Network Working Group
Internet Draft
Document: draft-mirtorabi-ospf-tunnel-adjacency-01.txt
Expiration Date: June 2004

Sina Mirtorabi
Peter Psenak
Cisco Systems, Inc

Acee Lindem
Redback Networks

December 2003

OSPF Tunnel Adjacency
draft-mirtorabi-ospf-tunnel-adjacency-01.txt

Status of this Memo

Abstract

The OSPF specification requires that intra-area paths are always preferred over inter-area paths, regardless of the path's cost. In some situations this can lead to an inefficient usage of network resources. This document describes a solution that helps to address this problem by creating adjacencies through backbone area that belong to non-backbone areas.

1. Motivation

There could be a requirement to prefer an inter-area path over an intra-area path. For example, in order to utilize a high bandwidth

backbone path to transit the intra-area traffic from a non-backbone area. The current OSPF specification does not provide any generic mechanism to achieve this. In some situations, Virtual Links (VLs) can help. However, there are some restrictions associated with VLs:

a) Transit area must be a non-backbone regular area

b) VLs prevent summarization of backbone prefixes into their associated transit area

c) VLs cannot be configured through Stub or NSSA [2] areas

2. Proposed Solution

The Tunnel Adjacency (TA) proposal uses a concept similar to virtual links by forming an adjacency (possibly multihop) between two ABRs through a transit area. However, TAs can be configured for any non-backbone area with the backbone as the transit area.

Tunnel Adjacencies operate similar to VLs in adjacency establishment, sending unicast OSPF packets, and datbase synchronization. Data packet forwarding between the endpoint ABRs is different from VLs in that the packets are tunneled if the TA's path spans multiple hops. This removes the requirement for routers internal to the transit area to have the TA area's unsummarised intra-area routes. The rest of this document describes the TA specification.

3. Bringing up the tunnel adjacency

TAs are configured between two ABRs attached to the backbone. Similiar to virtual links, TAs are identified by the Router ID of the endpoint. Once a tunnel adjacency for a given area is configured and an intra-area path exists between the two ABRs through the backbone an adjacency can be formed as specified in OSPF [1].

The interface MTU should be set to 0 in Database Description packets sent over TAs as is done with virtual links. TAs can be configured as a Demand Circuits (DC) in order to reduce Hello exchange and periodic LSA flooding.

4. Tunnel adjacency encapsulation

User traffic routed based on the presence of the TA will be encapsulated on the TA endpoints in the following way:

a) If both ends of the TA are directly connected to the same network and the best intra-area path over the backbone is via this direct network connection, no additional encapsulation is needed.

b) Otherwise, the traffic is further encapsulated (tunneled) and sent directly to the TA endpoint. The encapsulation type is left to the implementation and different encapsulation types could be specified through configuration. However, in order to have interoperability between vendors all implementations should support GRE encapsulation [3].

5. Advertising tunnel adjacency

TAs are announced as unnumbered point-to-point links. Once a router's TA reaches the FULL state a type 1 link will be added to the Router LSA with:

```
Link ID   = Remote's Router ID
Link Data = Router's own IP address associated with TA
Cost      = Intra-area cost to the TA endpoint via the backbone area
            or the configured cost
```

The IP address specified in the link data is computed during the routing table build process for the backbone.

6. Tunnel adjacency interface data structure

The TA interface data structure is the same as specified in section 9 of OSPF [1]. An OSPF interface data structure is created for each configured tunnel adjacency. The cost of the TA is configurable allowing a traffic path to be selected independent of the intra-area path cost. The default cost is equal to the intra-area cost to reach the TA endpoint through the backbone.

Topologically, a TA is identical to an unnumbered point-to-point interface.

7. Tunnel adjacency interface FSM

The TA Interface FSM is the same as specified in section 9.3 of OSPF [1]. The InterfaceUp event for TA interfaces is generated once the remote end of the TA becomes reachable through the backbone via an intra-area path.

Similiarly, the InterfaceDown event is generated for TA interfaces when the remote end of the TA is no longer reachable through the backbone via an intra-area path.

8. Tunnel adjacency neighbor data structure

The TA neighbor data structure is identical to the neighbor data structure for standard OSPF adjacencies as specified in section 10 of OSPF [1].

9. Tunnel adjacency neighbor FSM

The TA neighbor FSM is identical to the neighbor FSM for a standard OSPF point-to-point adjacency as specified in section 10.3 of OSPF [1].

10. Tunnel adjacency OSPF control packet processing

OSPF control packet processing is specified in OSPF [1] section 8. This section is modified as follow:

[...]

The IP source address should be set to the IP address of the sending interface. Interfaces to unnumbered point-to-point networks have no associated IP address. On these interfaces, the IP source should be set to any of the other IP addresses belonging to the router. For this reason, there must be at least one IP address assigned to the router. Note that, for most purposes, virtual links and tunnel adjacency act precisely the same as unnumbered point-to-point networks.

However, each virtual link or tunnel adjacency does have an IP interface address belonging to a transit area or backbone (discovered during the routing table build process) which is used as the IP source when sending packets over the virtual link or tunnel

adjacency. If there is not at least one IP address belonging to Transit area or the backbone and a virtual link or TA is configured, a router could advertise any of its attached IP address as a stub link (Link ID set to the router's own IP interface address, Link Data set to the mask 0xffffffff) to the transit area.

[...]

Receiving protocol packets as described in 8.2 is changed as follow:

Next, the OSPF packet header is verified. The fields specified in the header must match those configured for the receiving interface. If they do not, the packet should be discarded:

o The version number field must specify protocol version 2.

o The Area ID found in the OSPF header must be verified. If all of the following cases fail, the packet should be discarded. The Area ID specified in the header must either:

(1) Match the Area ID of the receiving interface. In this case, the packet has been sent over a single hop. Therefore, the packet's IP source address is required to be on the same network as the receiving interface. This can be verified by

comparing the packet's IP source address to the interface's IP address, after masking both addresses with the interface mask. This comparison should not be performed on point-to-point networks. On point-to-point networks, the interface addresses of each end of the link are assigned independently, if they are assigned at all.

(2) Indicate a non-backbone area. In this case, the packet has been sent over a tunnel adjacency. The receiving router must be an area border router, and the Router ID specified in the packet (the source router) must be the other end of a configured tunnel adjacency. The receiving interface must also attach to the backbone. If all of these checks succeed, the packet is accepted and is from now on associated with the tunnel adjacency for that area.

(3) Indicate the backbone. In this case, the packet has been sent over a virtual link. The receiving router must be an

area border router, and the Router ID specified in the packet (the source router) must be the other end of a configured virtual link. The receiving interface must also attach to the virtual link's configured Transit area. If all of these checks succeed, the packet is accepted and is from now on associated with the virtual link (and the backbone area).

- Packets whose IP destination is AllDRouters should only be accepted if the state of the receiving interface is DR or Backup (see Section 9.1).

[...]

11. Tunnel adjacency next hop calculation

The next-hop to reach the TA endpoint is equal to the next-hop associated with the TA endpoint via the backbone area.

Data packet forwarding between the two ABRs is different from a VL in that the packets are tunneled if the TA path spans multiple hops. This removes the requirement for routers internal to the backbone area to have the TA area's unsummarised intra-area routes.

12. Virtual link - tunnel adjacency comparison

Virtual links are part of area 0 and must transit through a regular non-backbone area and are configured to avoid backbone partitioning. Conversely, tunnel adjacencies can be part of any type non-backbone area and use the backbone as a transit area. Hence, TAs complement

virtual links and address the following requirements (refer to the applications section for more information):

a) Preference of a high speed backbone area link for non-backbone traffic.

b) On-demand (automatic) partition repair for non-backbone areas.

c) Multiple TAs could be configured over a backbone path, each (TA) belonging to a different area in order to provide an intra-area path for each area and saving the cost of an additional link.

An additional advantage is the cost of TA is configurable allowing a traffic path to be selected independent of the intra-area path cost. This allows an alternate traffic path to be forced.

13. Applications

In this section we give a few examples of how TAs can be used.

13.1 Prefer Inter-area Path over intra-area Path

It is a common requirement that users would like to prefer the high bandwidth part of the backbone for traffic that can be strictly routed inside the non-backbone area.

Consider the following topology:

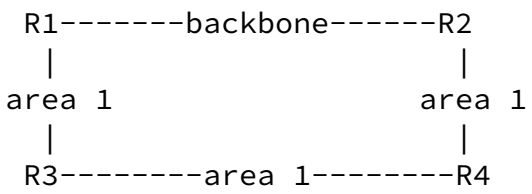

Fig.1

The backbone link between R1 and R2 is a high speed link and could be used to forward part of the area 1's traffic between R1 and R2. In the current OSPF specification, intra-area paths are preferred over inter-area paths. As a result, R1 will always route traffic to R4 through area 1 over the lower speed links. Even to reach networks connected to R2 that belong to area 1, R1 will use the intra-area path over area 1.

By configuring a TA between R1 and R2, a P2P link will be advertised into area 1 making the TA a topological part of area 1 with a lower cost than than the low speed links.

Note that the above scenario can not be solved using a VL since the link between R1 and R2 belongs to the backbone area and it is not desirable to move this backbone link in a non-backbone area.

It should also be noted that the connection between R1 and R2 in the backbone area could be multiple hops away. In other words, TAs are not limited to directly connected topologies.

13.2 On-demand partition avoidance for summarized non-backbone area

In general when a non-backbone area is partitioned there is no need for partition repair as intra-area routes will be replaced by inter-area routes for the partiationed area. However, this is not true if the area is summarized into the backbone. Consider the following topology:

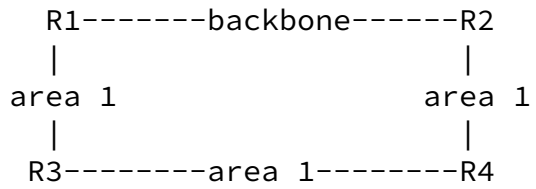

Fig.3

R1 and R2 are summarizing area 1 into the backbone area. When area 1 becomes partitioned due the R3-R4 going down, R1 and R2 continue to summarize area 1 into the backbone area. This can lead to blackholing of the traffic. The reason is that after the area partitioning, R1 or R2 will only have knowledge of their attached area partitions. When R1 or R2 receives a packet that does not belong to its attached partitioned area (as a result of advertising a summary) the packet will be discarded.

Note that R1 and R2 will install a discard route for the configured summary range. If the destination doesn't match an intra-area route in R1 or R2 area partition, the destination will match on the less specific discard route.

By configuring an on-demand TA for area 1 through the backbone, R1 and R2 will establish an adjacency if area 1 becomes partitioned.

When a TA is configured between the two ABRs, a configuration option (automatic) will be used to not start sending Hellos unless the other ABR is not reachable via area 1.

The cost of on-demand TA should automatically be set to maximum cost LSInfinity (16-bit value 0xFFFF). The reason to set the cost of TA to 0xFFFF is to make it easier to detect that the area is no longer partitioned. During the SPF, only the shortest path to the remote end of the TA is discovered and making the TA cost the maximum reachable cost will allow partition repair to be detected as a natural side effect of the intra-area SPF calculation.

13.3 Saving additional link between ABRs in a Hub and Spoke environment

Consider the typical hub and spoke topology in figure 4.

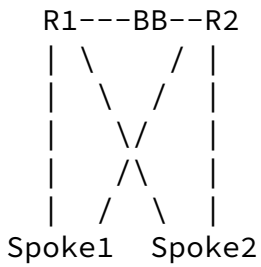

Fig.4

Only two spokes are represented in figure 4, but in general we may have N spokes similar to Spoke1.

R1 and R2 are ABRs and can be multiple hops away over the backbone area (BB). Further, the ABRs are summarizing IP prefixes from all the attached areas into the backbone.

Case 1: Spoke1 and Spoke2 are in different area

Since both R1 and R2 are summarizing, there is a need for a link between R1 and R2 in each connected area. This is to guarantee an alternative path when the link between a spoke and hub becomes unavailable.

For example, imagine a network X advertised by Spoke1 and summarized by both R1 and R2. Later the link between R1 and Spoke1 goes down. When a packet arrives at R1 to be forwarded to Spoke1, R1 cannot send the packet to Spoke1 since the link is not available.
Since R1 is summarizing this route it may have installed a discard route for summarized range (here we assume the range is still 'active', as there may be other spokes in the same area as Spoke1 that are still attached to R1 and advertising prefixes that fall in the same range as X). Hence, R1 will not use an inter-area path over R2. A link between R1

and R2 inside the same area as the link between R1 and Spoke1 would prevent this problem.

Case 2: Spoke1 and Spoke2 are in the same area

Link between R1 and Spoke1 is broken. The path from R1 to Spoke1 is R1-Spoke2-R2-Spoke1 instead of R1-R2-Spoke1.

In general, for N areas being attached to the hub routers, there is a need for N links between hub routers. Multiple TAs could be used through the backbone between the hub routers to avoid using multiple physical links between ABRs (each belonging to a different non-backbone area)

14. Tunnel adjacency parameters

Tunnel adjacencies can be configured in a non-backbone areas between area border routers having at least one backbone conection. A tunnel adjacency is defined by the following configurable parameters:

- The Router ID of the Tunnel adjacency's endpoint.
- The area where the tunnel adjacency resides.

Optionally, the following parameters are configurable:

- The cost of the tunnel adjacency which will override intra-area cost between the two TA endpoints.
- The encapsulation type to be used when the two TA endpoints are not directly connected. The default is GRE.
- The 'automatic' option used for on on-demand partition repair.

15. Tunnel adjacency in OSPFv3

All mechanisms described in this document for OSPFv2 applies also to OSPFv3 [4] with the following exceptions:

- The IPv6 interface address of a tunnel adjacency must be an IPv6 address having a global scope, instead of the link-local addresses

used by other interface types. This address is used as the IPv6 source for OSPF protocol packets sent over the tunnel adjacency.

- Likewise, the tunnel adjacency neighbor's IPv6 address is an IPv6 address with global scope.

- Like all other IPv6 OSPF interfaces, tunnel adjacency are assigned unique (within the router) Interface IDs. These are advertised in Hellos sent over the tunnel adjacency and specified for links in the router's router-LSAs.

16. Compatibility issues

All mechanisms described in this document are backward-compatible with standard OSPF implementations.

17. Security

Tunnel adjacencies as specified in this document do not raise any security issues that are not already covered in [1].

18. Acknowledgments

Authors would like to thank Abhay Roy, Liem Nguyen, Pat Murphy, and Alex Zinin for their comments on the document.

19. Reference

[1] Moy, J., "OSPF Version 2", STD 54, RFC 2328, April 1998.

[2] Murphy, P., "The OSPF Not-So-Stubby Area (NSSA) Option", RFC 3101, January 2003.

[3] D. Farinacci, T. Li, S. Hanks, D. Meyer and P. Traina, "Generic Routing Encapsulation (GRE), RFC 2784, March 2000.

[4] R. Coltun, D. Ferguson and J. Moy, "OSPF for IPv6", RFC 2740, December 1999.

20. Authors' address

Sina Mirtorabi
Cisco Systems

225 West Tasman drive
San Jose, CA 95134
E-mail: sina@cisco.com

Peter Psenak
Cisco Systems
Parc Pegasus,
De Kleetlaan 6A
1831 Diegem
Belgium
E-mail: ppsenak@cisco.com

Acee Lindem
Redback Networks
102 Carric Bend Court
Cary, NC 27519
Email: acee@redback.com