Netwok Working Group
Internet Draft

Danny McPherson
Amber Networks
Tony Przygienda
Redback

draft-mcpherson-ospf-transient-00.txt

July 2000

OSPF Transient Blackhole Avoidance

Status of this Memo

Abstract

This document describes a simple, interoperable mechanism that can be employed in OSPF networks in order to decrease data loss associated with deterministic blackholing of packets during transient network conditions. The mechanism proposed here requires no OSPF protocol changes and is completely interoperable with the existing OSPF specification.

Specification of Requirements

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

Table of Contents

1. Overview

When an OSPF router that was previously a transit router becomes unavailable as a result of some transient condition such as a reboot, other routers within the routing domain must select an alternative path to reach destinations which had previously transited the failed router. Presumably, the newly selected router(s) comprising the path have been available for some time and, as a result, have complete forwarding information bases (FIBs) which contain a full set of reachibility information for both internal and external (e.g. BGP) destinations.

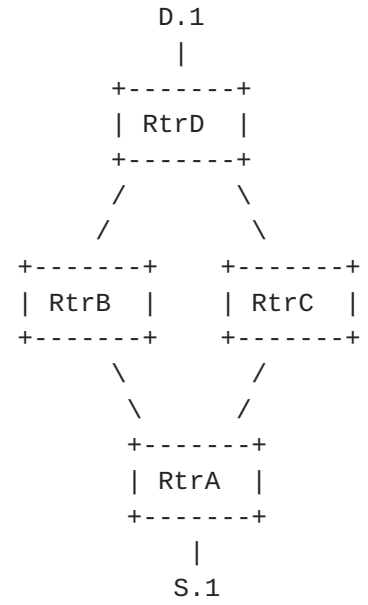
When the previously failed router becomes available again, in only a few seconds paths that had previously transited the router are again selected as the optimal path by the IGP. As a result, forwarding tables are updated and packets are once again forwarded along the path. Unfortunately, external destination reachibility information (e.g. learned via BGP) is not yet available to the router, and as a result, packets bound for destinations not learned via the IGP are unnecessarily discarded.

A mechanism to alleviate the offshoot associated with this deterministic behavior is discussed below.

2. Discussion

This document describes a simple, interoperable mechanism that can be employed in OSPF [RFC2328] networks in order to avoid transition to a newly available path until other associated routing protocols such as BGP have had sufficient time to converge.

The benefits of such a mechanism can realized when considering the following scenario.



Host S.1 is transmitting data to destination D.1 via a primary path of RtrA->RtrB->RtrD. Routers A, B and C learn of reachibility to destination D.1 via BGP from RtrD. RtrA's primary path to D.1 is selected because when calculatng the path to BGP NEXT_HOP of RtrD the sum of the OSPF link costs on the RtrA-RtrB-RtrD path is less than the sum of the costs of the RtrA-RtrC-RtrD path.

Assume RtrB becomes unavailable and as a result the RtrC path to RtrD is used. Once RtrA's FIB is updated and it begins forwarding packets to RtrC everything should behave properly as RtrC has existing forwarding information regarding destination D.1's availability via BGP NEXT_HOP RtrD.

Assume now that RtrB comes back online. In only a few seconds OSPF neighbor state is been established with RtrA and RtrD and database synchronization has occurred. RtrA now realizes that the best path to destination D.1 is via RtrB, and therefore updates it FIB appropriately. RtrA begins to forward packets destined to D.1 to RtrB. Though, because RtrB has yet to establish and synchronization it's BGP neighbor relationship and routing information with RtrD, RtrB has no knowledge regarding reachibililty of destination D.1, and therefore discards the packets received from RtrA.

If RtrB were to temporarily set it's link costs to 0xFFFF while synchronizing with BGP tables with it's neighbors, RtrA would continue to use the working RtrA->RtrC->RtrD path. Upon intial synchronization of BGP tables with neighboring router, RtrB would generate a new LSA describing the actual link costs associated with each connection, and RtrA could again begin using the optimal path via RtrB.

However, if no alternative path were available to destination D.1 RtrA would still be able to use the path via RtrB, and manipulation

of the link costs would result in no adverse effect.

3. Deployment Considerations

Such a mechanism increases overall network availablity and allows network operators to alleviate the deterministic blackholing behavoir introduced in this scenario. The IS-IS overload bit has been employed in IS-IS routing domains to achieve similar behavior.

Triggers for setting the link costs as described are left to the implementor. Some potential triggers could include N seconds after booting, or N number of BGP prefixes in the BGP Loc-RIB.

Also, understand that this mechanism assumes actual deployments assign substantially lower values for link costs (and the sum of subsequent path costs), and that a value of 0xFFFF for an individual link within a path would be sufficiently large enough to discourage transit traffic from entering the router.

4. Security Considerations

Security issues are not discussed in this memo.

5. Acknowledgements

The authors would like to acknowledge John Moy of Sycamore Networks for his valuable input.

6. References

[RFC2328] Moy, J., "OSPF Version 2", RFC 2328, April 1998.

7. Authors' Address

Danny McPherson
Amber Networks
2465 Augustine Drive
Santa Clara, CA 95054
Phone: +1 408.486.6336
Email: danny@ambernetworks.com
John Moy
Sycamore Networks

Tony Przygienda
Redback

350 Holger Way
San Jose, CA 95134
Email: prz@redback.com