the interaction between BFD and the upper-layer protocol. If IS-IS has been deployed, several extensions to existing commands have been implemented.

- The **show clns interface** Cisco IOS command and the **show isis interface** Cisco IOS XR command will now show if an interface is enabled for BFD.
- The **show clns neighbors detail** Cisco IOS command and the **show isis neighbor** Cisco IOS XR command have been enhanced to display if BFD is monitoring a particular neighbor.
- The **debug isis adj-packets** Cisco IOS command and the **debug isis adjacencies** Cisco IOS XR command will incorporate some additional output to show when IS-IS registers or deregisters with BFD.

Nonstop Forwarding for IS-IS



Cisco IOS XR Software

```
router isis iosxr
nsf ietf
nsf lifetime 200
```

Cisco IOS Software

```
router isis 1
nsf ietf interval 200
```

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0—324

The following example enables IETF NSF for IS-IS on the Cisco IOS and Cisco IOS XR router. The restart interval has been changed from the 120-second default value to 200 seconds.

Summary

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

Summary

- A state of convergence is achieved once all routing protocol-specific information has been distributed to all routers participating in the routing protocol process.
- Convergence time is a measure of how fast a group of routers reach the state of convergence.
- The routing protocol performance characteristics can be tested by using ping or IP SLA.
- To improve convergence in the OSPF or IS-IS based network, adjust hello and fast timers, and enable and adjust BFD and Cisco NSF parameters.

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-325

Lesson Self-Check

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

- Q1) What is the default value for an OSPF hello timer on broadcast and point-to-point links? (Source: Testing OSPF)
- A) 40 seconds
 - B) 10 seconds
 - C) 20 seconds
 - D) 30 seconds
- Q2) By default, the dead timer interval is how many times the hello timer interval for OSPF? (Source: Testing OSPF)
- A) 3 times
 - B) 2 times
 - C) 5 times
 - D) 4 times
- Q3) What is the default value for the IS-IS hello timer? (Source: Testing IS-IS)
- A) 10 seconds
 - B) 5 seconds
 - C) 15 seconds
 - D) 20 seconds
- Q4) What is the minimum value that can be set for a hello multiplier in IS-IS? (Source: Testing IS-IS)
- A) 2
 - B) 4
 - C) 3
 - D) 5
- Q5) How quickly is failure detected in OSPF and IS-IS when fast hello packets are enabled? (Source: Testing OSPF and Testing IS-IS)
- A) in 1 second
 - B) in 2 seconds
 - C) in 0.5 second
 - D) in 3 seconds
- Q6) Which feature enables OSPF and IS-IS to converge after failure in less than one second? (Source: Testing OSPF and Testing IS-IS)
- A) fast hello packets
 - B) adjusting timers
 - C) BFD
 - D) Cisco NSF

Lesson Self-Check Answer Key

Q1)	B
Q2)	D
Q3)	A
Q4)	C
Q5)	A
Q6)	C

Testing External Routing

Overview

This lesson describes testing processes to measure Border Gateway Protocol (BGP) convergence time. Convergence and the convergence process in the BGP-based network are described in detail.

The lesson defines convergence time and describes the main design goals when deploying BGP in the network. BGP tuning options are explained, including BGP timers and additional features that are used to improve BGP convergence time.

Objectives

Upon completing this lesson, you will be able to identify performance issues for BGP on planned and executed tests and propose changes to improve BGP performance. This ability includes being able to meet these objectives:

- Describe BGP performance characteristics
- Describe the BGP convergence process
- Test EBGP and IBGP convergence
- Propose improvements that will improve BGP convergence

Testing BGP

This topic describes general BGP performance characteristics, testing procedures for BGP performance, and improving BGP convergence time.

BGP Performance Characteristics

- BGP is designed to scale to huge internetworks.
- Reliable, no periodic updates.
- Some design tradeoffs were made:
 - BGP uses TCP for reliable transport (CPU-intensive).
 - Scalability is the top priority (slower convergence).
- Periodic keepalives verify TCP connectivity.
- Triggered updates are batched and rate-limited.
- BGP scans the routing table for changes every 60 seconds.

© 2010 Cisco Systems, Inc. All rights reserved.

MSRPv1.0-3.3

The designers of the BGP have succeeded in creating a highly scalable routing protocol that can forward reachability information between autonomous systems. The designers had to consider an environment with an enormous number of reachable networks and complex routing policies.

The reliable transport mechanism that BGP uses is standard TCP. Because BGP uses a reliable transport, the sender knows that the receiver has actually received the transmitted information. This capability makes periodic updates unnecessary. Deciding to choose TCP kept BGP simple, but it increased the CPU resource requirements for routers running BGP. Whenever there was a choice between fast convergence and scalability, scalability was the top priority. The batching of updates and the relatively low frequency of keepalive packets are examples of designers placing convergence time second to scalability.

A router that has received reachability information from a BGP peer must be sure that the peer router is still there. Otherwise, the router could route traffic toward a next-hop router that is no longer available, causing IP packet loss. TCP does not provide the service to signal that the TCP peer has been lost, unless some application data is actually transmitted between the peers. In an idle state, when there is no need for BGP to update its peer, the peer could be unreachable without TCP detecting it. Therefore, BGP takes care of detecting the presence of neighbors by periodically sending small BGP keepalive packets to them. These packets are considered application data by TCP and therefore must be transmitted reliably. According to the BGP specification, the peer router must also reply with a BGP keepalive packet. Keepalive messages are sent to the BGP neighbor every 60 seconds by default.

When BGP was created, an important design goal was the ability to manage enormous amounts of routing information in very large and complex networks. In this environment, many links could go up and down (link flapping), causing topology changes that the routing protocol should consider. But low convergence time and quick responses to topology changes require fast updates and high CPU power to process both incoming and outgoing updates. The larger the network, the more updates per second that can be expected if immediate response is required. The presence of too many updates in large networks can jeopardize network scalability.

The designers of BGP decided that scalability was a more important issue than fast convergence time, so BGP was designed to batch updates. Any changes that are received within the batch interval time are saved. At the end of the interval, only the remaining result is forwarded in an outgoing update. If the network flaps several times during the batch interval, only the state at the end of the interval is sent in an update. The batching feature avoids an uncontrolled flood of updates all over the Internet because batching procedure limits the number of updates.

The BGP scanner process walks (scans) the BGP table and confirms the reachability of next hops. Increasing the BGP scanner process frequency will make the router find a changed status more quickly, but it will also consume more CPU resources.

BGP Convergence

- As the number of routes in the Internet routing table grows, the time it takes for BGP to converge increases.
- Network convergence times can range from 10 minutes to more than 1 hour.
- BGP is considered converged when:
 - All routes have been accepted.
 - All routes have been installed in the routing table.
 - The table version for all peers equals the table version of the BGP table.
 - The input and output queues for all peers are empty.
- Lowering BGP timers speeds up convergence.
- To test and measure convergence, use ping or IP SLA.
- **Reduced BGP timers increase CPU usage.**

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRPv1.0-3-4

As the number of routes in the Internet routing table grows, service providers and large enterprise customers are experiencing a dramatic increase in the time that BGP takes to converge. Networks that converged in 10 or 15 minutes may now need as much as 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 specific conditions have been met.

- All routes have been accepted.
- All routes have been installed in the routing table.
- The table version for all peers equals the table version of the BGP table.
- The input and output queues for all peers are empty.

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

You can test and measure BGP convergence using ping or IP Service Level Agreements (IP SLA). Be aware that decreasing BGP timers will speed up convergence but will increase CPU load.

BGP Timers

- **Keepalive** BGP messages provide verification of neighbor existence.
- **Keepalive** messages are sent every **60** seconds.
- Default BGP **holdtime** value is **180** seconds.
- **Advertisement interval timer** acts as a form of rate-limiting on a per-destination basis:
 - Triggered updates are batched and rate-limited.
 - Default is **30 seconds** for **EBGP** neighbors.
 - Default is **5 seconds** for **IBGP** neighbors.
- **Scan** time specifies scanning interval of BGP routing information.
- Reduced timers increase CPU usage!

© 2010 Cisco Systems, Inc. All rights reserved.

MSRPv1.0-3-5

BGP uses certain timers to control periodic activities, such as sending keepalive messages, and the interval after not receiving a keepalive message after the peer is declared dead. TCP-based BGP sessions do not provide any means of verifying the presence of a BGP neighbor. After BGP has established the TCP session, the only method of verifying that neighbor presence is for the neighbor to actually send BGP traffic. BGP traffic is sent over the TCP session with acknowledgments (ACKs) and is therefore reliable. Successfully sending BGP traffic confirms the existence of a BGP neighbor.

However, there are often long periods when no BGP traffic is sent between neighbors. During those periods, TCP does not implement a mechanism to check for the existence of the configured neighbor. BGP neighbors could therefore be easily disconnected during times of session inactivity. This situation would lead to incorrect routing information on the other side of the BGP session.

To avoid routing packets to a router that is no longer there, BGP needs an additional mechanism to make sure that a neighbor exists. BGP sends special keepalive messages during every keepalive interval to inform its peer of its presence. By default, this interval is every 60 seconds. If no BGP traffic is received within the selected holdtime interval, the BGP router sends a BGP notification message to the inactive peer and tears down the BGP session between them. The default BGP holdtime value is 180 seconds. You can adjust those timers. When a connection is started, BGP will negotiate the hold time with the neighbor. The smaller of the two hold times will be chosen. The keepalive timer is then set based on the negotiated hold time and the configured keepalive time.

When changing the default values of keepalive and holdtime intervals, you must take care not to configure too large a keepalive interval in comparison to the hold time. Too large of a difference could result in resetting the BGP session after only one keepalive message has been missed, making a network unstable. The suggested ratio of keepalive to holdtime interval is 1:3.

When a BGP-speaking router sends a route update to a neighbor for a specific destination, it is not allowed to send another update to the neighbor about the same destination until a period equal to the advertisement interval has elapsed. In that way, the advertisement interval timer acts as a form of rate limiting on a per-destination basis, even though the value of the advertisement interval is configured for each neighbor.

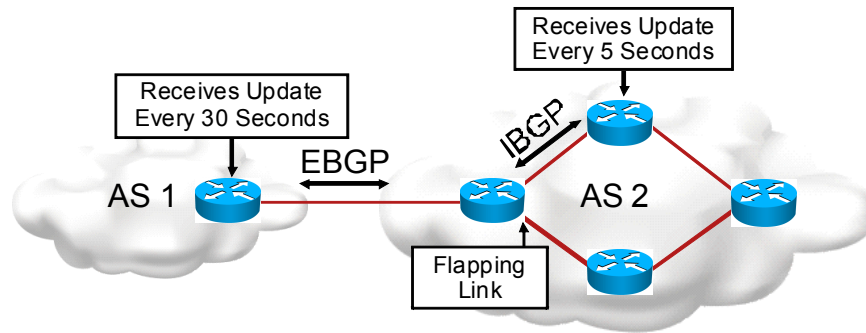
It is important to note that while the router is waiting for the advertisement interval timer to expire, the router can still receive and process route updates from BGP neighbors. The advertisement interval does not rate-limit BGP route selection (inbound updates and subsequent processing), but it does limit the rate of outgoing route advertisements. When faster propagation of successive BGP updates (which are batched and rate-limited) is required, the network administrator can lower the default value of the advertisement interval, thus improving convergence.

Scan time specifies the scanning interval at which BGP routing information and next-hop reachability is checked.

In routers that manage large BGP tables and less stable networks, lowering the timers could potentially lead to consuming large portions of router resources, especially CPU.

Link Flapping and Path Recalculation

- **Link flapping:** Link repeatedly goes up and down.
- If a route **flaps** several times within the batch period, only the last update is being sent in the batch.
- Less best-path recalculations for BGP routers.



Many links can go up and down (flapping) in very large and complex networks, which service provider networks are. So-called flapping is causing topology changes, which routing protocols should consider. Every time a BGP router sends an update, it starts a 5-second timer for internal neighbors and a 30-second timer 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 repeatedly 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. When faster propagation of successive BGP updates (which are batched and rate-limited) is required, the network administrator can lower the default value of the advertisement interval, thus improving convergence.

CPU Effects of BGP Processes

- BGP scanner process:
 - Walks BGP table and confirms reachability of the next hops every 60 seconds.
 - High CPU utilization stemming from the BGP scanner process can be expected for short durations.
 - While the BGP scanner runs, low priority processes need to wait a longer time to access the CPU.
- BGP router process:
 - Calculates the best BGP path and processes any route changes every second.
 - BGP router process consumes all free CPU cycles.
- Check the CPU load caused by BGP processes.

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRPv1.0-3-7

The BGP scanner process “walks” (scans) the BGP table and confirms next-hop reachability. The BGP scanner also checks conditional advertisement to determine whether BGP should advertise conditional prefixes. On routers that carry a large Internet routing table, high CPU utilization stemming from the BGP scanner process can be expected for short periods. Once per minute, the BGP scanner walks the BGP routing table and performs important maintenance tasks. These tasks include checking the next hop that is referenced in the router's BGP table and verifying that the next-hop devices can be reached. Thus, a large BGP table takes an equally large amount of walking and validation time.

The BGP scanner runs through the entire BGP table. Therefore, the duration of the high CPU utilization condition that the BGP scanner process causes, varies with the number of neighbors and of routes that are learned per neighbor. While the BGP scanner runs, low-priority processes need to wait longer to access the CPU. One low-priority process controls Internet Control Message Protocol (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 for the BGP scanner process to finish. The BGP scanner process runs for a period, is suspended; then ICMP runs, ICMP is suspended; 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 CPU processes directly on the router.

The BGP router process calculates the best BGP path and processes any route changes. It also sends and receives routes, establishes peers, and interacts with the Routing Information Base. 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.

You can use the **show process cpu | include bgp** Cisco IOS Software command or the **show process cpu | include bgp** Cisco IOS XR Software command to see the volume of CPU resources that are consumed (utilized) when running BGP processes.

Improving BGP Convergence

- BGP convergence can be improved to some extent by:
 - Configuring a smaller interval for the BGP scanner process (scan time)
 - Configuring a smaller advertisement interval between BGP neighbors
 - Configuring a smaller keepalive and hold timer
- **Limitations:**
 - Not recommended in routers dealing with large BGP tables
 - Could lead to CPU or memory exhaustion

© 2010 Cisco Systems, Inc. All rights reserved.

MSRPv1.0—3-8

The convergence times were not considered important in initial BGP implementations and were commonly sacrificed to reduce the CPU utilization that BGP imposed on the router. There are four BGP parameters that can influence BGP convergence speed.

- **Scan time:** Controls the BGP scanner process responsible for verifying information in the BGP table.
- **Advertisement interval:** Controls the rate at which successive advertisements are sent to a BGP neighbor.
- **Keepalive interval:** Verifies the presence of a BGP neighbor.
- **Holdtime interval:** If no BGP traffic is received within the selected holdtime interval, the BGP router sends a BGP notification message to the inactive peer and tears down the BGP session between them.

Network administrators must take care when configuring these four 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.

Improving BGP Convergence (Cont.)

- **Fast-external-fallover for EBGp sessions:**
 - Immediately resets external BGP peering session if the link used to reach these peers goes down
 - Enabled by default
 - Configured globally
- **Fast-peering deactivation for EBGp and IBGP sessions:**
 - As soon as the IP address of the BGP peer disappears from the IP routing table, the session is deactivated.
 - Immediate convergence.
 - Configured per-neighbor.
- **Cisco Nonstop Forwarding Awareness**
- **BGP Support for Nonstop Routing with SSO**

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRPv1.0-3.9

There are some other improvements and tuning mechanisms that improve overall BGP convergence.

Fast external loss detection immediately resets external BGP peering sessions if the link used to reach these peers goes down. The fast external neighbor loss detection functionality is enabled by default, but you can disable it (for example, if the flapping external link continuously disrupts EBGp session). Interface flap no longer causes immediate loss of EBGp sessions. The EBGp sessions are disconnected only after the BGP hold time expires.

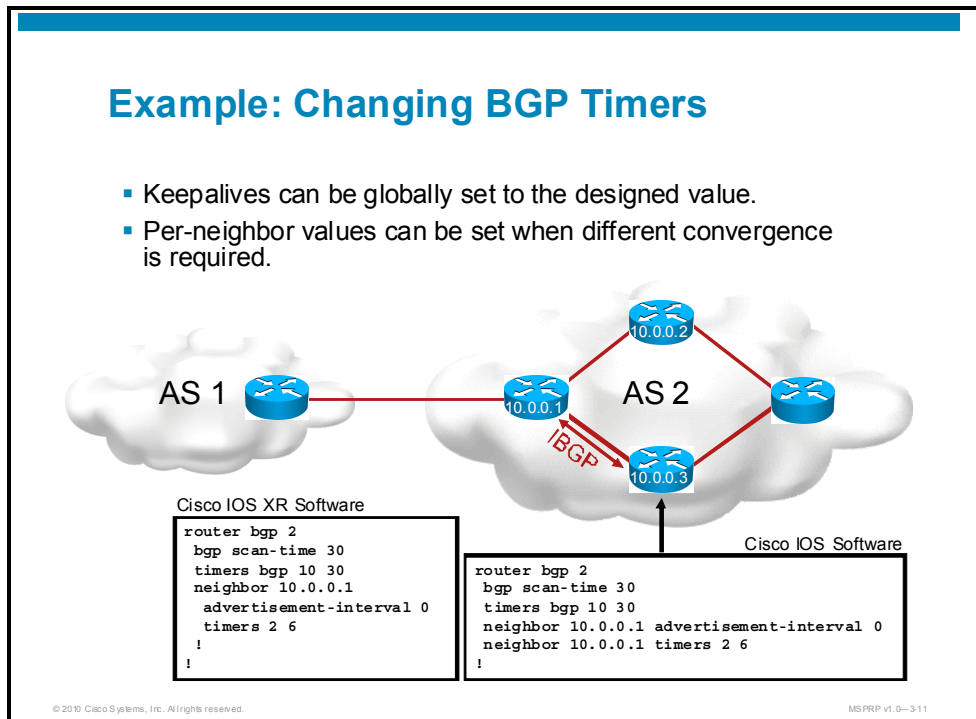
The BGP support for the Fast-Peering Session Deactivation feature introduces an event-driven notification system that allows a BGP process to monitor BGP peering sessions on a per-neighbor basis. This feature improves the response time of BGP to adjacency changes by allowing BGP to detect an adjacency change and deactivate the terminated session between standard BGP scanning intervals. Enabling this feature improves overall BGP convergence.

Cisco Nonstop Forwarding (Cisco NSF) awareness allows a Cisco NSF-aware router to assist Cisco NSF-capable and Cisco NSF-aware neighbors in continuing to forward packets during a switchover operation or during a well-known failure condition.

The BGP Support for Nonstop Routing (NSR) with Stateful Switchover (SSO) feature enables provider edge (PE) routers to maintain the BGP state with customer edge (CE) routers. NSR with SSO ensures continuous packet forwarding during a route processor (RP) switchover or during a planned In-Service Software Upgrade (ISSU) for a PE router. CE routers do not need to be Cisco NSF-capable or Cisco 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. BGP NSR with SSO, thus enables service providers to supply the benefits of Cisco NSF, with the additional benefits of NSR, without requiring CE router upgrades to support BGP graceful restart.

Testing IBGP

This topic describes configuration for improving IBGP convergence and testing procedures for IBGP performance.



You can use the **bgp scan-time** Cisco IOS command to configure the time interval for repetitions of the BGP scanner process and thus improve the convergence on IBGP sessions. The BGP scanner process walks the BGP table and confirms the reachability of next hops. A change of this status triggers a new BGP route selection for the network. The router then propagates changes to established BGP neighbors. Increasing the BGP scanner process frequency will make the router find a changed status more quickly, but it will also consume more CPU resources.

The minimum valid value that could be used for the scanning interval is five seconds. You can check the configured BGP scan interval using the **show ip bgp summary** command. The configured BGP scan interval will apply to the entire BGP routing protocol process.

With the help of the **neighbor advertisement-interval** command, network administrators can modify the default advertisement interval for a specific BGP peer. The router, which is using the advertisement interval timer, rate-limits BGP advertisements. When a BGP-speaking router sends a route update to a neighbor for a specific destination, it is not allowed to send another update to the neighbor about the same destination until a period equal to the advertisement interval has elapsed. Using the **neighbor advertisement-interval** command does not rate-limit BGP route selection (inbound updates and subsequent processing), but only the rate of outgoing route advertisements.

When faster propagation of successive BGP updates (which are batched and rate-limited) is required, the network administrator can lower the default value of the advertisement interval, thus improving convergence. You can examine the currently configured BGP advertisement interval with the **show ip bgp neighbors** command. The advertisement interval is defined for a specific neighbor in a specific BGP address family. Actual values of the advertisement interval are therefore stated under the specific address family portion of the neighbor output.

BGP uses certain timers to control periodic activities such as sending keepalive messages and the interval after not receiving a keepalive message, after which a peer is declared dead. By default, the keepalive timer is 60 seconds, and the hold timer is 180 seconds. You can adjust these timers. Reduced keepalive and holdtime values always result in increased CPU utilization. You can examine the currently configured BGP keepalive and holdtime interval with the **show ip bgp neighbors** command.

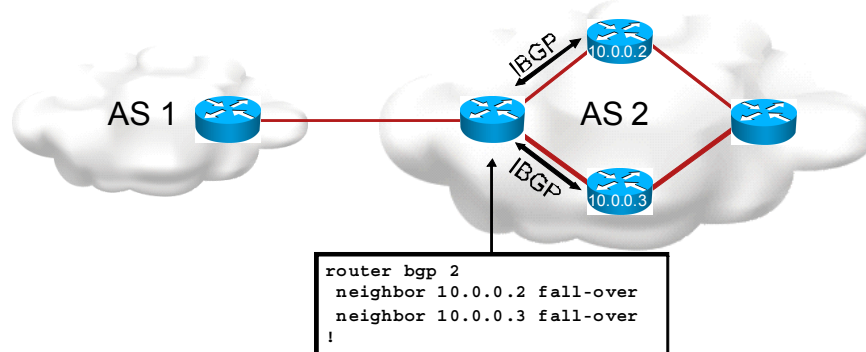
Configuration for changing default BGP timers in Cisco IOS XR Software is identical to Cisco IOS Software configuration. Therefore, you can use the **bgp scan-time** command to configure the time interval for repetitions of the BGP scanner process and thus improve the convergence on IBGP sessions. You can also modify the default advertisement interval for a specific BGP peer by using the **neighbor advertisement-interval** command. Another option of improving BGP convergence is adjusting default BGP timers by using the **timers bgp** command.

The example shows the Cisco IOS configuration for changing default values for BGP timers. The default advertisement interval for IBGP neighbors is five seconds.

IBGP Fast-Peering Deactivation

Cisco IOS Software

- As soon as the IP address of the BGP peer disappears from the IP routing table, the BGP session with the internal peer is deactivated.

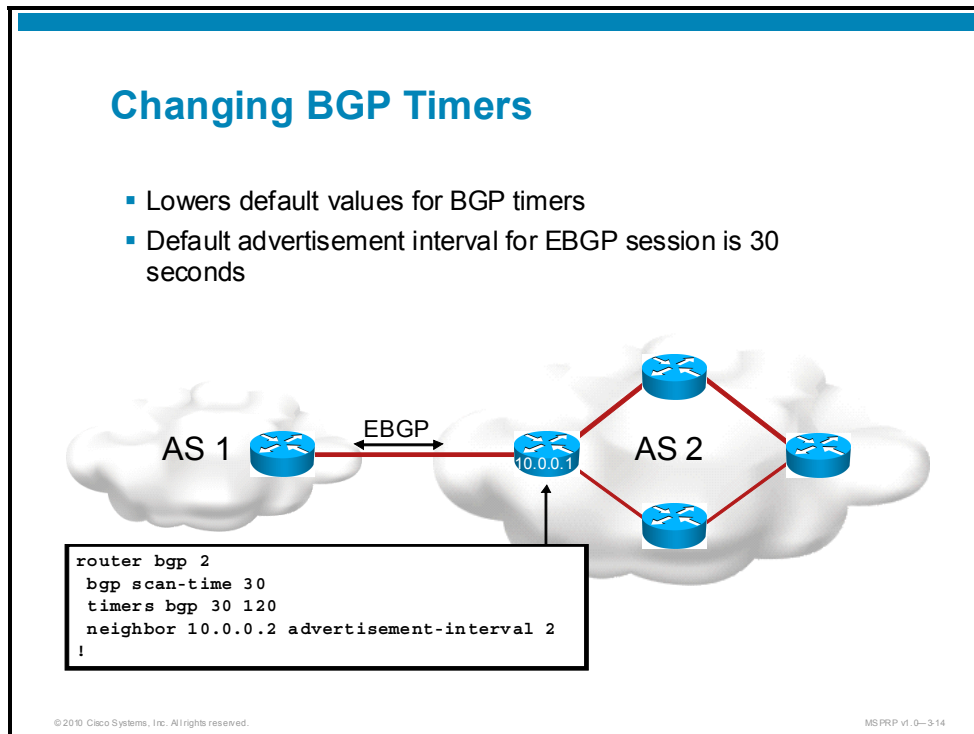


The **neighbor fall-over** Cisco IOS command is used to enable the BGP Fast-Peering Session Deactivation. BGP Fast-Peering Session Deactivation improves BGP convergence and response time to adjacency changes with BGP neighbors. This feature is event-driven and configured on a per-neighbor basis. When this feature is enabled, BGP will monitor the peering session with the specified neighbor. Adjacency changes are detected and terminated peering sessions are deactivated between the default or configured BGP scanning interval.

You should use the fast session deactivation only if you are running a fast converging interior gateway protocol (IGP). Your IGP should be able to find an alternate route to the BGP peer (if one exists) immediately. If there is even a slight interruption between the moment the original route to the BGP peer is lost until an alternate route is inserted in the IP routing table, the BGP session would already be disconnected. (There is no holddown or delay in BGP session deactivation.)

Testing EBG

This topic describes the testing procedures for EBG performance and the configuration for improving EBG convergence.

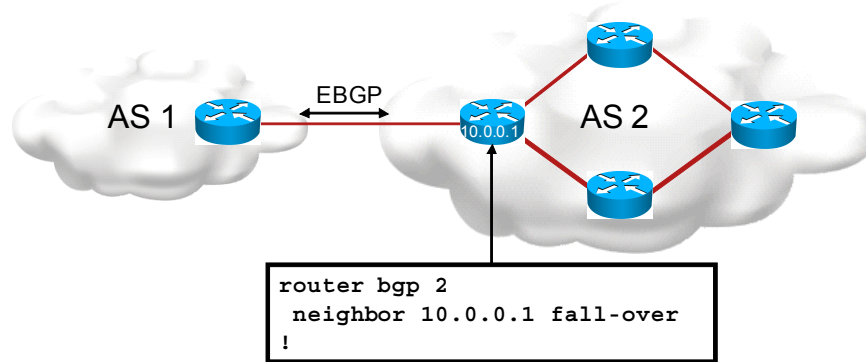


You can change default BGP timers using the same commands that are explained in the testing IBGP topic. The example shows the configuration for changing the default values for BGP timers. The default advertisement interval for EBG neighbors is 30 seconds.

EBGP Fast-Peering Deactivation

Cisco IOS Software

- You can quickly detect failures of EBGP sessions established between **loopback** interfaces of EBGP peers.
- Configured per neighbor.



The Cisco IOS **bgp fast-external-fallover** command is used to disable or enable fast external fallover for BGP peering sessions with directly connected external peers. The session is immediately reset if the link goes down. Only directly connected peering sessions are supported.

If BGP fast external fallover is disabled, the BGP routing process will wait until the default hold timer expires (three keepalives) to reset the peering session. BGP fast external fallover can also be configured on a per-interface basis using the **ip bgp fast-external-fallover interface** configuration command.

Fast external fallover functionality is enabled by default. You can disable it globally with the **bgp fast-external-fallover disable** Cisco IOS XR command (for example, if flapping external links continuously disrupt EBGP sessions). After the **bgp fast-external-fallover disable** has been configured, the interface flap no longer causes immediate loss of EBGP sessions.

The BGP fast-session deactivation Cisco IOS XR command **neighbor fall-over** works on all BGP sessions. You can use it to quickly detect failures of EBGP sessions that are established between loopback interfaces of EBGP peers (a common scenario in which to implement simple EBGP load sharing). Or, you can use it to detect EBGP neighbor loss when you disable fast external fallover.

Fast external fall-over is a global setting, whereas fast session deactivation is configured per-neighbor. You can thus disable fast external fallover with the **no bgp fast-external-fallover** Cisco IOS XR command and retain the quick response to interface failures for a selected subset of EBGP. The EBGP fast session deactivation is identical to the IBGP case; the configuration command to use is **neighbor fall-over**.

BGP Cisco NSF Awareness

- Cisco NSF allows a Cisco NSF-aware router to assist Cisco NSF-capable neighbors in continuing to forward packets during an SSO operation.
- SSO feature is a prerequisite of Cisco NSF.
- It minimizes the amount of time a network is unavailable following an RP switchover.
- Device is Cisco NSF-aware if it is running Cisco NSF-compatible software (incapable of performing an SSO operation).
- Device is Cisco NSF-capable if it has been configured to support Cisco NSF (capable of performing an SSO operation).
- All BGP neighbors must be Cisco NSF-capable or Cisco NSF-aware.

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-3.16

Cisco NSF awareness allows a router to assist Cisco NSF-capable neighbors to continue forwarding packets during an SSO operation. The BGP Cisco NSF Awareness feature allows a Cisco NSF-aware router that is running BGP to forward packets along routes that are already known for a router that is performing an SSO operation. This capability allows the BGP peers of the failing router to retain the routing information that the failing router advertises. They can continue to use this information until the failed router has returned to normal operating behavior and is able to exchange routing information. The peering session is maintained throughout the entire Cisco NSF operation.

SSO is a prerequisite of Cisco NSF. Cisco NSF works with 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 an RP switchover.

A networking device is said to be Cisco NSF-aware if it is running Cisco NSF-compatible software. A device is said to be Cisco NSF-capable if it has been configured to support Cisco NSF. A Cisco NSF-capable device would rebuild routing information from Cisco NSF-aware or Cisco NSF-capable neighbors. All neighboring devices participating in BGP Cisco NSF must be Cisco NSF-capable or Cisco NSF-aware, having been configured for BGP graceful restart.

BGP depends on Cisco Express Forwarding to continue forwarding packets during switchover while the BGP rebuilds the Routing Information Base (RIB) tables. Once the BGP has converged, Cisco Express Forwarding updates the FIB table and removes stale route entries. Cisco Express Forwarding then updates the line cards with the new FIB information.

BGP support for Cisco NSF requires that neighbor routers are Cisco NSF-aware or Cisco NSF-capable. The Cisco NSF graceful restart mechanism enables Cisco NSF awareness in BGP. A router that is Cisco NSF-aware functions like a router that is Cisco NSF-capable with one exception: a Cisco NSF-aware router is incapable of performing an SSO operation. However, a router that is Cisco NSF-aware is capable of maintaining a peering relationship with a Cisco NSF-capable neighbor during a Cisco NSF SSO operation. In addition, a Cisco NSF-aware router can hold routes for this neighbor during the SSO operation.

The BGP Cisco NSF Awareness feature makes a Cisco NSF-aware router capable of detecting a neighbor that is undergoing an SSO operation, maintaining the peering session with this neighbor, retaining known routes, and continuing to forward packets for these routes. Deploying BGP Cisco NSF awareness can minimize the effects of RP failure conditions and improve overall network stability by reducing the amount of resources that are normally required for reestablishing peering with a failed router.

Configuring BGP Cisco NSF Awareness

Cisco IOS and Cisco IOS XR Software

```
router(config-router)#
```

```
bgp graceful-restart restart-time 130  
bgp graceful-restart stalepath-time 350
```

```
RP/0/RP0/CPU0:CRS(config-bgp)#
```

```
bgp graceful-restart restart-time 130  
bgp graceful-restart stalepath-time 350
```

- Cisco NSF awareness for BGP is not enabled by default.
- Enables Cisco NSF awareness on router that is running BGP.
- BGP graceful restart timers are configurable.
- Restart time is, by default, 120 seconds.
- Stale path time is, by default, 360 seconds.

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-3.17

Cisco NSF awareness for BGP is not enabled by default. The **bgp graceful-restart** command is used to enable Cisco NSF awareness on a router that is running BGP. BGP Cisco NSF awareness allows Cisco NSF-aware routers to support Cisco NSF-capable routers during an SSO operation. Cisco NSF awareness is not enabled by default and should be configured on all neighbors that participate in BGP Cisco NSF. If you enter this command after the BGP session has been established, you must restart the session for the exchange of capability with the BGP neighbor.

If the **bgp graceful-restart** command has been issued after the BGP session has been established, you must reset by issuing the **clear ip bgp *** command or by reloading the router before graceful restart capabilities will be exchanged.

BGP graceful restart timers are configurable. There are two BGP graceful restart timers.

- The **restart-time** argument determines how long peer routers will wait to delete stale routes before a BGP open message is received. The default value is 120 seconds.
- The **stalepath-time** argument determines how long a router will wait before deleting stale routes after an end of record message is received from the restarting router. The default value is 360 seconds.

The configuration of the restart and stalepath timers is not required to enable the BGP graceful restart capability. The default values are optimal for most network deployments, and the experienced network operator should adjust them.

BGP Support for NSR with SSO

- There are situations in which PE routers in EBGp sessions do not support BGP graceful restart.
- Enables PE routers to maintain the BGP state with CE routers and ensure continuous packet forwarding.
- CE routers do not need to be Cisco NSF-capable or Cisco NSF-aware.
- PE routers have to support BGP NSR with SSO.
- Eliminates route flaps by keeping BGP state on both active and standby RPs and ensures minimal packet loss during RP failovers.

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-318

The BGP Support for NSR with SSO feature enables PE routers to maintain the BGP state with CE routers. NSR with SSO ensures continuous packet-forwarding during an RP switchover or during a planned ISSU for a PE router. CE routers do not need to be Cisco NSF-capable or Cisco 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. BGP NSR with SSO enables service providers to supply the benefits of Cisco NSF with the additional benefits of NSR, without requiring an upgrade to CE routers to support BGP graceful restart.

Prior to the introduction of BGP NSR with SSO, BGP required that all neighboring devices participating in BGP Cisco NSF be configured to be either Cisco NSF-capable or Cisco NSF-aware (by configuring the devices to support the BGP graceful restart mechanism). BGP Cisco NSF required that all neighboring devices be upgraded to a version of Cisco IOS or Cisco IOS XR Software that supports BGP graceful restart. However, in many Multiprotocol Label Switching (MPLS) virtual private network (VPN) deployments, there are situations in which PE routers engage in EBGp peering sessions with CE routers that do not support BGP graceful restart. These routers 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 RP. 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. Additionally, the BGP NSR with SSO feature dynamically detects Cisco NSF-aware peers and runs graceful restart with those CE routers. For EBGp peering sessions with Cisco NSF-aware peers and for IBGP sessions with BGP route reflector routers in the service provider core, the PE uses Cisco NSF to maintain the BGP state.

BGP NSR with SSO reduces the impact on customer traffic during RP switchovers (scheduled or unscheduled events), extending high-availability deployments and benefits at the edge. BGP NSR with SSO also eliminates route flaps by keeping the BGP state on both active and standby route processors and ensures continuous packet-forwarding with minimal packet loss during RP failovers.

Configuring BGP for NSR with SSO

Cisco IOS and Cisco IOS XR Software

```
router(config)#
```

```
router bgp 1
  bgp graceful-restart
  neighbor 10.1.1.1 ha-mode sso
```

- Supported in BGP peer, BGP peer group, and BGP session template.
- On PE router, configure BGP graceful restart and support for BGP NSR with SSO.

```
RP/0/RP0/CPU0:CRS(config)#
```

```
router bgp 1
  nsr
```

- Supported in Cisco IOS XR Software Release 3.8.0 and later.

© 2010 Cisco Systems, Inc. All rights reserved.

MS PRP v1.0-319

Prior to the introduction of BGP NSR with SSO in Cisco IOS Release 12.2(28)SB, BGP required that all neighboring devices participating in BGP Cisco NSF be configured to be either Cisco NSF-capable or Cisco NSF-aware.

BGP NSR with SSO is supported in the BGP peer, BGP peer group, and BGP session template configurations. To configure support for BGP NSR with SSO in BGP peer and BGP peer group configurations, use the **neighbor ha-mode sso** command in BGP configuration mode. To include support for Cisco BGP NSR with SSO in a peer session template, use the **ha-mode sso** command in session-template configuration mode.

You must perform several tasks on a PE router if you want to configure a BGP peer to support BGP NSR with SSO. In router configuration mode, enable BGP graceful restart capability and BGP Cisco NSF awareness using the **bgp graceful-restart** command. Configure the neighbor to support BGP NSR with SSO using the **neighbor ha-mode sso** command.

Before Cisco IOS XR Release 3.8.0, the NSR feature was not supported. In Cisco IOS XR Release 3.8.0 software, the NSR feature is configured in BGP configuration mode using the **nsr** command.

Summary

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

Summary

- BGP is designed to scale to huge internetworks (slower convergence).
- Lowering BGP timers speeds up convergence.
- BGP performance characteristics can be tested by using ping or IP SLA.
- Features to improve convergence:
 - **fast-external-fallover**
 - Fast peering deactivation
 - Cisco NSF awareness
 - BGP Support for Nonstop Routing with SSO

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0–3.20

Lesson Self-Check

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

- Q1) Which packets are sent to verify TCP connectivity? (Source: Testing BGP)
- A) hello
 - B) holdtime
 - C) keepalive
 - D) scan time
- Q2) What can you achieve by reducing default BGP timers? (Source: Testing BGP)
- A) slow down convergence and increase CPU usage
 - B) speed up convergence and increase CPU usage
 - C) speed up convergence and decrease CPU usage
 - D) slow down convergence and decrease CPU usage
- Q3) What are default values for advertisement interval timer? (Source: Testing BGP)
- A) 30 seconds for EBGp neighbors and 5 seconds for IBGP neighbors
 - B) 5 seconds for EBGp neighbors and 30 seconds for IBGP neighbors
 - C) 60 seconds for EBGp neighbors and 180 seconds for IBGP neighbors
 - D) 180 seconds for EBGp neighbors and 60 seconds for IBGP neighbors
- Q4) Which two features can speed up BGP convergence? (Choose two.) (Source: Testing BGP)
- A) Fast peering deactivation
 - B) BGP support for fast convergence
 - C) BGP support for NSR with SSO
 - D) Fast TCP session deactivation
- Q5) What was the most important parameter when designing BGP? (Source: Testing BGP)
- A) convergence
 - B) low latency
 - C) high throughput
 - D) scalability

Lesson Self-Check Answer Key

- Q1) C
- Q2) B
- Q3) A
- Q4) A, C
- Q5) D

Module Summary

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

Module Summary

- Monitoring routing protocol performance includes these issues:
 - Impact of protocols on the control plane and consequently on data plane
 - Impact of protocol configuration on network performance (link utilization)
 - Impact of protocols on convergence
- Changes can be proposed and implemented based on network audit:
 - Optimize convergence
 - Optimize resource utilization
- Care should be taken when tuning timers:
 - Too aggressive: high CPU utilization, unstable network
 - Too long: slow convergence

© 2010 Cisco Systems, Inc. All rights reserved.

MSRPv1.0-3-1

Each network device in the service provider network has finite hardware resources. Therefore, it is important to monitor CPU, memory, link utilization, and other manual audit parameters. When high CPU and memory utilization are caused by routing protocols, take the actions described in this module to optimize routing operations. For instance, when link utilization is high or unequal, change routing paths.

You can test routing protocol performance characteristics by measuring routing protocol convergence using a ping or IP Service Level Agreement (IP SLA). You can improve the convergence of routing protocols by adjusting hello and fast timers, by enabling Bidirectional Forwarding Detection (BFD), and by applying the other tools described in this module.

Fault Management for Routing Protocols

Overview

Typical failures in service provider networks stem from connectivity issues or device misconfigurations. After configuration changes or the introduction of new equipment in a network, some routing issues may arise. In order to detect and solve problems that might arise, you have to understand what can go wrong in service provider networks. You also have to know how to gather information about a problem and how to solve it. This module describes how to troubleshoot a problem by indicating the problem, diagnosing a cause, and implementing a solution.

Module Objectives

Upon completion of this module, you will be able to perform fault management for routing protocols using various tools such as **show** commands and monitoring tools. This ability includes being able to meet these objectives:

- Troubleshoot routing issues in the service-provider core network by interpreting the monitoring systems, and then propose solutions to fix the problems
- Troubleshoot routing issues at the service-provider edge network by interpreting the monitoring systems or customer-reported issues, and then propose solutions to fix the problems

Troubleshooting Routing Issues in the Network

Overview

After configuration changes or introducing new equipment in a network, some routing issues may arise due to misconfiguration. Failures can also arise because of connectivity issues between devices. In order to detect and solve problems that might arise, you have to understand what can go wrong in service provider networks. You also have to know how to gather information about a problem and how to solve it.

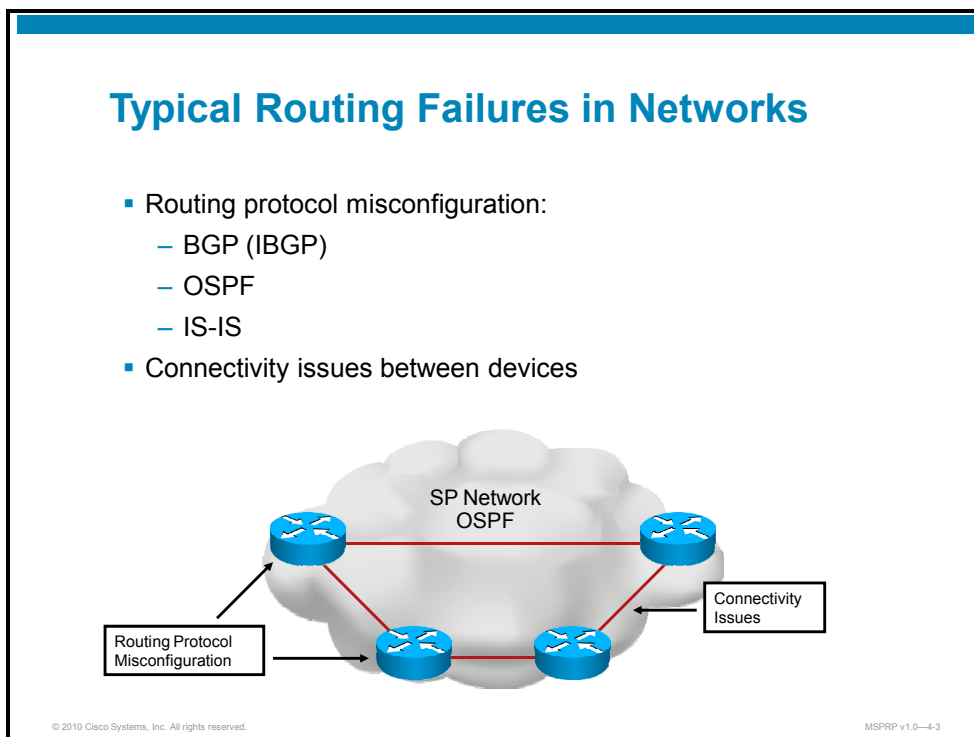
Objectives

Upon completing this lesson, you will be able to identify and solve routing issues in service provider networks. This ability includes being able to meet these objectives:

- Identify Border Gateway Protocol (BGP) routing failures
- Identify Open Shortest Path First (OSPF) and Intermediate System-to-Intermediate System (IS-IS) routing failures
- Use a structured approach to troubleshooting
- Gather and analyze information about routing failures

Typical Routing Failures in Service Provider Environments

This topic describes typical BGP, OSPF, and IS-IS misconfigurations and failures in service provider networks.



In service provider networks, typical routing failures stem from routing protocol misconfiguration or connectivity issues between devices. In service provider networks, BGP is used together with one of the internal gateway protocols (IGPs), such as OSPF or IS-IS. Inside service provider networks, Internal BGP (IBGP) is used to carry BGP information across the service provider network from one customer to another or to upstream service providers. When you install new equipment or make configuration changes, IBGP misconfigurations can occur that can prevent customer connectivity through the service provider network.

Typical BGP issues are BGP authentication mismatches, BGP peering IP address issues, and issues with the availability of the next hop of a BGP route.

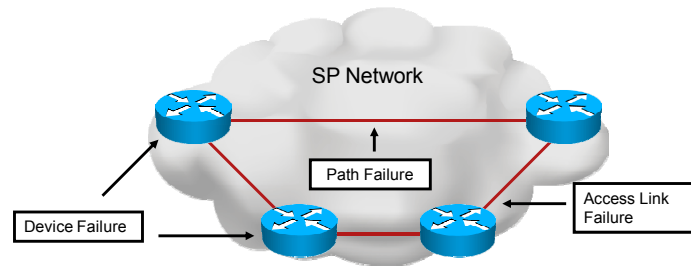
Since BGP operations depend heavily on IGP to resolve nondirectly connected next hops or nondirectly connected neighbors, IGP issues can manifest in BGP problems. Some typical misconfigurations when using OSPF are authentication mismatches, maximum transfer unit (MTU) size mismatches and timer mismatches.

When using IS-IS as the IGP, typical misconfigurations are authentication mismatches and maximum transfer unit (MTU) size mismatches.

In many cases, a problem may look like a BGP problem, but after troubleshooting, you may learn that the problem is actually an IGP or even connectivity issue. The reason is that IBGP relies on IGP to carry BGP routing information across the service provider network.

Connectivity Issues

- Connectivity issues between devices:
 - Device failure
 - Access link failure
 - Path failure



© 2010 Cisco Systems, Inc. All rights reserved.

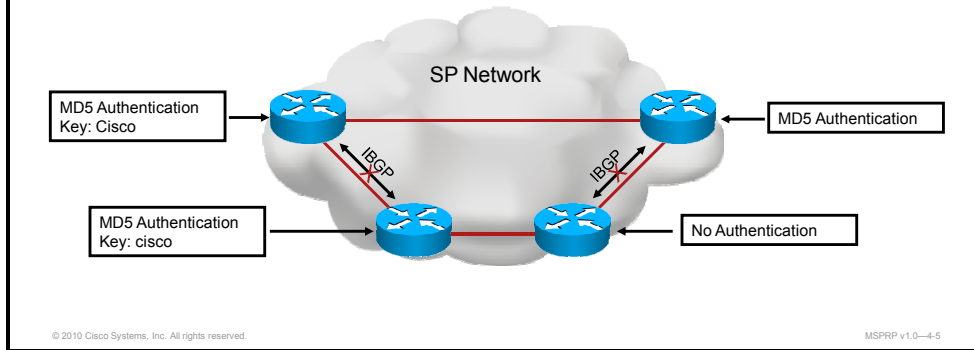
MSRP v1.0—4-4

Failures in service provider networks can also occur due to connectivity issues between devices. Network connectivity consists of links, devices, and paths across networks. A path interconnects two routers using a number of devices, such as switches and links. Any of these intermediate devices can fail and render a path unavailable.

When you eliminate connectivity issues, you should start troubleshoot routing failures that are caused by misconfigurations.

BGP Authentication Mismatch

- A BGP authentication mismatch results in a BGP session not being established:
 - Authentication configured on one side only
 - Authentication key mismatch

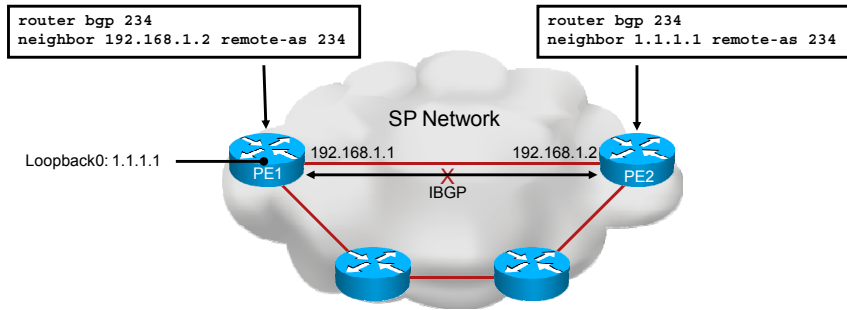


BGP supports MD5 authentication. When configuring authentication, you must make sure that you enable authentication on both BGP neighbors. If you enable BGP authentication only on one side or you misconfigure the password, the BGP session between the two neighbors will not be established. Routers will get stuck in the active state, which is not the desired state in BGP.

If you want to change the password between two BGP peers with an existing session, the local router will not tear down the existing session after you configure a new password. The local router will maintain the session using the new password until the BGP hold-down timer expires. The timer is 180 seconds by default. If the password is not changed or entered on the other router before the hold-down timer expires, the session times out.

BGP Neighbor Peering Issue

- By default, the source IP address of a BGP packet is defined by the outgoing interface.
- The source IP address of a BGP packet must match the address that the other router has in its neighbor statement.
- The source address of a BGP packet is compared to the list in the neighbor statement:
 - If a match is found, a relationship is established.
 - If no match is found, the packet is ignored.



When a router creates a BGP packet for a neighbor, it checks the routing table for the destination network to reach that network. The IP address of the outbound interface is used as the source IP address by default. The destination IP address is selected from the BGP neighbor statement.

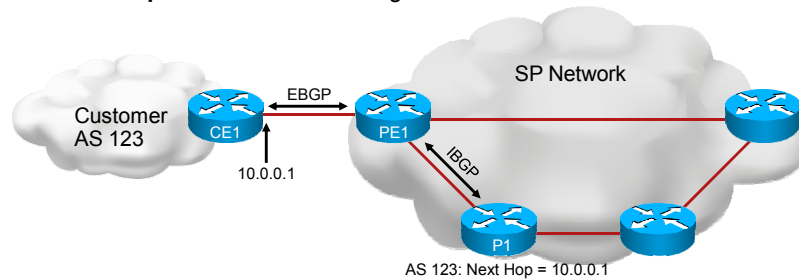
When the other router receives the BGP packet, it checks the source IP address. If the source IP address does not match the address that is defined in the corresponding neighbor statement, the routers will not establish a BGP session. They will remain in the active state, which is not the desired state.

In this example, when PE1 generates a BGP packet for the PE2 router, it uses 192.168.1.1 as the source IP and 192.168.1.2 as the destination IP address. The PE2 router compares the source IP address of the packet to neighbor statements. Since there is no match, the packet will be discarded and a BGP session will not be established.

To correct this problem, you should change the source of the BGP packets on the PE1 router using the **neighbor ip_address update-source loopback 0** command in the BGP router configuration mode.

BGP Next-Hop Behavior

- In BGP, the next hop does not mean the next router; it means the router in the next AS.
- For EBGp, the next hop is the address of the router which sent the update.
- For IBGP, the next hop advertised by BGP should be carried into IBGP.
- If an access link address space is not advertised into IGP, use the **next-hop-self** command on edge routers.



BGP is an external routing protocol that informs the next autonomous system (AS) about networks in other autonomous systems. For external BGP (EBGP), the next hop for a network is the IP address of the router that sent the update. Internal BGP (IBGP) operates differently. The next hop of a network that is learned by an IBGP neighbor is not the IP address of the router that sent the update. The next hop of a network that is learned by an IBGP neighbor is the IP address of the first router in the next AS on the path to the destination network.

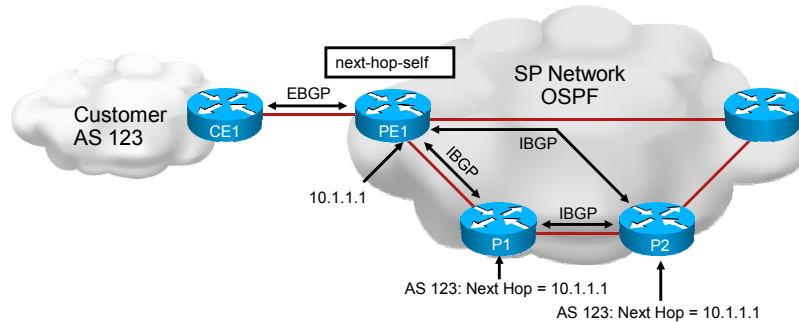
In the example, two autonomous systems are shown. Networks in AS 123 are advertised to the service provider autonomous system by the CE1 router through the EBGp session. Those networks are then further advertised to service provider routers through the IBGP session. BGP states that the next hop advertised by EBGp should be carried into IBGP. This is a reason why P1 router sees the network in AS 123 as available through the next hop of 10.0.0.1, which is the router in AS 123.

For cases in which IGP fails or does not advertise access links, service provider routers do not know how to reach the next hop of networks in other autonomous systems. If access links are not advertised into IGP, you must change the default next-hop behavior on PE routers using the **neighbor ip_address next-hop-self** command.

This is the first example for which it is shown that BGP relies on IGP.

BGP Next-Hop Behavior (Cont.)

- BGP heavily relies on IGP.
- The next hop of a network learned by BGP is not always directly connected and must be advertised using IGP.
- IGP failure results in an inaccessible next hop for a BGP route.



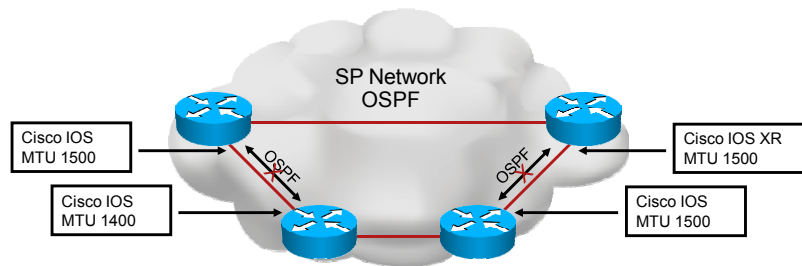
BGP states that an update learned from an IBGP neighbor should not be propagated to another IBGP neighbor, to prevent routing loops. For this reason, BGP requires full mesh IBGP between routers to advertise BGP routing updates across the service provider network.

The example shows a service provider network in which full mesh IBGP is established between the PE1, P1, and P2 routers. For P2 to reach networks in AS 123, it has to send packets to the PE1 router. But the PE1 router is not directly connected to the P2. The IP address of the PE1 router should be advertised to the P2 router using IGP. If there is a failure in IGP, the next hop on the P2 router will not be accessible. The router will not be able to route packets that are destined for networks in AS 123.

This is the second example that shows how BGP heavily relies on IGP. In many cases, a routing failure happens because of IGP failure and this failure is then manifested in BGP.

OSPF MTU Mismatch

- The MTU command on Cisco IOS XR Software should include the Layer 2 header size.
- The IP MTU inherits its value from the interface MTU.
- The OSPF DBD packet contains an MTU field.
- An IP MTU mismatch between OSPF neighbors results in neighbors being stuck in the exstart/exchange state.



OSPF uses different types of packets for its operations. One of the packets is the database description (DBD) packet, which checks for link-state database synchronization between routers. This packet includes the MTU size field, which is populated by the IP MTU value of the OSPF-enabled interface. If the IP MTU value is not set explicitly, the value will be inherited from the interface MTU. If the received value is different from the value that is configured on the other router, the DBD packet will be discarded and neighbors will get stuck in the exstart/exchange state.

If you try to establish OSPF between a Cisco IOS router and a Cisco IOS XR router and you do not change MTU values, OSPF adjacency will not be established. The reason is that when you define the interface MTU size in Cisco IOS XR Software, you should specify the maximum size of a packet including the Layer 2 header. The Layer 2 header is in the case of Ethernet is 14 bytes long. When you define the interface MTU size in Cisco IOS Software, you should specify the maximum size of a packet without the Layer 2 fields.

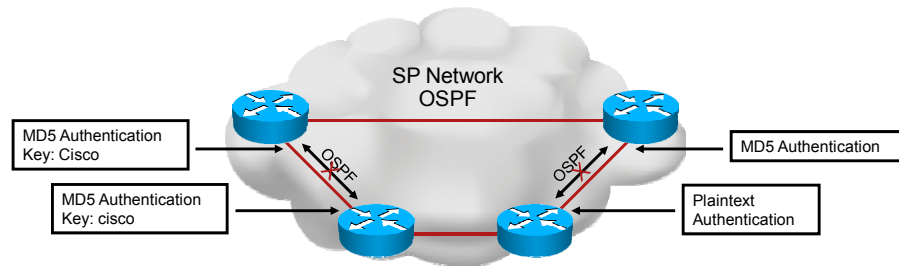
In the example, the interface MTU is set to 1500 bytes on the Cisco IOS router and Cisco IOS XR router. The IP MTU value will be inherited from the interface MTU value on both routers. However, in Cisco IOS XR Software, the interface MTU value size includes the Layer 2 header. That means that the inherited IP MTU will be 14 bytes lower, or 1486 bytes. On the Cisco IOS router, the inherited IP MTU value will be 1500 bytes. When you try to establish OSPF adjacency between the two routers, there will be a MTU mismatch and the routers will get stuck in the exstart/exchange state.

There are two solutions to this problem:

- **Correct the MTU values** on the interfaces.
- Configure **OSPF to ignore the MTU field in DBD packets** using the **ip ospf mtu-ignore** command in interface configuration mode on the Cisco IOS router. Use the **ip ospf mtu-ignore** command in router configuration mode under interface configuration mode in Cisco IOS XR Software.

OSPF Authentication Mismatch

- An authentication mismatch results in OSPF adjacency not being established:
 - Authentication type (plaintext or MD5) mismatch
 - Authentication key mismatch



© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-4-10

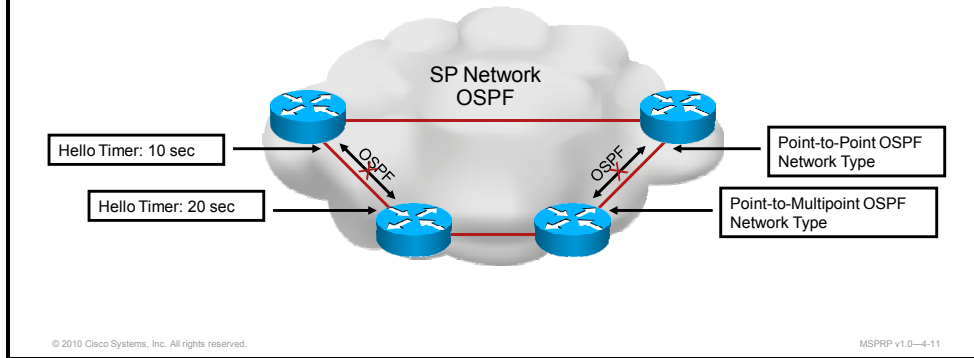
OSPF supports two types of authentication: MD5 and plaintext authentication. If you configure different types of authentication or you misconfigure keys that are used for authentication, routers will not establish OSPF adjacency.

If you use MD5 authentication, you can change the key used for authentication without OSPF adjacency being torn down. When you configure a new key for MD5 authentication, the local router will send multiple copies of OSPF packets, each authenticated by different keys. The local router will stop sending multiple copies of OSPF packets when it learns that the other router is also using the new key for authentication.

If you use plaintext authentication and would like to change the password that is used for authentication, you have to change the key on the other router before the hold-down timer expires. Otherwise, OSPF adjacency will be torn down.

OSPF Timer Mismatch

- Different OSPF network types result in different hello and dead timers.
- A hello timer and dead timer mismatch results in OSPF adjacency not being established.



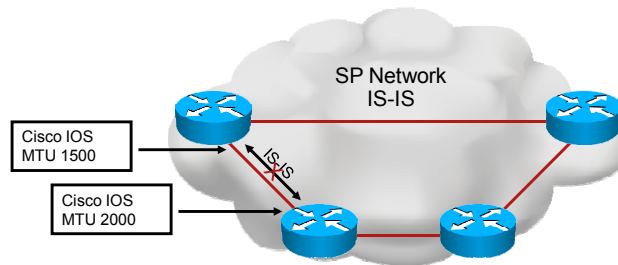
The hello timer specifies the frequency (in seconds) with which a router sends hello packets. The dead interval is the time (in seconds) that a router waits to hear from a neighbor before declaring the neighboring router out of service. The dead interval is four times the hello interval by default. These timers must match on both routers, or adjacency will not be established.

Different OSPF network types use different default values for hello and dead timers. Broadcast and point-to-point network types use a 10 second hello interval. Nonbroadcast, point-to-multipoint, and point-to-multipoint nonbroadcast OSPF network types use a 30-second hello interval. If you use different OSPF network types on connected routers, you must change the hello timer manually.

The figure shows two routers. The interface on one router is set to the point-to-point OSPF network type, and the interface on other router is set to the point-to-multipoint OSPF network type. The network type mismatch results in a hello timer mismatch, so OSPF adjacency will not be established. In this case, you have to manually set the hello timers to the same value or you have to change the OSPF network type to the same type on both routers.

IS-IS MTU Mismatch

- IS-IS hello packets are padded to the full MTU size.
- The router with the lower MTU will ignore IS-IS hello packets sent by the router with the higher MTU.
- An interface MTU mismatch between IS-IS neighbors causes the neighbor with the higher MTU to become stuck in the INIT state.



© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-4-12

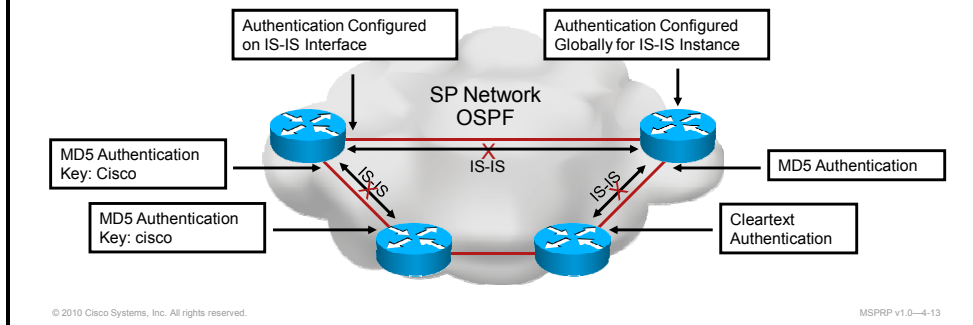
IS-IS hellos are, by default, padded to the complete interface MTU size. If there is an interface MTU mismatch on the two routers, the router with the higher interface MTU will pad IS-IS hello packets to the complete MTU size. This hello packet will be discarded on the other router's interface, which uses a smaller interface MTU. MTU mismatch will result in the router with the higher MTU on interface becoming stuck in the INIT state, while the router with the lower interface MTU will not know anything about the other IS-IS router.

To resolve this problem, you must make sure that both sides of the link use the same MTU.

IS-IS hello packet padding can be disabled using the **no isis hello padding** interface configuration command. A user would disable hello padding in order to avoid wasting network bandwidth if the MTU on both interfaces is the same. Although hello padding is disabled, routers still send the first five IS-IS hellos padded to the complete MTU size. Consecutive hellos are no longer padded. For this reason, you cannot disable IS-IS hello padding to resolve an MTU mismatch issue.

IS-IS Authentication Mismatch

- An authentication mismatch results in IS-IS adjacency not being established.
 - Authentication configuration type mismatch (IS-IS interface or IS-IS instance)
 - Authentication type (clear text or MD5) mismatch
 - Authentication key mismatch



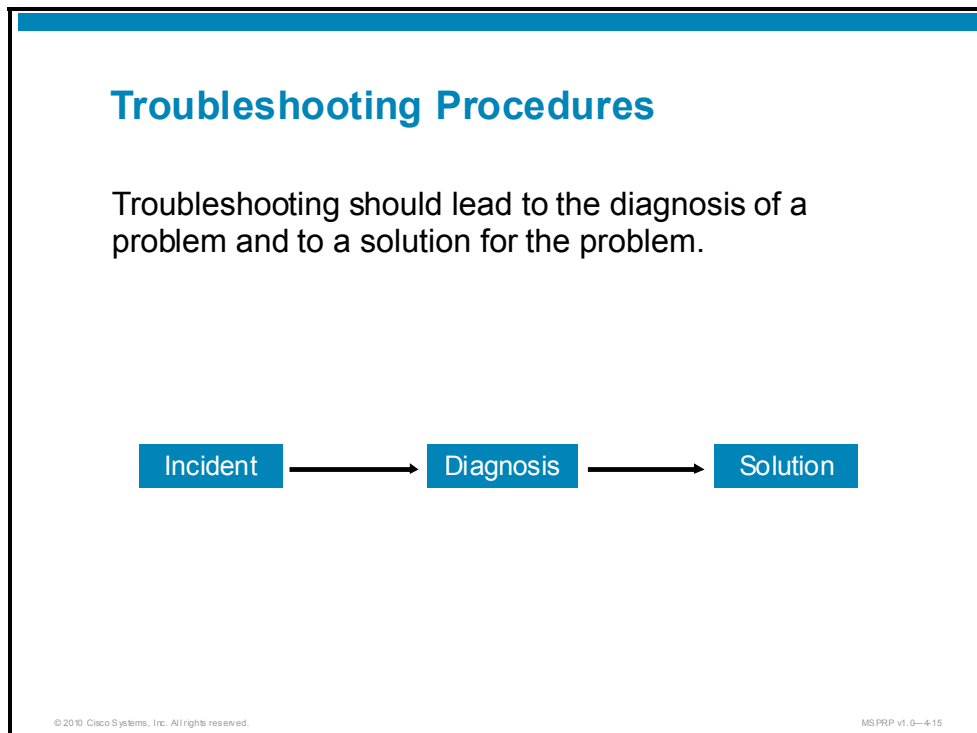
IS-IS supports MD5 and cleartext authentication. You have to make sure that you use the same type of authentication and the same keys on both sides of the link. Otherwise, routers will get stuck in the INIT state and IS-IS adjacency will not be established.

IS-IS authentication can be configured separately for each interface that is enabled for IS-IS or it can be configured globally for an IS-IS instance. However, authentication on an interface configures authentication of IS-IS hello packets, while authentication configured globally enables authentication of IS-IS link state packets (LSP). To successfully configure IS-IS hello packet authentication, you have to configure authentication on both routers on each IS-IS interface. To configure IS-IS authentication of LSP packets, you have to configure authentication on both routers globally for an IS-IS instance.

Authentication can be configured separately for level 1 or level 2, so you also have to be careful to enable authentication for the appropriate level.

Troubleshooting Techniques

This topic describes a structured approach that leads to diagnosing and solving problems.



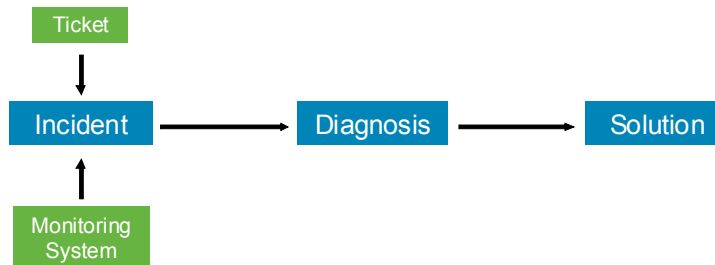
The troubleshooting process leads to the diagnosis of a problem and to a solution for the problem. Troubleshooting in general starts when somebody reports and defines a problem. Next, the diagnosis of the problem starts. During this process, information is gathered and analyzed, the problem definition is refined, and possible causes for the problem are proposed. Eventually, this process should lead to a diagnosis of the root cause of the problem.

When the root cause has been found, possible solutions need to be proposed and evaluated. After the best solution is selected, that solution should be implemented. In some cases, the solution cannot be implemented immediately and a workaround will have to be proposed until the actual solution can be implemented. The difference between a solution and a workaround is that a solution resolves the root cause of the problem, while a workaround only remedies or alleviates the symptoms of the problem.

Problem Detection

A problem can be:

- Detected by a monitoring system (SNMP, syslog messages)
- Reported by a customer



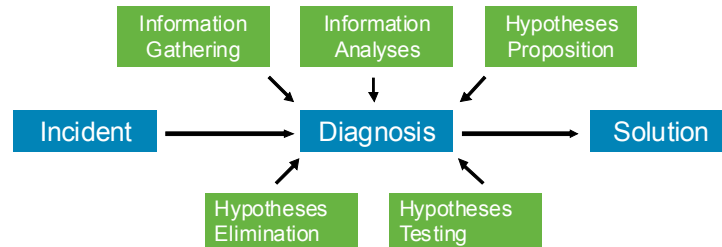
© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0—4-16

Troubleshooting starts with the reporting of a problem. Problems in service provider networks can be detected by a monitoring system that collects syslog messages or SNMP traps, or that periodically polls an SNMP agent on monitored devices. A customer can also detect a problem when it loses connectivity to other networks. The customer can report the problem by opening a ticket in a ticketing system. Regardless of how you learn about a problem, you, as network administrator, should immediately start to gather information and propose a possible solution.

Diagnostic Phases

Diagnosis should lead to the nature and cause of a problem.



© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-4-17

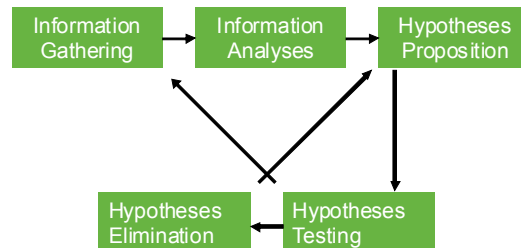
Diagnosis is the process of identifying the nature and cause of a problem. Essential phases of this process are as follows:

- **Information and symptoms gathering:** Gathering information about what is happening is essential to the troubleshooting process. Usually the problem reports do not contain enough information for you to formulate a good hypothesis without first gathering more information. You can gather information directly (by observing processes), or indirectly (by executing tests).
- **Information analyses:** Gathered information should be analyzed. You should compare the symptoms against your knowledge of the system, processes, and baselines to separate normal behavior from abnormal behavior.
- **Hypotheses proposition:** After analyzing information you will be left with one or more possible problem causes. The most likely cause should be proposed as the hypothetical cause of the problem.
- **Hypotheses testing:** By testing the hypothesis, you can confirm or deny that the hypothetical cause is the actual cause of the problem. The simplest way to do hypothesis testing is by proposing a solution based on this hypothesis, implementing that solution, and verifying if the solution has solved the problem. If this method is impossible or disruptive, the hypothesis can be strengthened or invalidated by gathering and analyzing more information. The hypothesis can be also tested in a test bed in a lab environment.
- **Hypotheses elimination:** by observing the results of hypothesis testing, you can eliminate some hypotheses.

Troubleshooting Methods

The following methods are guiding principles for how you should move through different troubleshooting phases:

- A troubleshooting method based on TCP/IP layers
- The “shoot from the hip” method (use previous experience)
- The “follow the path” method (trace the path of packets)
- The “spot the difference” method (compare with working devices)
- The “move the problem” method (swap components)



© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-4-18

In a structured troubleshooting process, you should follow diagnostic phases in specified order. However, when troubleshooting the process for complex problems, you will continually move between different phases. You will gather some information and analyze it, then formulate, test, and eliminate hypotheses. After that you will propose a new hypothesis, test it again, gather new information, and so on.

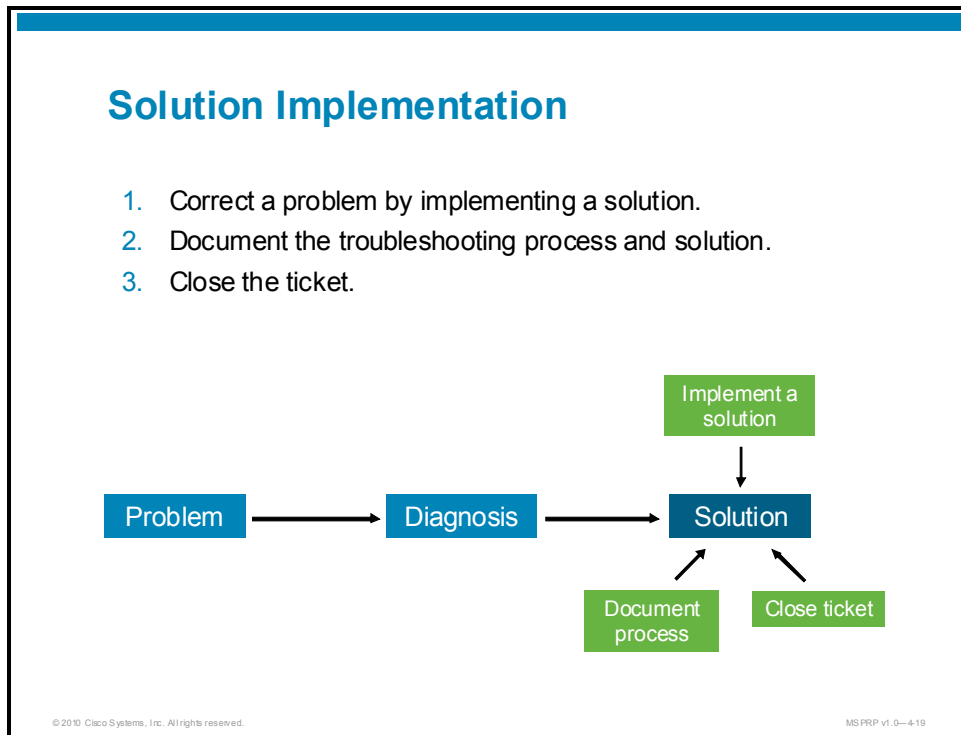
A structured approach leads to a solution faster than moving randomly between different phases. If there is no structure to your approach, it is also impossible to hand the troubleshooting over to someone else without losing all the progress that was made so far. This even applies to resuming your own troubleshooting process.

There are some proposed methods for how to move through different diagnosis phases:

- **“Shoot from the hip” method:** In this method, the troubleshooter, after a very short period of information gathering, quickly makes changes to see if doing so solves the problem. This method may seem like a random method of troubleshooting, but usually the guiding principles for this method are great knowledge of common symptoms and experiences with troubleshooting network problems.
- **Troubleshooting based on TCP/IP layers:** To start troubleshooting, you can follow the TCP/IP model layers and start the troubleshooting from the physical to the application layer. You can also start the troubleshooting process from the application to the physical layer. Another option is to start troubleshooting in the middle level (usually the Internet layer) and then go up or down based on the results.
- **“Follow the path” method:** In this method, you determine the path packets take from source to destination to eliminate irrelevant links and devices from the troubleshooting process.
- **“Spot the difference” method:** In this method, you compare a malfunctioning device or process to a functioning device and try to spot a difference. If you compare in this way, you may discover a solution or workaround without even understanding the root cause of the problem.

- **“Move the problem” method:** In this method, you swap a nonworking component with a working component and observe if the problem stays, moves, or disappears. This method isolates a problem to a single device. This method has a drawback, in that it cannot detect problems that stem from a particular combination of several devices.

In many practical cases, almost all of the previous methods can be combined in a single troubleshooting process.



After you determined the root cause of a problem, you have to implement a solution to eliminate the problem. After you make sure that the solution solved the problem and that the solution did not introduce new problems, you have to document the troubleshooting process for future use. If the problem was reported by a customer using the ticketing system, you also have to close the ticket.

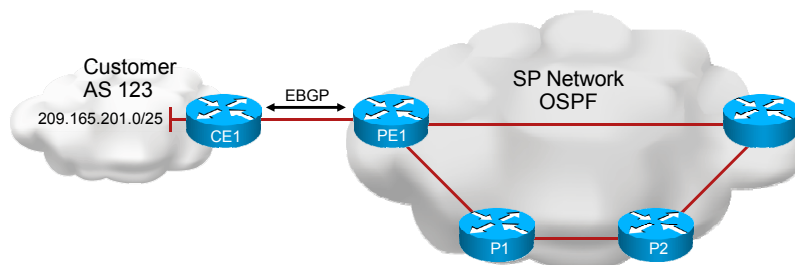
However, if you were not able to solve the problem, you will have to escalate the problem to other persons.

Troubleshooting Typical Failures

This topic describes procedures for troubleshooting typical routing failures. Based on different monitoring outputs, various troubleshooting procedures are proposed.

Troubleshooting Routing Failures

- You receive a ticket from a customer indicating that its network cannot communicate with other networks.
- Start gathering information:
 - Know the topology, routing, and addressing
 - Use various **show** and **debug** commands



© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0—4-21

In this example, a customer opens a ticket indicating that it cannot communicate with other networks and the Internet. You, as network administrator, should troubleshoot the problem and provide a solution.

First you have to start gathering information. The first thing that you should be familiar with is the network topology, the routing protocols that are used in the network and in addressing. If you are not familiar with these things, troubleshooting is practically impossible. Next, you should start gathering additional information about the problem using various **show** and **debug** commands on Cisco devices.

In the figure, in this example, the customer exchanges routing information with the service provider using BGP. IBGP is used inside the service provider network to advertise the customer network across the service provider network to the Internet or upstream SPs. The service provider uses OSPF as an internal routing protocol.

Gather and Analyze Information

- Use ping and traceroute to determine the point of failure.

```
CE1#ping 10.0.0.2
Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 10.0.0.2, timeout is 2 seconds:
.....
Success rate is 0 percent (0/5)
```

```
CE1#traceroute 10.0.0.2
Tracing the route to 10.0.0.2
 0 172.16.0.2 8 msec 64 msec 64 msec
 1 192.168.1.2 60 msec 128 msec 44 msec
 2 * * *
```

- Check for any link, path, or device failures and correct any you find.
- Verify the routing table for customer BGP routes.

```
P2#show ip route bgp
```

- There are no customer routes on SP routers—returning traffic cannot reach the customer.

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-422

The first thing that you would usually do is verify connectivity using the ping tool from the customer side to the service provider network. When you confirm that there is a problem with connectivity, you should use the traceroute tool to determine the point of failure. Then you should check for any link, path, or device failures. If there are any, you should correct them and verify whether doing so solved the problem.

If there were no connectivity issues, you should start troubleshooting routing protocols. If BGP is used to advertise customer networks, you should start troubleshooting BGP. First you should verify whether there are any customer routes in routing tables on service provider routers. You can verify these tables using the **show ip route bgp** command on a Cisco IOS router or **show route bgp** command on a Cisco IOS XR router.

In this example, you can see that there are no BGP routes present on the service provider router. Obviously there is a problem with BGP inside the service provider network. The customer can reach other networks. The problem is with returning traffic, because there are no customer routes present on service provider routers.

Gather and Analyze Information (Cont.)

Example 1

- Verify the BGP table for the customer network.

```
P2#show ip bgp
BGP table version is 1, local router ID is 10.0.0.2
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete
```

Network	Next Hop	Metric	LocPrf	Weight	Path
* i209.165.201.0/25	192.168.1.1	0	100	0	123 i

- The customer route is learned by BGP, but the route is not put into the routing table.

```
P2#show ip bgp 209.165.201.0
BGP routing table entry for 209.165.201.0/25, version 0
Paths: (1 available, no best path)
Not advertised to any peer
123
192.168.1.1 (inaccessible) from 10.0.0.1 (192.168.1.2)
Origin IGP, metric 0, localpref 100, valid, internal
Originator: 192.168.1.1, Cluster list: 192.168.1.2
```

- The next hop of the route is not accessible. The problem is very likely the OSPF.

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-423

The next step is to verify if there are any routes in the BGP table using the **show ip bgp** command in Cisco IOS Software. To verify if there are any routes in the BGP table on a Cisco IOS XR router, use the **show bgp** command. In the example, you can see that the customer network (209.165.201.0/25) is present in the BGP table on the service provider router. But that route is not selected and has not been put into the routing table.

Note Recall that the selected BGP route is marked with the right angle bracket (>) character.

To see details about the customer route, use the **show ip bgp prefix** command in Cisco IOS Software or the **show bgp prefix** on a Cisco IOS XR router. In the printout, you can see that the route is valid, but the next hop (192.168.1.1) is not accessible. The unavailability of the next hop is a reason that the BGP route has not been put into the routing table.

The unavailability of a next hop is very likely an OSPF problem, and the next step is to troubleshoot the OSPF protocol.

Gather and Analyze Information (Cont.)

Example 1 (Cont.)

- Verify the next hop of the BGP route.

```
P2#show ip route
<...output omitted...>
Gateway of last resort is not set

C    192.168.5.0/24 is directly connected, GigabitEthernet2/0
    10.0.0.0/8 is variably subnetted, 2 subnets, 2 masks
O    10.0.0.10/32 [110/2] via 192.168.5.1, 00:00:39, GigabitEthernet2/0
C    10.0.0.0/24 is directly connected, GigabitEthernet1/0
```

- The next hop is not present in the routing table—troubleshoot OSPF.

```
P2#show ip ospf neighbor

Neighbor ID    Pri   State           Dead Time   Address        Interface
192.168.5.1    1     FULL/DR         00:00:33    192.168.5.1    GigabitEthernet2/0
192.168.1.2    1     EXCHANGE/DR     00:00:38    10.0.0.1       GigabitEthernet1/0
```

- The OSPF neighbor is in the exchange state—the next hop of the BGP route cannot be learned by OSPF.

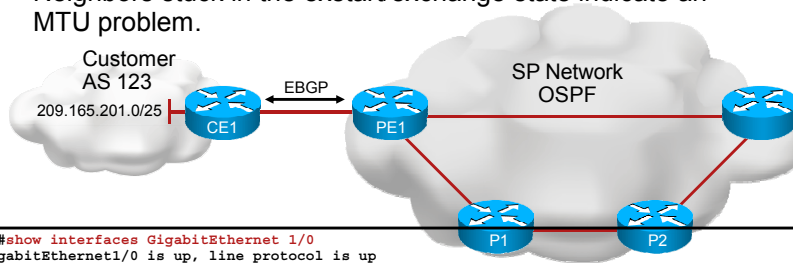
To troubleshoot the unavailability of the next hop of a BGP route, first verify the routing table on the service provider router using the **show ip route** command. You can see in the example that the network, including the BGP next hop (192.168.1.1), is not present in the routing table.

The next step is to verify the OSPF neighbors using the **show ip ospf neighbors** command in Cisco IOS Software or **show ospf neighbor** in Cisco IOS XR Software. In this way, you can see whether adjacency has been established with the router, which should advertise the next-hop network. In the printout, you can see that the OSPF neighbor 192.168.1.2 is in the exchange state, which means that networks cannot be learned from that neighbor. You should further troubleshoot OSPF to find a reason why the OSPF neighbor is stuck in the exchange state.

Propose and Test Hypothesis

Example 1 (Cont.)

- Neighbors stuck in the exstart/exchange state indicate an MTU problem.



```
P2#show interfaces GigabitEthernet 1/0
GigabitEthernet1/0 is up, line protocol is up
Hardware is 82543, address is ca02.19cc.001c (bia ca02.19cc.001c)
Internet address is 10.0.0.2/24
MTU 1700 bytes, BW 1000000 Kbit, DLY 10 usec,
reliability 255/255, txload 1/255, rxload 1/255
<...output omitted...>
```

```
P1#show interfaces GigabitEthernet 1/0
GigabitEthernet1/0 is up, line protocol is up
Hardware is 82543, address is ca02.19cc.001c (bia ca02.19cc.001c)
Internet address is 10.0.0.2/24
MTU 1500 bytes, BW 1000000 Kbit, DLY 10 usec,
reliability 255/255, txload 1/255, rxload 1/255
<...output omitted...>
```

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-4-25

After you gather all the information, you should analyze it and propose a hypothesis. One of the reasons an OSPF neighbor might be stuck in the exstart state is an interface MTU problem. To verify the interface MTU size, use the **show interfaces interface** command. You can see in the example that the P1 router and the P2 router use different MTU sizes on the connected link. Different MTU sizes on the two routers results in OSPF being stuck in the exstart state.

Note Recall that an OSPF DBD packet includes the MTU size. If the MTU size does not match the MTU size on other side of a link, the DBD packet will be discarded and neighbors will get stuck in the exstart/exchange state.

To test the hypothesis, change the MTU size on both interfaces to the same value or configure OSPF to ignore the MTU value in DBD packets. If this change solved the problem, you are done with troubleshooting OSPF. When you see the OSPF neighbor in the FULL state, verify the routing table again for the next-hop network. If this network is present in the routing table, verify if the customer route learned by BGP is now available in the routing table.

Gather and Analyze Information

Example 2

- Verify the next hop of the BGP route.

```
P2#show ip route
<...output omitted...>
Gateway of last resort is not set

C    192.168.5.0/24 is directly connected, GigabitEthernet2/0
    10.0.0.0/8 is variably subnetted, 2 subnets, 2 masks
O    10.0.0.10/32 [110/2] via 192.168.5.1, 00:00:39, GigabitEthernet2/0
C    10.0.0.0/24 is directly connected, GigabitEthernet1/0
```

- The next hop is not present in the routing table—troubleshoot OSPF.

```
P2#show ip ospf neighbor

Neighbor ID    Pri   State           Dead Time   Address        Interface
192.168.5.1    1     FULL/DR         00:00:33    192.168.5.1    GigabitEthernet2/0
```

- There is one OSPF neighbor missing—the next hop of the BGP route cannot be learned by OSPF.

© 2010 Cisco Systems, Inc. All rights reserved.

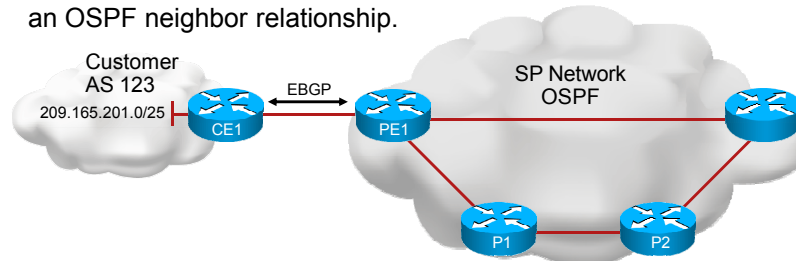
MSRP v1.0-426

Another possibility when troubleshooting OSPF is that the OSPF neighbor will not be present in the OSPF neighbor table at all. Again, the next hop that is used by a BGP route cannot be learned. As you can see in the example, OSPF neighbor 192.168.1.2 is not present in the OSPF neighbor table and the service provider router cannot learn the BGP next hop.

Propose and Test Hypothesis

Example 2 (Cont.)

- Authentication or timer mismatch results in failure to establish an OSPF neighbor relationship.



```
P1#show ip ospf interface gigabitEthernet 1/0 | include Hello|auth|Network
Process ID 1, Router ID 192.168.1.2, Network Type BROADCAST, Cost: 1
Timer intervals configured, Hello 10, Dead 40, Wait 40, Retransmit 5
Hello due in 00:00:01
Simple password authentication enabled
```

```
P2#show ip ospf interface gigabitEthernet 1/0 | include Hello|auth|Network
Process ID 1, Router ID 10.0.0.2, Network Type BROADCAST, Cost: 1
Timer intervals configured, Hello 5, Dead 20, Wait 20, Retransmit 5
Hello due in 00:00:01
Message digest authentication enabled
```

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0—4-27

After you gather all the information, you have to analyze it and propose a hypothesis. In this case, an OSPF adjacency is not being established. One reason for OSPF adjacency not being established might be an authentication mismatch or OSPF timer mismatch.

Note For an OSPF adjacency to be established, authentication and timers should match on both sides of a link.

Another reason could be different types of OSPF networks on both sides of a link, because different OSPF network types use different hello and dead timers. To verify OSPF authentication, timers, and network type, use the **show ip ospf interface interface** command on a Cisco IOS router and **show ospf interface interface** command on a Cisco IOS XR router. Optionally, you can provide the | (pipe) character, followed by the **include** statement to show only relevant information.

In the figure, you can see that the GigabitEthernet1/0 interface is on both routers, configured as the broadcast OSPF network type. You can also see that there is an OSPF hello and dead timer mismatch and authentication-type mismatch. On the P1 router, the hello timer is set to 10 seconds, the dead timer is set to 40 seconds, and the configured authentication type is plaintext authentication. On the P2 router, the hello timer is set to 5 seconds, the dead timer is set to 20 seconds, and the configured authentication type is MD5 authentication. The reason that the OSPF adjacency is not being established is an authentication-type mismatch and timer mismatch.

To test the hypothesis, correct the authentication type and timers on both interfaces. If making these corrections solved the problem, you are done with troubleshooting OSPF. When you see the OSPF neighbor in the FULL state, verify the routing table again for the next-hop network. If this network is present in the routing table, verify if the customer route learned by BGP is now available in the routing table.

Gather and Analyze Information

Example 3

- Verify the BGP table for the customer network.

```
P2#show ip bgp
```

- No routes are learned by BGP—troubleshoot BGP.

```
R2#show ip bgp summary
BGP router identifier 192.168.5.2, local AS number 234
BGP table version is 1, main routing table version 1

Neighbor      V    AS  MsgRcvd  MsgSent   TblVer   InQ  OutQ  Up/Down  State/PfxRcd
10.0.0.1      4    234     342     339       0     0     0 00:04:03  Active
```

- The BGP neighbor is in the active state—the established state is indicated by the number in the State/PfxRcd column.

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-428

In the previous two examples, the problems were with OSPF, but manifested as BGP problems, because of the unavailability of the next hop. Routing problems can also stem from BGP misconfiguration. In the third example, the customer network is again not present in the routing table. When you verify the BGP table, you see that no routes are present, which indicates a BGP problem. To further troubleshoot BGP, you should verify if the BGP neighbor is in an established state. You can verify the neighbor state with the **show ip bgp summary** command on Cisco IOS routers or the **show bgp summary** command on Cisco IOS XR routers.

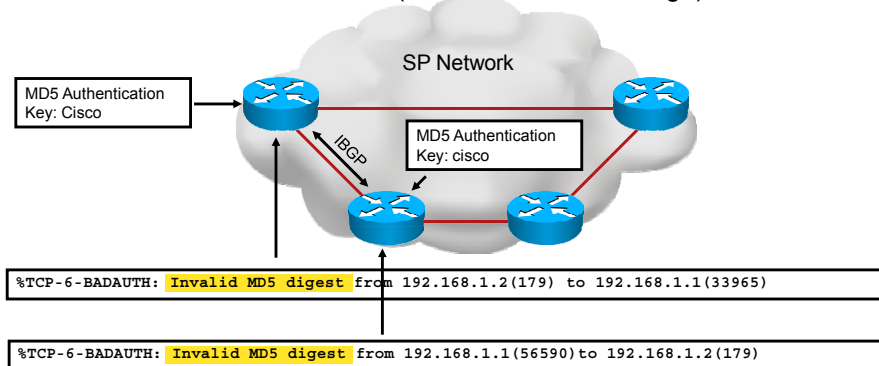
The sample printout indicates that the BGP neighbor is in the active state, which is not the desired state. The desired state is the established state, which is indicated by a number under the State/PfxRcd column. This number indicates the number of prefixes that are received from the BGP neighbor.

Propose and Test Hypothesis

Example 3 (Cont.)

BGP neighbor in active state indicates:

- Router does not have a route to the IP address specified by the **neighbor** command (troubleshoot IGP problem)
- Authentication mismatch (routers issue a message)



© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-4-29

The reason the BGP neighbor is in the active state could be that the router does not have a route to the IP address specified by the **neighbor** command. Again, this problem is an IGP problem and you should troubleshoot it as described in the first two examples.

Another reason for a BGP neighbor to be in the active state could be an authentication mismatch. In the example, MD5 authentication is configured on both routers, but the authentication key is mismatched. Routers will issue the following console message when trying to establish a BGP session:

```
%TCP-6-BADAUTH: Invalid MD5 digest from 192.168.1.2(179) to
192.168.1.1(33965)
```

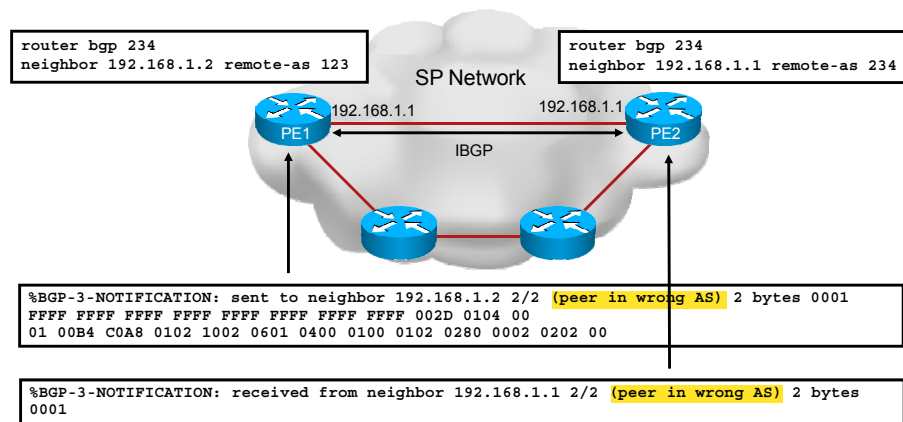
If one router is configured for authentication, but the other is not, a message such as the following will appear on first routers console:

```
%TCP-6-BADAUTH: No MD5 digest from 192.168.1.2(179) to
192.168.1.1(33965)
```

Propose and Test Hypothesis

Example 3 (Cont.)

- AS number misconfiguration (routers issue a message)



Another misconfiguration that can cause a BGP neighbor to be in the active state is an AS number misconfiguration. Routers will issue a console message in this case, indicating that one of the peers is configured with the wrong AS number.

At the router with the wrong AS number in the neighbor statement, you will see the following message:

```
%BGP-3-NOTIFICATION: sent to neighbor 192.168.1.2 2/2 (peer in  
wrong AS) 2 bytes 0001 FFFF FFFF FFFF FFFF FFFF FFFF FFFF  
002D 0104 0001 00B4 C0A8 0102 1002 0601 0400 0100 0102 0280  
0002 0202 00
```

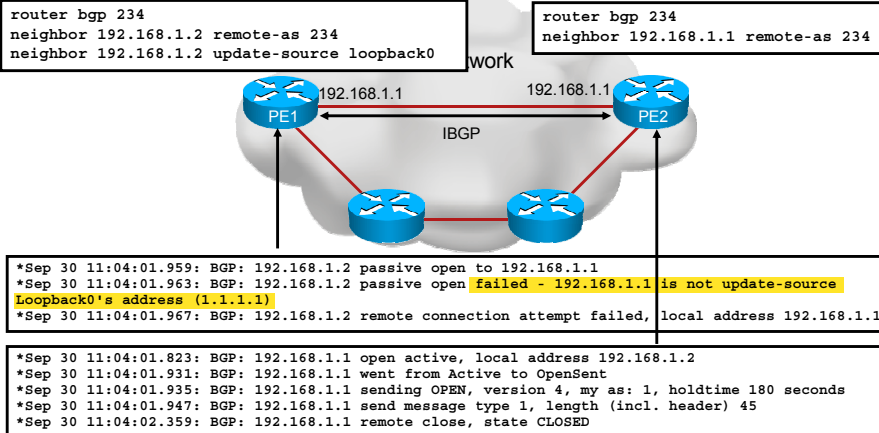
At the other router, you will see the following message:

```
%BGP-3-NOTIFICATION: received from neighbor 192.168.1.1 2/2  
(peer in wrong AS) 2 bytes 0001
```

Propose and Test Hypothesis

Example 3 (Cont.)

- BGP neighbors are peering with the wrong addresses
- Use the **debug ip bgp** command



Another reason BGP neighbors might be in the active state is if they peer with the wrong addresses.

Note

When a local router creates a BGP packet, the outgoing interface IP address is used as the source address by default. When the BGP packet reaches the remote router, the source IP address of the packet should match the IP address specified in the neighbor statement. Otherwise, the packet is discarded.

In this example, routers are peering with the wrong addresses. On the PE1 router, it is specified that the loopback's IP address should be used as the source address of BGP packets that are sent to the PE2 router. On the PE2 router, the wrong IP address is specified in the neighbor statement.

To discover this problem, use the **debug ip bgp** command.

For BGP neighbors to be in the established state, you should verify all of the mentioned problems and solve them. Next you should verify whether you receive any BGP routes from the other router.

Troubleshooting Other Routing Failures

- Log adjacency changes for BGP, OSPF and IS-IS
- Use other **show** and **debug** commands:
 - **show running-config**
 - **show ip route**
 - **show ip bgp**
- OSPF:
 - **show ip ospf neighbor**
 - **debug ip ospf adj**
- IS-IS:
 - **show isis neighbors**
 - **debug isis adj-packets**

© 2010 Cisco Systems, Inc. All rights reserved.

MS PRP v1.0-4-32

The recommended practice is to configure logging of adjacency changes for all routing protocols that are used. In this way, you will be immediately notified when there are changes in routing protocol adjacencies.

To troubleshoot other routing failures, you should use other **show** and **debug** commands. To verify OSPF adjacencies on a Cisco IOS router, use the **show ip ospf neighbor** command. To display information on related adjacency events on a Cisco IOS router, use the **debug ip ospf neighbor** command.

To verify OSPF adjacencies on a Cisco IOS XR router, use the **show ospf neighbor** command. To display information on related adjacency events on a Cisco IOS XR router, use the **debug ospf process_name adj** command.

To verify IS-IS adjacencies on a Cisco IOS router, use the **show isis neighbors** command. To display information on all adjacency-related activity on a Cisco IOS router, use the **debug isis adj-packets** command.

To verify IS-IS adjacencies on a Cisco IOS XR router, use the **show isis neighbors** command. To display information on all adjacency-related activity on a Cisco IOS XR router, use the **debug isis adj-packets** command.

Summary

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

Summary

- Typical failures in service provider networks stem from connectivity issues or device misconfigurations.
- Typical BGP misconfigurations are authentication mismatches, wrong peering IP addresses, and the unavailability of a next hop or peering IP address because of an IGP failure.
- Typical IGP misconfigurations are authentication mismatches, MTU mismatches, and timer mismatches.
- Troubleshoot problems by indicating the problem, diagnosing a cause, and implementing a solution.
- To diagnose the cause of a problem, gather and analyze information and propose and test hypotheses.
- Use various **show** and **debug** commands to gather and analyze information.

© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-433

Lesson Self-Check

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

- Q1) What are two BGP misconfigurations that happen in service provider networks when you introduce new devices? (Choose two.) (Source: Typical Routing Failures in Service Providers Environments)
- A) Different authentication types are configured on both sides of a link.
 - B) BGP neighbors peer with the wrong IP addresses.
 - C) Authentication is configured on one side only.
 - D) Hello timers are mismatched.
- Q2) Why does the interface MTU cause a problem when using the OSPF routing protocol? (Source: Typical Routing Failures in Service Providers Environments)
- A) The IP MTU size is sent to other routers in OSPF hello packets.
 - B) The interface MTU is sent to other routers in OSPF DBD packets.
 - C) The IP MTU is sent to other routers in OSPF DBD packets.
 - D) The OSPF hello packet is padded to the complete MTU interface size and the packet might be discarded.
- Q3) For which two reasons does BGP rely on the internal routing protocol? (Choose two.) (Source: Typical Routing Failures in Service Providers Environments)
- A) BGP uses IGP to resolve the path to the directly connected next hop of a BGP route.
 - B) BGP uses IGP to resolve the path to the nondirectly connected BGP neighbor that is specified in the neighbor statement.
 - C) BGP uses IGP to resolve the path to the nondirectly connected next hop of a BGP route.
 - D) BGP uses IGP to resolve the path to the directly connected BGP neighbor that is specified in the neighbor statement.
- Q4) What could be a reason for OSPF neighbors to get stuck in the exstart/exchange state? (Source: Typical Routing Failures in Service Providers Environments)
- A) authentication type mismatch
 - B) different OSPF network types on both ends of a link
 - C) IP MTU mismatch on both ends of a link
 - D) area number mismatch on both routers
- Q5) How can you solve the MTU mismatch problem when using the IS-IS routing protocol? (Source: Typical Routing Failures in Service Providers Environments)
- A) disable IS-IS hello packets padding
 - B) make sure both ends of the link use the same interface MTU
 - C) configure IS-IS to ignore the MTU size in IS-IS hello packets

Lesson Self-Check Answer Key

- Q1) B, C
- Q2) C
- Q3) B, C
- Q4) C
- Q5) B

Troubleshooting External Routing Issues

Overview

After configuration changes or introducing new equipment in a network, some routing issues may arise due to misconfiguration. Failures can also arise because of connectivity issues between devices. In order to detect and solve problems that might arise, you have to understand what can go wrong in networks of service providers. You also have to know how to gather information about a problem and how to solve it.

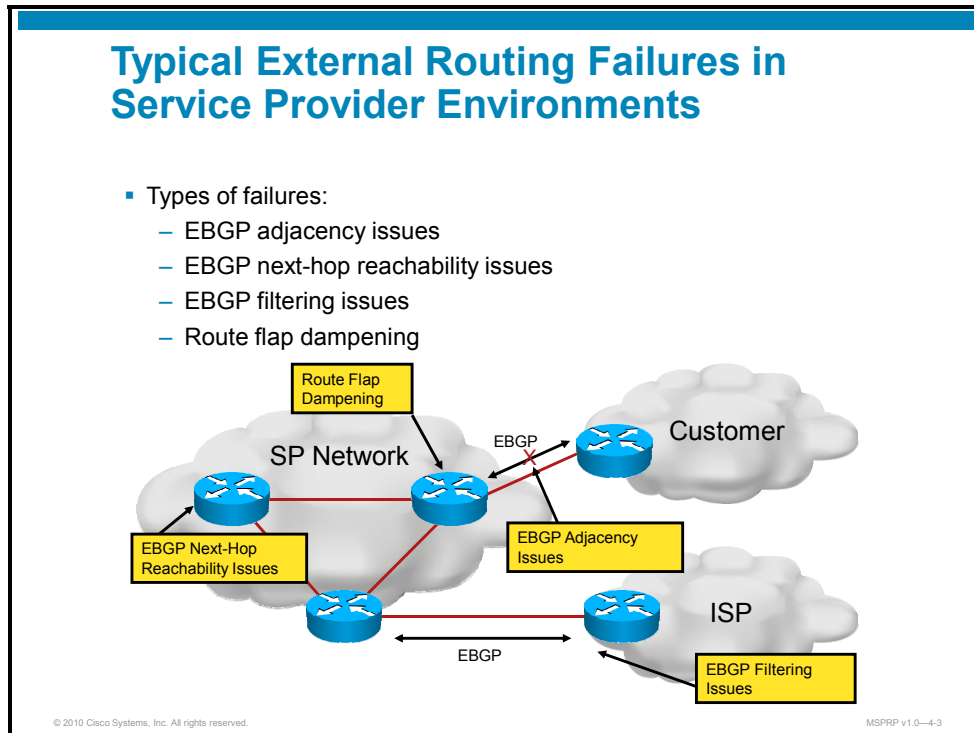
Objectives

Upon completing this lesson, you will be able to troubleshoot External Border Gateway Protocol (EBGP) routing issues at the service provider edge network, interpret the monitoring systems or customer-reported issues, and propose solutions to fix the problems. This ability includes being able to meet these objectives:

- Describe the typical external routing failures in service provider environments
- Use monitoring and troubleshooting commands to troubleshoot routing issues

Typical External Routing Failures in Service Provider Environments

This topic describes the typical external routing failures in service provider environments.



In service provider networks, typical routing failures stem from routing protocol misconfiguration or connectivity issues between devices. In these networks, EBGP is used to carry Border Gateway Protocol (BGP) information from the customers to other customers and other upstream service providers. When you or your customer install new equipment or make configuration changes, EBGP misconfigurations can occur that can prevent customer connectivity through the service provider network.

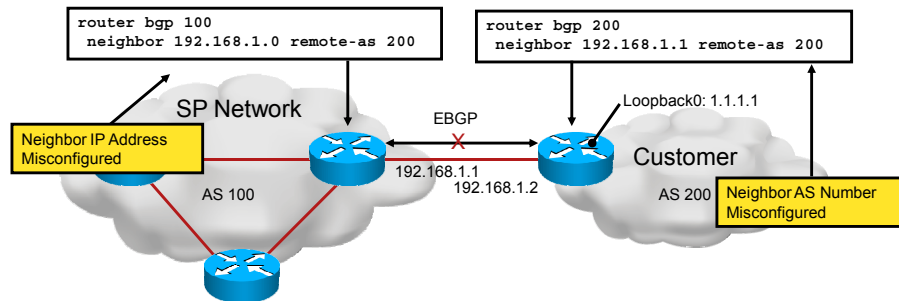
Typical BGP misconfigurations are BGP authentication mismatches, BGP neighbor IP address issues, and issues with the availability of a next hop of BGP route.

You may face these types of failures with EBGP:

- EBGP adjacency issues
- EBGP next-hop reachability issues
- EBGP filtering issues
- Route flap dampening issues

EBGP Adjacency Issues

- The wrong neighbor AS number or IP address causes the BGP session to not be established.



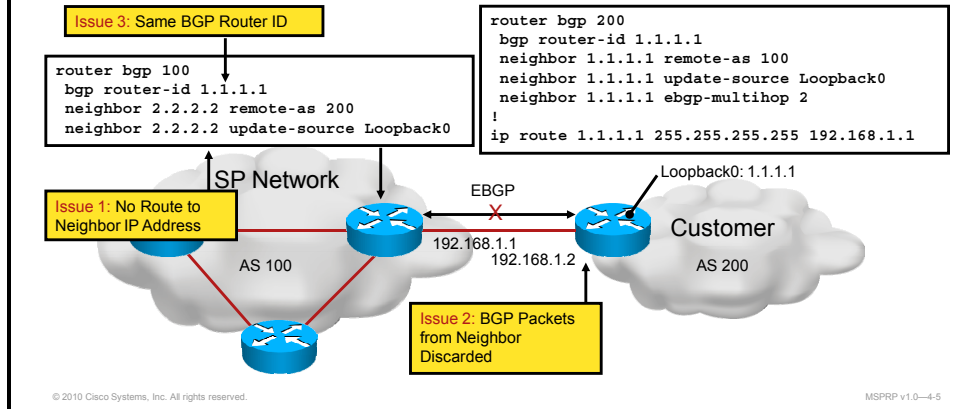
When routers are establishing an EBGP session, the neighbor IP address in the neighbor configuration should be the active IP address of the neighbor.

The example shows that the IP address in the service provider router is not correct and should be 192.168.1.2.

The neighbor autonomous system (AS) number should be correct. The example shows that the AS number in the neighbor configuration on the customer router is wrong. It should be 100, which is the AS number of the service provider.

EBGP Adjacency Issues (Cont.)

- Remote network (loopback) used for the EBGP session:
 - No route to neighbor loopback IP address
 - BGP packets from neighbor discarded because the TTL has been exceeded
- Same BGP router ID on EBGP neighbors



When a EBGP session is not enabled between adjacent IP addresses, an explicit route should be configured. An External BGP session will not be established if a router has no route to the neighbor IP address. Because there is no internal routing protocol between two adjacent autonomous systems, routers are not aware of remote IP addresses.

In the example, the service provider router has no route to the customer router loopback. To solve this issue, add a static route on the service provider router, similar to the configuration on the customer router.

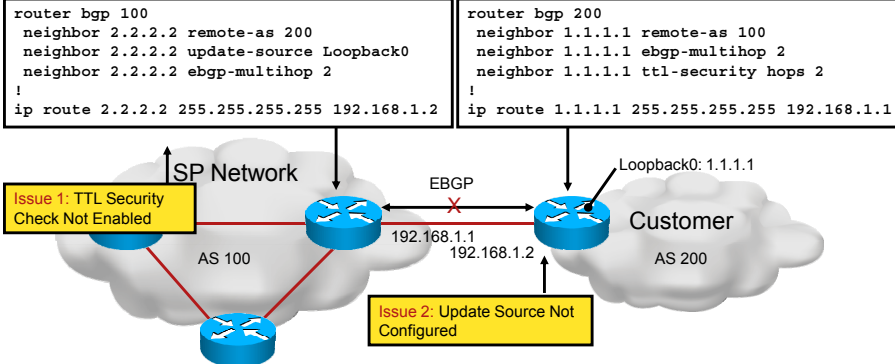
In order to successfully establish a BGP session between loopbacks of directly connected neighbors, you need to specify that the sessions is established across at least two hops (a loopback on the remote peer appears to be more than one hop away).

In the example, the service provider should set the **ebgp-multihop** value to 2.

If, by mistake, both administrators at the customer and service provider sites configure the BGP router ID to the same IP address, a session will not be established. The BGP router ID should be unique. If possible, use the active IP address on the router as the BGP router ID.

EBGP Adjacency Issues (Cont.)

- A TTL security check should be enabled on both routers.
- When loopback addresses are being employed, use the **update source** command.



BGP Support for the TTL Security Check feature introduces a lightweight security mechanism to protect EBGP peering sessions from CPU utilization-based attacks using forged IP packets. By enabling this feature, you can prevent attempts to hijack the EBGP peering session by hosts on network segments that are not part of either BGP network or between the EBGP peers.

You enable this feature by configuring a minimum TTL value for incoming IP packets that are received from a specific EBGP peer. When this feature is enabled, BGP will establish and maintain the session only if the TTL value in the IP packet header is equal to or greater than the TTL value configured for the peering session. If the value is less than the configured value, the packet is silently discarded and no ICMP message is generated.

For an EBGP session to be established, both routers should have equal BGP TTL security parameters configured. In the example, the **neighbor ip-address ttl-security hops hops** Cisco IOS command is missing on the service provider router.

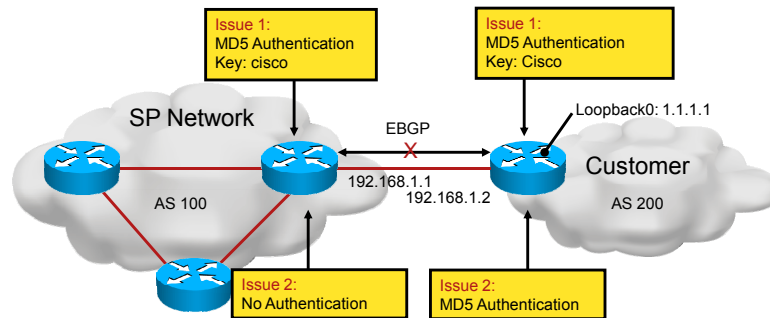
When a router creates a BGP packet for a neighbor, it checks the routing table for the destination network to reach that network. The IP address of the outbound interface is used as the source IP address, by default. The destination IP address is selected from the BGP neighbor statement.

When the other router receives a BGP packet, it checks the source IP address. If the source IP address does not match the address defined in the corresponding neighbor statement, the routers will not establish BGP session. They will remain in an active state, which is not a desired state.

To correct this problem, the customer should change the source of the BGP packets on its router.

EBGP Adjacency Issues (Cont.)

- A BGP authentication mismatch causes a BGP session to not be established:
 - Authentication configured on one side only
 - Authentication key mismatch

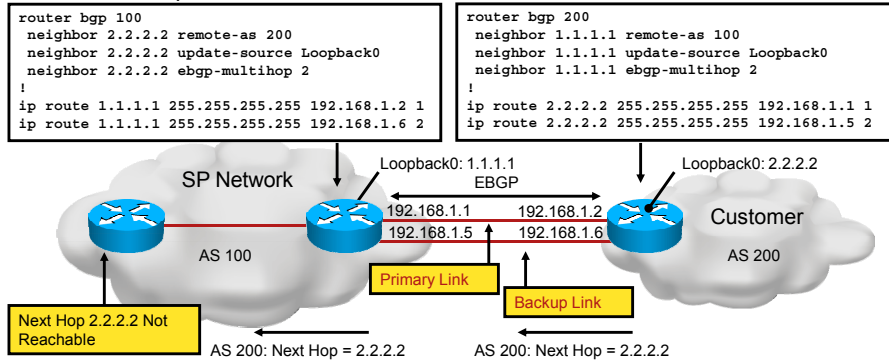


BGP supports Message Digest 5 (MD5) authentication. When configuring authentication, make sure that you enable authentication on both BGP neighbors. If you enable BGP authentication on one side or you misconfigure the passwords, the BGP session between the two neighbors will not be established. Routers will get stuck in the active state, which is not the desired state in BGP.

If you want to change the password between two BGP peers, remember that the local router will not tear down the existing session after you configure a new password. The local router will maintain the session using the new password until the BGP hold-down timer expires. This timer is 180 seconds by default. If the password is not changed or entered on the other router before the hold-down timer expires, the session times out.

EBGP Next-Hop Reachability Issues

- For EBGP, the next hop is the address of the router that sent the update:
 - By default, the next hop is set to the IP address from the connected interface.
 - In the example above, the next hop is set to the IP from loopback interface.



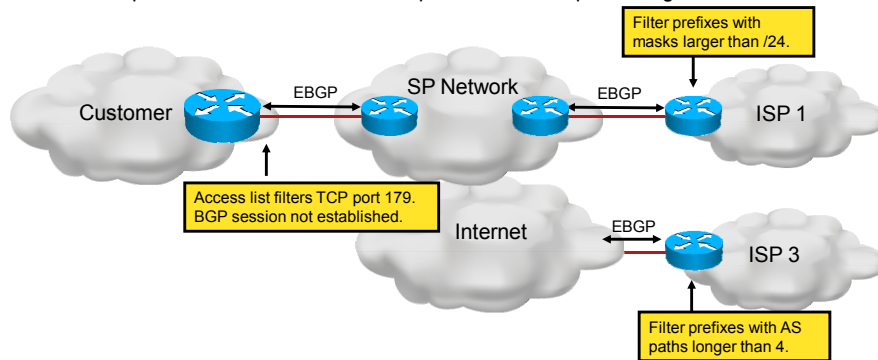
BGP is an external routing protocol that informs the next AS about networks in other autonomous systems. For EBGP, the next hop is the IP address of the router that sent the update. Internal Border Gateway Protocol (IBGP) operates differently. The next hop that is learned by an IBGP neighbor is not the IP address of the router that sent the update. The next hop that is learned by an IBGP neighbor is the IP address of the first router in the next AS on the path to the destination network.

In the example above, two autonomous systems are shown. Networks in AS 200 are advertised to the service provider autonomous system by the customer router through the EBGP session. Those networks are then further advertised to service provider routers through the IBGP session. BGP states that the next hop advertised by EBGP should be carried into IBGP. For this reason, the internal service provider router sees the network in AS 200 as available through the next hop of 2.2.2.2, which is the router in AS 200.

In the example, interior gateway protocol (IGP) should not advertise the loopback address of the customer into the service provider network. To solve this issue, you have to change the default next hop behavior on the provider edge router using the **neighbor ip-address next-hop-self** command.

EBGP Filtering Issues

- Bidirectional access list filtering of TCP port 179 will prevent a BGP session from being established.
- Prefix lists that filter out prefixes larger than /24:
 - Customer route has to be summarized
 - Customer load balancing may not be possible
- AS path access lists that filter out prefixes with AS paths longer than 4



A BGP session is a TCP session using port number 179. Two routers establishing a BGP session will first establish two sessions and later keep only one for communication. If there are filters (access lists or firewalls) that prohibit TCP port 179 in both directions, the BGP session will not be established.

Make sure to open TCP port number 179 at least in one direction.

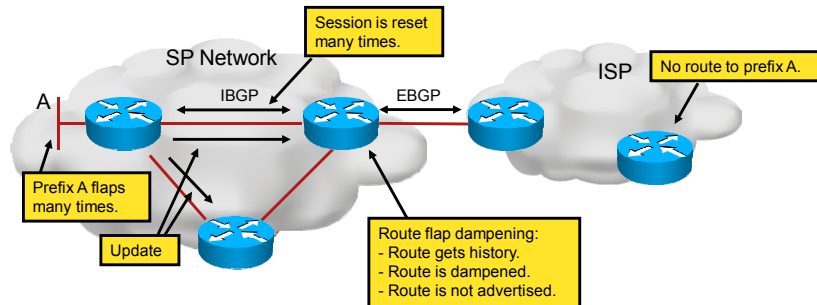
At many points, ISPs may set various filters for BGP routing information. The main reasons to set such filters are to minimize the BGP routing table, control CPU utilization, control memory utilization, and make the network more stable and predictable.

The following can be used to filter BGP routing information:

- **Prefix lists that deny routes with masks larger than /24:** More specific routes should be summarized or they will be filtered out.
- **Prefixes with AS paths that are too long:** These may be filtered out to limit the size of the BGP routing table. If service providers or customers are using the AS path prepend feature, routes may be filtered out.

Route Flap Dampening

- Route may be dampened:
 - Prefix flaps many times.
 - BGP session resets many times.
- A dampened route may not be reachable for up to one hour.



© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-4-10

When a customer tests or changes the BGP configuration, the customer may use the **clear ip bgp** Cisco IOS command to push a new policy in to work. Use of this command will reset the BGP session and cause all updates to flap. If such flaps occur many times in a short period, the service provider router may dampen all customer prefixes. Dampened routes are no longer advertised to the rest of the Internet. The customer may lose connectivity for up to 1 hour.

To overcome this issue, you may use the route refresh feature to push a new policy in to work, by using the **clear ip bgp soft** Cisco IOS command.

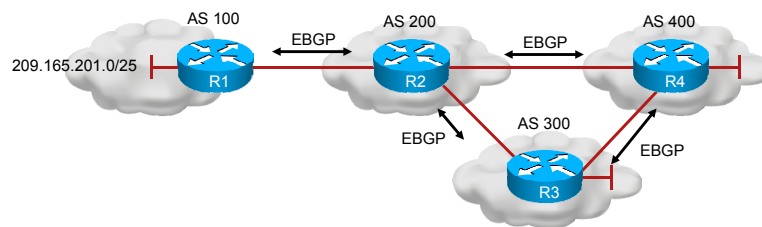
The route flap dampening feature is disabled by default.

Troubleshooting Typical Failures

This topic describes procedures for troubleshooting typical EBGP routing failures. Based on different monitoring outputs, various troubleshooting procedures are proposed.

Troubleshooting Routing Failures

- You receive a ticket from a customer indicating that its network cannot communicate with other networks.
- Start gathering information:
 - Know the topology, routing, and addressing
 - Use various **show** and **debug** commands



© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0—4-12

In this example, a customer cannot communicate with other networks or the Internet. The customer opens a ticket. You, as network administrator, should troubleshoot the problem and provide a solution.

First you must start gathering information. You should be familiar with the customer network topology, routing protocols that are used in the network, and addressing. If you are not familiar with the customer topology, troubleshooting is practically impossible. Next, you should start gathering additional information about the problem using various **show** and **debug** commands on Cisco devices.

In the example in this figure, the customer (AS 100) exchanges routing information with service provider (AS 200) using BGP. The service provider is running an EBGP session with other upstream service providers (AS 300 and AS 400).

Gather and Analyze Information

- Use ping and traceroute to determine the point of failure

```
R1#ping 210.166.4.1 source 209.165.201.1
Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 210.166.4.1, timeout is 2 seconds:
Packet sent with a source address of 209.165.201.1
.....
Success rate is 0 percent (0/5)
```

```
R1#traceroute 210.166.4.1 source 209.165.201.1

Type escape sequence to abort.
Tracing the route to 210.166.4.1
 0  *  *  *
 1  *  *  *
 2  *  *  *
```

- Check for link, path, or device failures, and correct any you find
- Verify the routing table for customer BGP routes

```
R1#show ip route bgp
```

- There are no BGP routes on SP routers. The router is not able to route to the Internet.

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-413

You should verify connectivity using the ping tool from the customer side to the service provider network. When you confirm the existence of a problem with connectivity, you should use the traceroute tool to determine the point of failure. Then you should check for any link, path, or device failures. If there are any, you should correct them and verify whether doing so solved the problem.

If there are no connectivity issues, you should start troubleshooting routing protocols. If BGP is used to advertise customer networks, you should start troubleshooting BGP. First you should verify whether there are any customer routes in routing tables on service provider routers. Use the **show ip route bgp** command on a Cisco IOS router or the **show route bgp** command on a Cisco IOS XR router.

In the example above, you can see that there are no BGP routes present on the service provider router. Obviously there is a problem with BGP inside the service provider network. In this example, the customer can reach other networks. The problem is with returning traffic, because there are no customer routes present on service provider routers.

Gather and Analyze Information

BGP Session

- Verify the BGP summary:

```
R1#show ip bgp summary
BGP router identifier 209.165.201.1, local AS number 100
< text omitted >

Neighbor      V    AS MsgRcvd MsgSent  TblVer  InQ  OutQ  Up/Down  State/PfxRcd
192.168.1.2    4    200     42     36      0    0    0 00:44:34 Active
```

- The neighbor is active, but no BGP session is established.

```
R1#show ip bgp neighbors 192.168.1.2
BGP neighbor is 192.168.1.2, remote AS 200, external link
  BGP version 4, remote router ID 0.0.0.0
  BGP state = Active
< text omitted >
Last reset 00:45:09, due to BGP Notification received, hold time expired
No active TCP connection
```

- There is no established TCP session to the BGP neighbor.

© 2010 Cisco Systems, Inc. All rights reserved.

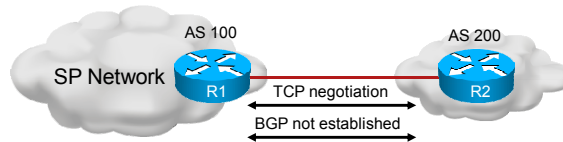
MSPRP v1.0—4-14

The next step is to verify if there a BGP session is established by using the **show ip bgp summary** and **show ip bgp neighbor** commands on Cisco IOS Software. To verify BGP sessions on a Cisco IOS XR router, use the **show bgp summary** and **show bgp neighbor** commands. In the example, you can see that no BGP session to the neighbor 192.168.1.2 is established. The reason is that the hold time has expired.

Gather and Analyze Information

Debug TCP

- MD5 password authentication is configured on the remote (R2) router:



```
R1#debug ip tcp transactions
TCP special event debugging is on
*Oct 6 09:45:27.531: MD5 received, but NOT expected from 192.168.1.2:38004 to
192.168.1.1:179
*Oct 6 09:45:27.535: TCP0: bad seg from 192.168.1.2 -- bad sequence number: port 179
seq 4257218527 ack 0 rcvnxt 4257218528 rcvwnd 16384 len 0
*Oct 6 09:45:27.671: TCP0: timeout #3 - timeout is 15988 ms, seq 2371990894
*Oct 6 09:45:27.675: TCP: (179) -> 192.168.1.2(38004)
*Oct 6 09:45:34.919: Released port 13780 in Transport Port Agent for TCP IP type 1
delay 240000
*Oct 6 09:45:34.923: TCP0: state was SYNSENT -> CLOSED [13780 -> 192.168.1.2(179)]
*Oct 6 09:45:34.931: TCB 0x65E87E30 destroyed
```

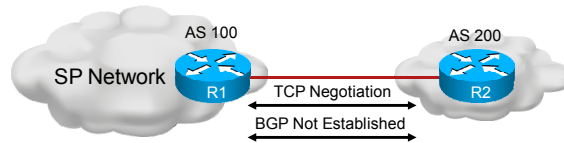
© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-4-15

To further troubleshoot the BGP session, you may use **debug** commands. The **debug ip tcp transactions** Cisco IOS command will display the state of TCP sessions from and to a router. Since BGP is using the TCP protocol, you can see in the example that packets have been received using authentication.

Gather and Analyze Information

Debug TCP (Cont.)



```
R1#show running-config | section router bgp
router bgp 100
  no synchronization
  bgp log-neighbor-changes
  network 209.165.201.0 mask 255.255.255.128
  neighbor 192.168.1.2 remote-as 200
  no auto-summary
```

© 2010 Cisco Systems, Inc. All rights reserved.

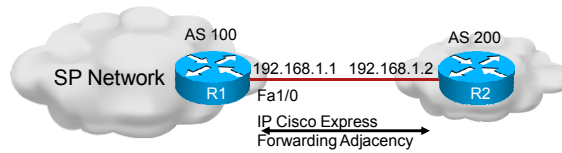
MSPRP v1.0—4-16

By verifying router configuration, you can see that no BGP authentication is configured on the local router.

For BGP authentication to be properly configured, both routers should have authentication enabled and the same password configured. The most common mistake is typing the password and leaving a space at the end of the **neighbor ip-address password level password** command. The space becomes part of the password.

Verify IP Cisco Express Forwarding

- Verify IP Cisco Express Forwarding adjacency:



```
R1#show ip cef adjacency FastEthernet 1/0 192.168.1.2
%CEF not running
Prefix      Next Hop      Interface
```

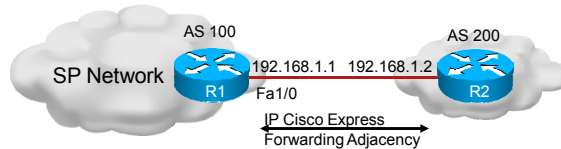
© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0-4-17

Cisco Express Forwarding must be enabled for faster packet forwarding and for many features to work properly (for example, Multiprotocol Label Switching [MPLS]). To check if IP Cisco Express Forwarding is enabled, use the **show ip cef adjacency** Cisco IOS command or the **show cef adjacency** Cisco IOS XR command.

Verify IP Cisco Express Forwarding (Cont.)

- Enable IP Cisco Express Forwarding:



```
R1#show ip route
< text omitted >
209.165.201.0/25 is subnetted, 1 subnets
C    209.165.201.0 is directly connected, Loopback0
B    210.166.4.0/24 [20/0] via 192.168.1.2, 00:09:15
B    211.167.3.0/24 [20/0] via 192.168.1.2, 00:09:15
C    192.168.1.0/24 is directly connected, FastEthernet1/0
```

```
R1#show ip cef adjacency FastEthernet 1/0 192.168.1.2
Prefix      Next Hop      Interface
192.168.1.2/32 192.168.1.2    FastEthernet1/0
210.166.4.0/24 192.168.1.2    FastEthernet1/0
211.167.3.0/24 192.168.1.2    FastEthernet1/0
```

© 2010 Cisco Systems, Inc. All rights reserved.

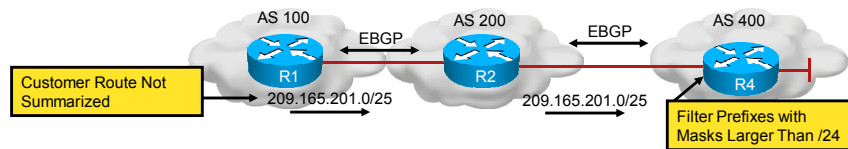
MSRP v1.0-4-18

In the example, after you enable IP Cisco Express Forwarding on the router, you may see all three prefixes reachable via Fast Ethernet 1/0 interface shown in the IP Cisco Express Forwarding adjacency table.

Filter Prefixes

- Display the routes advertised to a BGP neighbor:

```
RI#show ip bgp neighbors 192.168.1.2 advertised-routes
BGP table version is 8, local router ID is 209.165.201.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete
              Network        Next Hop          Metric LocPrf Weight Path
*> 209.165.201.0/25 0.0.0.0          0           32768 i
Total number of prefixes 1
```



© 2010 Cisco Systems, Inc. All rights reserved.

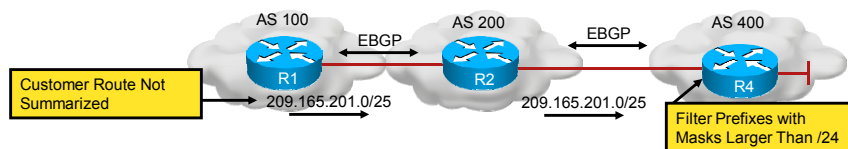
MSPRP v1.0—4-19

In the example, the customer is sending a too-specific route with the mask /25. An upstream provider has a prefix list configured to filter out routes with masks larger than /24. For the customer to get Internet reachability, either the customer or the service provider should aggregate the route.

Filter Prefixes (Cont.)

- Aggregate the customer route to prefix with mask length 20:

```
RI(config-router)#aggregate-address 209.165.192.0 255.255.240.0
RI(config-router)#do show ip bgp neighbors 192.168.1.2 advertised-routes
BGP table version is 9, local router ID is 209.165.201.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete
              Network        Next Hop          Metric LocPrf Weight Path
*> 209.165.192.0/20 0.0.0.0          0           32768 i
*> 209.165.201.0/25 0.0.0.0          0           32768 i
Total number of prefixes 2
```



© 2010 Cisco Systems, Inc. All rights reserved.

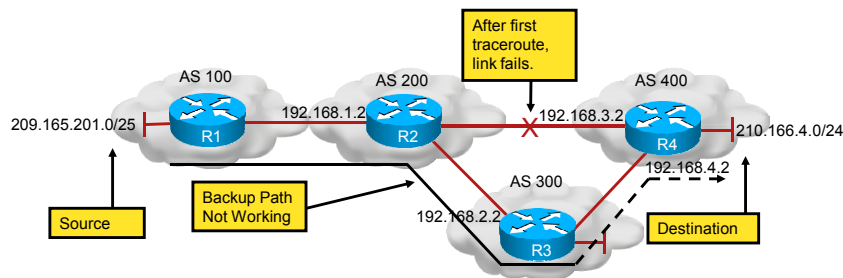
MSPRP v1.0—4-20

In the example, the customer prefix is aggregated with the mask /20, thus allowing the customer to communicate with the rest of the Internet.

Multiple BGP Paths

- Traceroute via the primary path (the R2-R4 link is active):

```
R1#traceroute 210.166.4.1 source 209.165.201.1
Type escape sequence to abort.
Tracing the route to 210.166.4.1
 1 192.168.1.2 120 msec 168 msec 120 msec
 2 192.168.3.2 124 msec 180 msec *
```

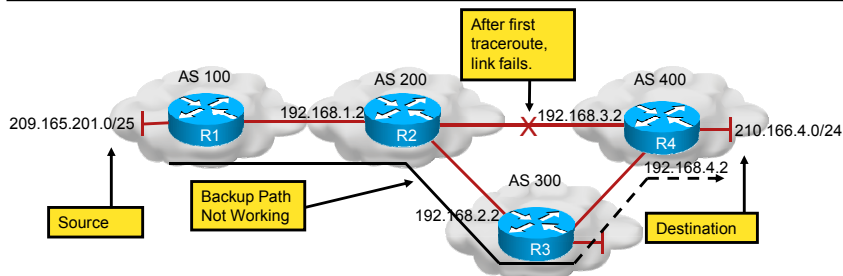


This shows the situation with the primary path active (the link between R2 and R4 active) and the situation with the primary path inactive. Packets should be routed via a second upstream service provider in AS 300.

Multiple BGP Paths (Cont.)

- Traceroute via the backup path (R3 the is last router to respond):

```
R1#traceroute 210.166.4.1 source 209.165.201.1 numeric
Type escape sequence to abort.
Tracing the route to 210.166.4.1
 1 192.168.1.2 128 msec 76 msec 12 msec
 2 192.168.2.2 132 msec 60 msec 60 msec
 3 * * *
```

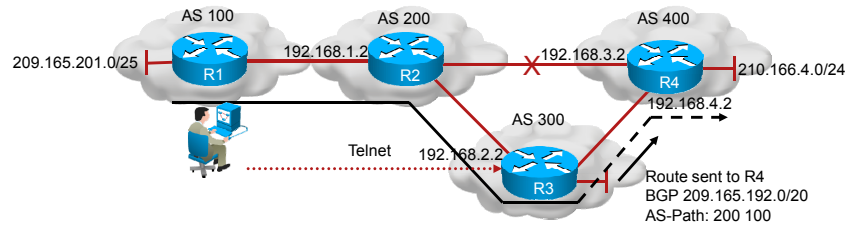


Traceroute shows that the router in AS 300 is the last to respond. To find the root of this problem, you should concentrate on the BGP session between AS 300 and AS 400.

BGP Looking Glasses

- Connect via Telnet to the R3 router.
- Verify that the customer route is present:

```
R3#show ip bgp 209.165.192.0/20
BGP routing table entry for 209.165.192.0/20, version 7
Paths: (1 available, best #1, table Default-IP-Routing-Table)
  Advertised to update-groups:
    2
  200 100, (aggregated by 100 209.165.201.1)
    192.168.2.1 from 192.168.2.1 (192.168.3.1)
      Origin IGP, localpref 100, valid, external, atomic-aggregate, best
```



© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0—4-23

To get some information from distant autonomous systems (AS 300), you should use BGP looking glasses.

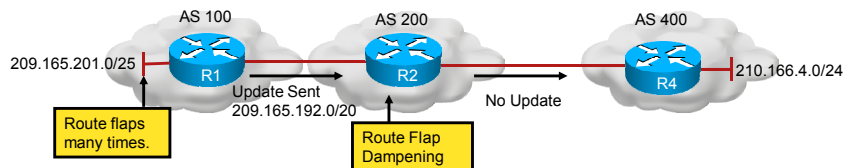
This shows a Telnet connection to the BGP looking glass server in AS 300. You can see that the route from the customer (209.165.201.0/20) is present.

This is the only information that you can gather from distant autonomous systems using BGP looking glasses.

Route-Flap Dampening

- No route received on the R4 router:

```
R4#show ip bgp | include 209.165.192.0
R4#
```



© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-4-24

If a customer prefix in AS 100 flaps many times in a short period, customer prefixes may be dampened. Prefixes may also be dampened if a BGP session between the service provider (AS 200) and customer is reset for testing purposes many times.

Route-Flap Dampening (Cont.)

- Flap statistics on the R2 router:

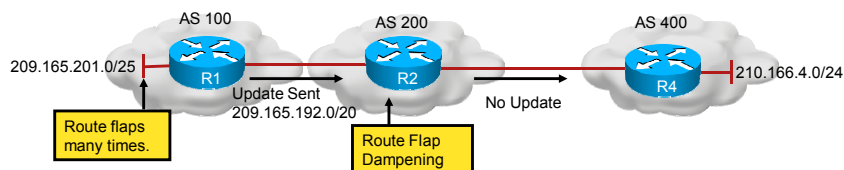
```
R2#show ip bgp dampening flap-statistics
BGP table version is 19, local router ID is 192.168.3.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete
```

Network	From	Flaps	Duration	Reuse	Path
*d 209.165.192.0/20	192.168.1.1	7	00:36:51	00:05:09	100

Route Damped

Number of Flaps

Reused in 5 Minutes and 9 Seconds



© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-4-25

In the service provider autonomous system, the R2 router is configured with the route flap dampening feature. By using the **show ip bgp dampening flap-statistics** Cisco IOS command, you can verify which route is dampened. You can also verify the number of flaps and when the route will be reused and sent further to upstream service providers.

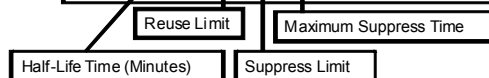
Route-Flap Dampening Information

- A BGP route gets a history with penalty points:

```
R2#show ip bgp 209.165.192.0/20
BGP routing table entry for 209.165.192.0/20, version 30
Paths: (1 available, no best path)
Flag: 0x820
Not advertised to any peer
100, (aggregated by 100 209.165.201.1), (suppressed due to dampening) (history
entry)
192.168.1.1 from 192.168.1.1 (209.165.201.1)
Origin IGP, metric 0, localpref 100, external, atomic-aggregate
Dampinfo: penalty 6052, flapped 7 times in 00:08:47, reuse in 00:01:28
```

- BGP route-flap dampening parameters (default):

```
R2#show ip bgp dampening parameters
dampening 15 750 2000 60 (DEFAULT)
Half-life time : 15 mins      Decay Time      : 2320 secs
Max suppress penalty: 12000    Max suppress time: 60 mins
Suppress penalty : 2000       Reuse penalty   : 750
```



© 2010 Cisco Systems, Inc. All rights reserved.

MSPRP v1.0-426

You can use these additional commands to see route dampened statistics and parameters:

- The **show ip bgp prefix** Cisco IOS command shows that a route is suppressed due to dampening. It also shows penalty points, the number of flaps, and when the route will be reused.
- The **show ip bgp dampening parameters** Cisco IOS command lets you verify dampening parameters. In the previous example, you can see default dampening parameters: half-life time, reuse limit, suppress limit, and maximum suppress time.

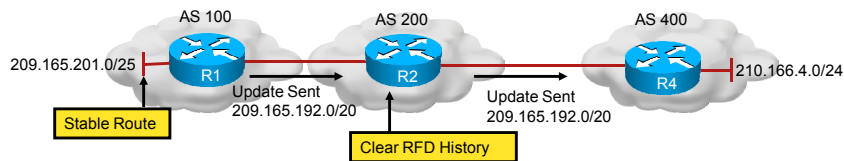
Clear Route-Flap Dampening History

- A BGP route gets a history with penalty points:

```
R2#clear ip bgp dampening 209.165.192.0 255.255.240.0
R2#
```

- The BGP route now has no route-flap dampening history:

```
R2#show ip bgp 209.165.192.0/20
BGP routing table entry for 209.165.192.0/20, version 2
Paths: (1 available, best #1, table Default-IP-Routing-Table)
Flag: 0x820
Advertised to update-groups:
  1
  100, (aggregated by 100 209.165.201.1)
    192.168.1.1 from 192.168.1.1 (209.165.201.1)
      Origin IGP, metric 0, localpref 100, valid, external, atomic-aggregate, best
```

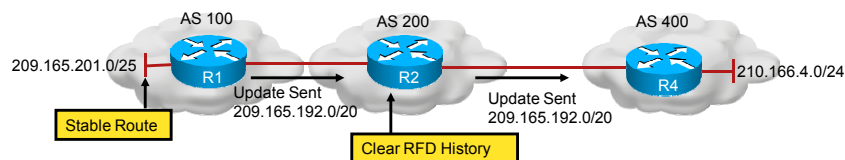


To clear route flap dampening history, use the **clear ip bgp dampening prefix mask** Cisco IOS command. This command will delete the route flap dampening statistics and the route will be immediately sent further to the upstream service provider.

Clear Route-Flap Dampening History (Cont.)

- Flap statistics on the R2 router have been deleted:

```
R2#show ip bgp dampening flap-statistics
R2#
```



After you clear route flap dampening history, the route has no flap dampening statistics.

Summary

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

Summary

- Typical failures in service provider networks stem from connectivity issues or device misconfigurations.
- Typical BGP misconfigurations are authentication mismatches, wrong peering IP addresses, and the unavailability of a next hop or peering IP address.
- Troubleshoot problems by indicating the problem, then diagnosing a cause and implementing a solution.
- To diagnose the cause of a problem, gather and analyze information and propose and test hypotheses.
- Use various **show** and **debug** commands to gather and analyze information.

© 2010 Cisco Systems, Inc. All rights reserved.

MSRP v1.0—429

Lesson Self-Check

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

- Q1) Which two EBGP misconfigurations may happen in networks of service providers when new devices are introduced? (Choose two.) (Source: Typical External Routing Failures in Service Provider Environments)
- A) different authentication types configured on both sides of a BGP session
 - B) different encapsulation types configured on both sides of a link
 - C) no keepalives configured on a link toward the upstream service provider
 - D) no route configured to a neighbor loopback IP address
- Q2) Which two EBGP filtering issues may happen in networks of service providers? (Choose two.) (Source: Typical External Routing Failures in Service Provider Environments)
- A) A prefix list filters out prefixes with large subnet masks.
 - B) An AS path access list filters out prefixes with long AS paths.
 - C) An access list filters out routing information.
 - D) BGP routes with communities are filtered out.
- Q3) Which command displays the reason why a BGP session is not established? (Source: Troubleshooting Typical Failures)
- A) **show ip bgp summary**
 - B) **show ip bgp neighbors** *ip-address*
 - C) **show running-config**
 - D) **show ip route bgp**
 - E) **ping** *ip-destination* **source** *ip-source*
- Q4) When using BGP looking glasses to a distant AS, which two commands may be issued on the remote system? (Choose two.) (Source: Troubleshooting Typical Failures)
- A) **show ip bgp summary**
 - B) **show ip bgp**
 - C) **show ip bgp neighbors** *ip-address*
 - D) **show ip bgp** *prefix/mask*
- Q5) What are two possible reasons for route flap dampening to occur? (Choose two.) (Source: Troubleshooting Typical Failures)
- A) A BGP prefix flaps many times because an interface flaps.
 - B) A BGP prefix flaps once only because an interface flaps.
 - C) IGP calculates a new best path in the service provider network.
 - D) A BGP session is reset many times for testing purposes.
 - E) The BGP attributes router does not like a prefix and suppresses it.

Lesson Self-Check Answer Key

- Q1) A, D
- Q2) A, B
- Q3) B
- Q4) B, D
- Q5) A, D

Module Summary

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

Module Summary

- Troubleshoot incidents:
 - Define incident parameters.
 - Diagnose causes.
 - Devise solutions (permanent or temporary).
- To diagnose the cause of an incident, gather and analyze information and propose and test hypotheses.
- Use workarounds to minimize downtime.
- Determine and propose a solution to an incident (especially if it is recurring) to prevent it from happening again.
- Use various **show** and **debug** commands to gather and analyze information.

© 2010 Cisco Systems, Inc. All rights reserved.

Course acronym vx.x--B.1

Typical failures in service provider networks stem from connectivity issues or device misconfigurations. After configuration changes or the introduction of new equipment in a network, some routing issues may arise. Examples of routing misconfigurations are an authentication mismatch, an MTU or timer mismatch, a wrong peering IP address, and the unavailability of a next hop. This module covered how to gather information about a problem and how to solve it. This module also described how to troubleshoot a problem by indicating the problem, diagnosing a cause, and implementing a solution. The module discussed how to diagnose the cause of a problem, gather and analyze information using various **show** and **debug** commands, and how to propose and test hypotheses.

