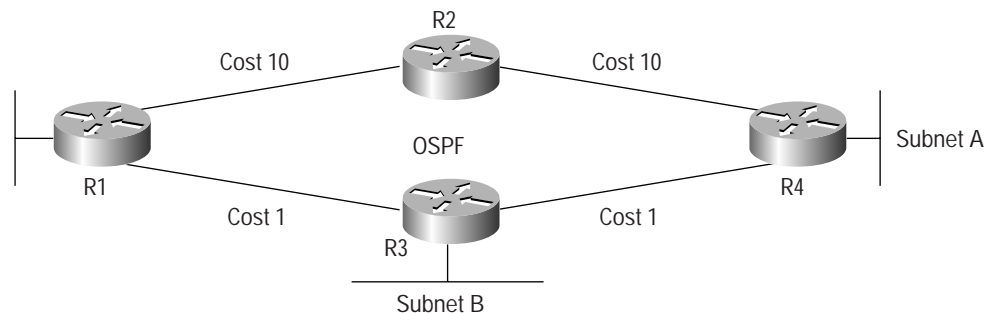


Using *OSPF* Stub Router Advertisement Feature

Overview

It can be desirable to have a router in a network that is not a transit router. A “transit router” forwards traffic that is not destined for networks directly connected to the router. A router that only forwards packets destined to its directly connected links is known as a “stub router”. In *OSPF* networks, a router could become a stub router by advertising large metrics for its connected links, so that the cost of a path through this router becomes larger than that of an alternative path. It does not matter what specific metrics are advertised in a stub network that is connected to the router, because data must reach the network via the stub router itself.

Figure 1
OSPF Networks



R1 uses R3 as next-hop to reach subnet A because of the *OSPF* cost (Figure 1). If R3 cannot act as a transit router, R3 can advertise a large metric for its connected links (except subnet B, a stub network). As a result, R1 will use R2 as next-hop to reach subnet A. R1 will still use R3 to reach subnet B, regardless of the metric associated with the link.

OSPF Stub Router Advertisement allows a router to advertise infinity metric (0xFFFF) for its connected links in Router-LSAs, and advertise normal interface cost if the link is a stub network. The feature also provides three options for administrators to determine when to advertise such infinity metric:

1. Always send route-LSAs with infinity metric after enabling this feature
2. Send route-LSAs with infinity metric after reboot for a specified time
3. Send route-LSAs with infinity metric after reboot until BGP is converged (or maximum timer is expired)



Benefit

- Graceful removal of an OSPF router from networks (i.e. to upgrade software or to perform maintenance) and reduction in packet drop
- Reduce black-hole packet dropping when interacting with BGP

The rest of this paper discusses scenarios for using this feature to achieve such benefits.

Command Syntax

Here is the command syntax for the three options of this feature. They are configured under “router ospf <pid>”.

1. To always originate router-LSAs with infinity metric after configuration
`router(config-router)#max-metric router-lsa`
2. To send router-LSAs with infinity metric after reboot in a specified time
`router(config-router)#max-metric router-lsa on-startup <announce-time>`
3. To send router-LSAs after reboot until BGP is converged (or maximum timer is expired – 600sec)
`router(config-router)#max-metric router-lsa on-startup wait-for-bgp`

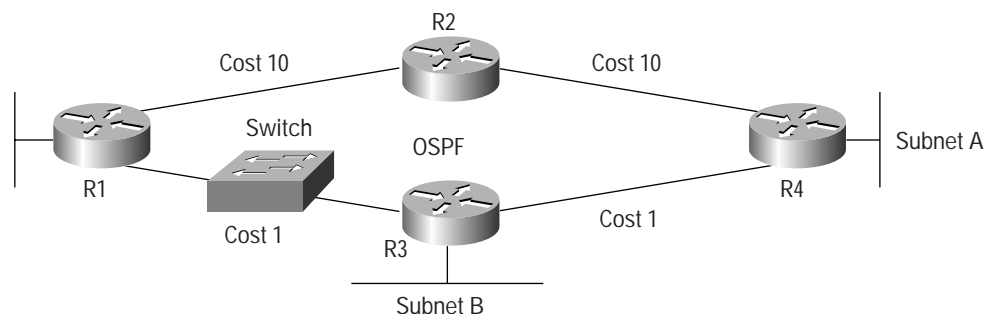
Usage Examples

Gracefully Remove OSPF Router from Networks and Reduce Number of Transit Packets Dropped

When an OSPF router is removed from networks and its neighbors cannot detect the physical interface is down (i.e., switch between the removed router and its neighbors), the neighbors will need to wait until the hold-timer expires to remove the adjacency and re-converge. This may cause packets to be dropped.

Consider the following example, which includes a switch between R1 and R3:

Figure 2
OSPF Network Example



Suppose that R3 must be removed from the network for maintenance. If it is removed simply by powering down, the OSPF process of R1 will not detect that R3 is unavailable until the holdtimer is expired. Within that period, R1 will continue forwarding packets destined to subnet A to the link that is connected to the switch. The switch will drop the packets. To reduce the number of packets dropped, R1 can use the path via R2, prior to removing R3 from the networks.



The following steps are necessary to using this feature of OSPF Stub Router Advertisement.

To gracefully remove R3, perform the following procedures:

1. Configure “max-metric router-lsa” under “router ospf <pid>” on R3
 - This step causes the router to generate new LSAs with infinity metric for the R1 to R3, and R3 to R4 links. Upon receiving the LSAs, R1 will perform SPF and use the R1-R2-R4 path for packets destined to subnet A.
2. Make sure R1 uses the R1-R2-R4 path for the transit traffic – i.e. using “show ip route” on R3. (in this example the transit traffic is the packets destined to subnet A.)
 - Step 2 occurs only for the purpose of confirmation.
3. Shut down R3 and remove it from the network
 - No packets that are destined to subnet A will then be dropped, since R3 is no longer forwarding transit traffic. R3 can now be gracefully removed from the networks.

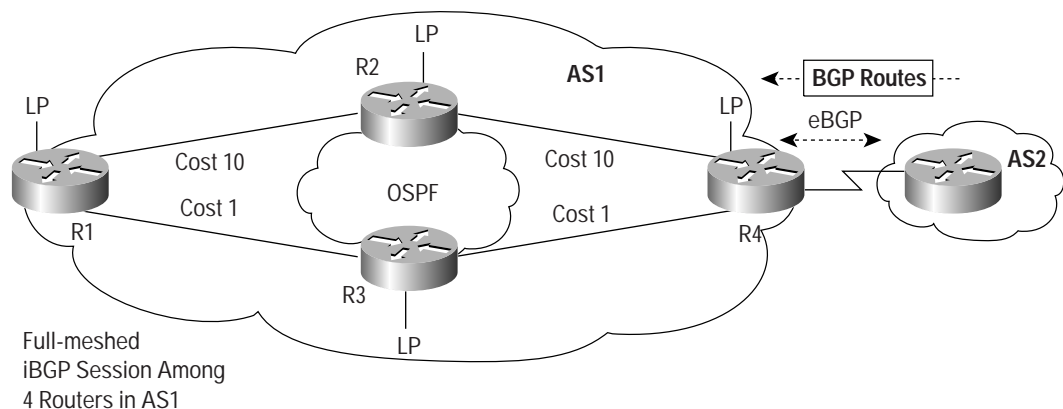
Avoid Routing Black-hole by Waiting for BGP to Converge

Rebooting Internet backbone routers can lead to a loss of Internet connectivity. The IGP may converge quickly after a router reboots, and other routers may try to forward traffic through the just rebooted router.

This router may still be building its BGP routing tables, and may not have fully converged yet. While this happens, it may drop many packets for destinations it has not learned via BGP yet. It would be far more beneficial for stability and availability if a router could build its BGP routing table while it is not yet fully participating in packet forwarding.

With this feature and its “wait-for-bgp” option, the router advertises its locally generated router LSAs with infinity metric until BGP is converged or after a default maximum timeout (600 sec) is expired. This action will allow the router to converge without attracting transit traffic if there are better alternate paths around this router (Figure 3).

Figure 3
BGP and OSPF Networks



In Figure 2, the R1, R2, R3, and R4 within AS1 are running full-meshed iBGP. A loopback interface on each of these routers is used as the BGP peering address, and the loopback interfaces are in OSPF domain. Therefore, if the R1-R3-R4 path is unavailable, the BGP session between R1 and R4 can still use R1-R2-R4 path, so the BGP session is assumed to be intact.



R4 has an eBGP session with a router in AS2, and it receives BGP routes from AS2. The “next-hop self” is configured for all iBGP neighbors on R4. All four of the routers have the BGP routes in the routing table. In a typical ISP environment, the following is recommended for all routers in the transit path that run BGP: no redistribution from BGP to IGP and the enabling of “no sync”.

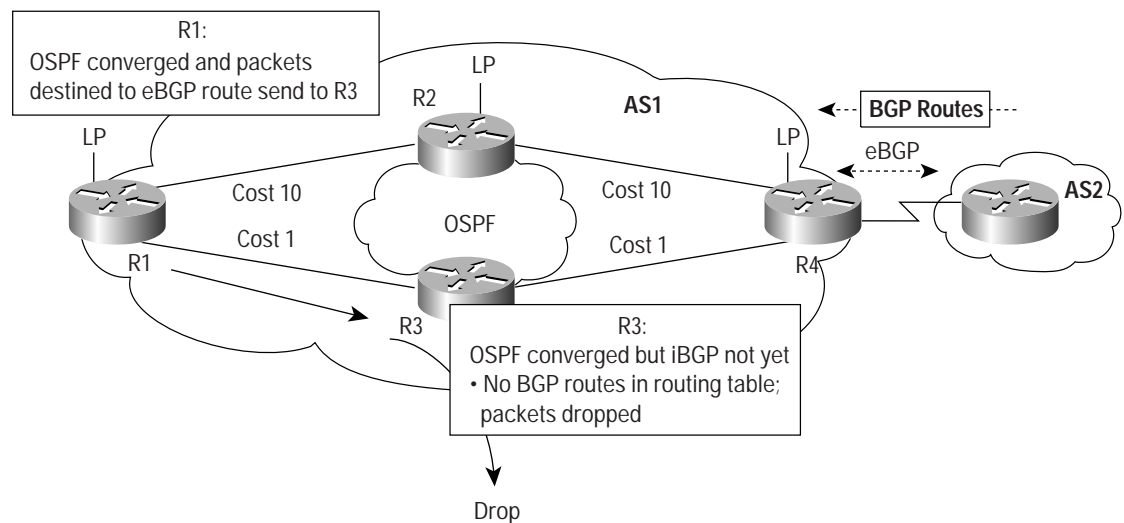
In normal operation, to reach the BGP routes in AS2, R1 performs a recursive lookup for the next-hop address of the BGP routes, which is the loopback interface of R4. As a result, R1 uses R1-R3-R4 path for routing packets destined to the BGP routes (R1-R3-R4 path cost 2 < R1-R2-R4 path cost 20).

If R3 were rebooted, then R1 would lose the OSPF adjacency to R3. To reach the next-hop address of the BGP routes, R1 would take the R1-R2-R4 path. Thus, the packets destined to the BGP routes are continuously forwarded to its destinations.

The OSPF and iBGP within AS1 try to converge after R3 boots. OSPF typically converges faster than BGP, so assume that occurs in this example. R1 would have the loopback interface of R4 in its routing table via R3 as result of SPF calculation.

Recall that the loopback interface is the next-hop address of the BGP routes (next-hop self). Consequently, R1 would use R1-R3-R4 path to reach the BGP routes. Since the iBGP is not converged yet, the R3 does not have the BGP routes in its routing table. When it receives the packets coming from R1 destined to the BGP routes, it will drop the packets until the BGP is converged (see figure below). This is sometimes referred to as routing black hole.

Figure 4
Routing Black-Hold Example



To avoid a routing black hold, the OSPF stub router advertisement feature can be used with “wait-for-bgp” option.

```
R3(config-router)# max-metric router-lsa on-startup wait-for-bgp
```



With this command, R3 advertises infinity metric for transit links in its router-LSA, which causes R1 to continue using the R1-R2-R4 path. After BGP is converged on R3, the OSPF process of R3 will receive notification that causes it to stop advertising route-LSAs with infinity metric and generate new LSAs with regular interface cost for the associated links. R1 will then receive the new LSA and start using R1-R3-R4 path for the packets destined to the eBGP routes as result of SPF calculation. This avoids the routing black hole that would otherwise occur as shown.

Other Situations to Use This Feature

- To avoid forwarding transit traffic to a router that is in a critical condition (i.e. a router has very high CPU load or does not have enough memory to store all LSAs or build the routing table)
- Graceful introduction of a router to the network (i.e. a test router in lab is going to connect to a production network, so it must “fit” properly before starting to forward transit traffic to it).

Recommendation

Assume there is redundant path in OSPF networks.

- When a router is running OSPF and BGP together and OSPF is used to resolve the next-hop address of the BGP learned routes, configure “max-metric router-lsa on-startup wait-for-bgp”
- When removing an OSPF router from networks, to avoid packet drop configure “max-metric router-lsa” first, then confirm the alternative path is used, finally remove the OSPF router.

Software

- Cisco IOS Software Release 12.0.15S or later
- Cisco IOS Software Release 12.0.16ST or later
- Currently scheduled for Release 12.1(8)E
- Currently scheduled for Release 12.2(3)T

Reference

- CSCdp82156
- RFC 3137—OSPF Stub Router Advertisement

Appendix—Example of Show Output

1. Output of “show ip ospf”

Note: The highlighted information above varies with configuration. See table1 below for details.

- Routing Process “ospf 1998” with ID 14.18.134.155
- Supports only single TOS(TOS0) routes
- Supports opaque LSA
- It is an area border and autonomous system boundary router
- Redistributing External Routes from,
- static, includes subnets in redistribution
- Originating router-LSAs with maximum metric, Time remaining: 00:01:18



- Condition: on startup while BGP is converging, State: active
- SPF schedule delay 5 secs, Hold time between two SPF's 10 secs
- Minimum LSA interval 5 secs. Minimum LSA arrival 1 secs
- Number of external LSA 7. Checksum Sum 0x47261
- Number of opaque AS LSA 0. Checksum Sum 0x0
- Number of DCbitless external and opaque AS LSA 0
- Number of DoNotAge external and opaque AS LSA 0
- Number of areas in this router is 2. 1 normal 0 stub 1 nssa
- External flood list length 0

Area BACKBONE(0)

- Number of interfaces in this area is 1
- Area has no authentication
- SPF algorithm executed 3 times
- Area ranges are
- Number of LSA 8. Checksum Sum 0x474AE
- Number of opaque link LSA 0. Checksum Sum 0x0

Table 1

Status	Command	Output
Active	max-metric router-lsa	Originating router-LSAs with maximum metric Condition: always, State: active
	max-metric router-lsa on-startup <time>	Originating router-LSAs with maximum metric, Time remaining: 00:01:51 Condition: on startup for 300 seconds, State: active
	max-metric router-lsa on-startup wait-for-bgp	Originating router-LSAs with maximum metric, Time remaining: 00:01:18 Condition: on startup while BGP is converging, State: active
Inactive	max-metric router-lsa on-startup <time>	Originating router-LSAs with maximum metric Condition: on startup for 300 seconds, State: inactive
	max-metric router-lsa on-startup wait-for-bgp	Originating router-LSAs with maximum metric Condition: on startup while BGP is converging, State: inactive

2. Output of "show ip ospf database router" command:

Exception Flag: Announcing maximum link costs

LS age: 68
Options: (No TOS-capability, DC)
LS Type: Router Links
Link State ID: 14.18.134.155
Advertising Router: 14.18.134.155
LS Seq Number: 80000002
Checksum: 0x175D
Length: 60
Area Border Router
AS Boundary Router
Number of Links: 3

Link connected to: a Transit Network

(Link ID) Designated Router address: 192.1.1.11

(Link Data) Router Interface address: 192.1.1.14
Number of TOS metrics: 0
TOS 0 Metrics: 65535 (metric used for local calculation: 10)

Link connected to: a Transit Network

(Link ID) Designated Router address: 10.1.145.11
(Link Data) Router Interface address: 10.1.145.14
Number of TOS metrics: 0
TOS 0 Metrics: 65535 (metric used for local calculation: 10)

Link connected to: a Stub Network

(Link ID) Network/subnet number: 9.11.12.0
(Link Data) Network Mask: 255.255.255.0
Number of TOS metrics: 0
TOS 0 Metrics: 1



Corporate Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
www.cisco.com
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 526-4100

European Headquarters
Cisco Systems Europe
11 Rue Camille Desmoulins
92782 Issy-les-Moulineaux
Cedex 9
France
www-europe.cisco.com
Tel: 33 1 58 04 60 00
Fax: 33 1 58 04 61 00

Americas Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
www.cisco.com
Tel: 408 526-7660
Fax: 408 527-0883

Asia Pacific Headquarters
Cisco Systems, Inc.
Capital Tower
168 Robinson Road
#22-01 to #29-01
Singapore 068912
www.cisco.com
Tel: +65 317 7777
Fax: +65 317 7799

Cisco Systems has more than 200 offices in the following countries and regions. Addresses, phone numbers, and fax numbers are listed on the
Cisco Web site at www.cisco.com/go/offices

Argentina • Australia • Austria • Belgium • Brazil • Bulgaria • Canada • Chile • China PRC • Colombia • Costa Rica • Croatia
Czech Republic • Denmark • Dubai, UAE • Finland • France • Germany • Greece • Hong Kong SAR • Hungary • India • Indonesia • Ireland
Israel • Italy • Japan • Korea • Luxembourg • Malaysia • Mexico • The Netherlands • New Zealand • Norway • Peru • Philippines • Poland
Portugal • Puerto Rico • Romania • Russia • Saudi Arabia • Scotland • Singapore • Slovakia • Slovenia • South Africa • Spain • Sweden
Switzerland • Taiwan • Thailand • Turkey • Ukraine • United Kingdom • United States • Venezuela • Vietnam • Zimbabwe

All contents are Copyright © 1992-2002, Cisco Systems, Inc. All rights reserved. Cisco, Cisco IOS, Cisco Systems, and the Cisco Systems logo are registered trademarks of Cisco Systems, Inc. and/or its affiliates in the U.S. and certain other countries.

All other trademarks mentioned in this document or Web site 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.
(0203R) 202822.A/ETMG 7/02