            Using a Link State Advertisement (LSA) Options Bit to
         Prevent Looping in BGP/MPLS IP Virtual Private Networks (VPNs)

Status of This Memo

Copyright Notice

Abstract

   This document specifies a procedure that deals with a particular
   issue that may arise when a Service Provider (SP) provides "BGP/MPLS
   IP VPN" service to a customer and the customer uses OSPFv2 to
   advertise its routes to the SP.  In this situation, a Customer Edge
   (CE) Router and a Provider Edge (PE) Router are OSPF peers, and
   customer routes are sent via OSPFv2 from the CE to the PE.  The
   customer routes are converted into BGP routes, and BGP carries them
   across the backbone to other PE routers.  The routes are then
   converted back to OSPF routes sent via OSPF to other CE routers.  As
   a result of this conversion, some of the information needed to
   prevent loops may be lost.  A procedure is needed to ensure that once
   a route is sent from a PE to a CE, the route will be ignored by any
   PE that receives it back from a CE.  This document specifies the
   necessary procedure, using one of the options bits in the LSA (Link
   State Advertisements) to indicate that an LSA has already been
   forwarded by a PE and should be ignored by any other PEs that see it.

Table of Contents

1.  Introduction

   [VPN] describes a method by which a Service Provider (SP) can use its
   IP backbone to provide an "IP VPN" service to customers.  In that
   sort of service, a customer's edge devices (CE devices) are connected
   to the provider's edge routers (PE routers).  Each CE device is in a
   single Virtual Private Network (VPN).  Each PE device may attach to
   multiple CEs of the same or of different VPNs.  A VPN thus consists
   of a set of "network segments" connected by the SP's backbone.

   A CE exchanges routes with a PE, using a routing protocol to which
   the customer and the SP jointly agree.  The PE runs that routing
   protocol's decision process (i.e., it performs the routing
   computation) to determine the set of IP address prefixes for which
   the following two conditions hold:

      -  Each address prefix in the set can be reached via that CE.

      -  The path from that CE to each such address prefix does NOT
         include the SP backbone (i.e., it does not include any PE
         routers).

   The PE routers that attach to a particular VPN redistribute routes to
   these address prefixes into BGP, so that they can use BGP to
   distribute the VPN's routes to each other.  BGP carries these routes
   in the "VPN-IPv4 address family", so that they are distinct from
   ordinary Internet routes.  The VPN-IPv4 address family also extends
   the IP addresses on the left so that address prefixes from two
   different VPNs are always distinct to BGP, even if both VPNs use the
   same piece of the private RFC 1918 address space.  Thus, routes from
   different VPNs can be carried by a single BGP instance and can be
   stored in a common BGP table without fear of conflict.

   If a PE router receives a particular VPN-IPv4 route via BGP, and if
   that PE is attached to a CE in the VPN to which the route belongs,
   then BGP's decision process may install that route in the BGP route
   table.  If so, the PE translates the route back into an IP route and

redistributes it to the routing protocol that is running on the link
to that CE.

This methodology provides a "peer model".  CE routers peer with PE
routers, but CE routers at different sites do not peer with each
other.

If a VPN uses OSPFv2 as its internal routing protocol, it is not
necessarily the case that the CE routers of that VPN use OSPFv2 to
peer with the PE routers.  Each site in a VPN can use OSPFv2 as its
intra-site routing protocol while using BGP or RIP (for example) to
distribute routes to a PE router.  However, it is certainly
convenient when OSPFv2 is being used intra-site to use it on the PE-
CE link as well, and [VPN] explicitly allows this.  In this case, a
PE will run a separate instance of OSPFv2 for each VPN that is
attached to the PE; the PE will in general have multiple VPN-specific
OSPFv2 routing tables.

When OSPFv2 is used on a PE-CE link that belongs to a particular VPN,
the PE router must redistribute to that VPN's OSPFv2 instance certain
routes that have been installed in the BGP routing table.  Similarly,
a PE router must redistribute to BGP routes that have been installed
in the VPN-specific OSPF routing tables.  Procedures for this are
specified in [VPN-OSPF].

The routes that are redistributed from BGP to OSPFv2 are advertised
in LSAs that are originated by the PE.  The PE acts as an OSPF border
router, advertising some of these routes in AS-external LSAs, and
some in summary LSAs, as specified in [VPN-OSPF].

Similarly, when a PE router receives an LSA from a CE router, it runs
the OSPF routing computation.  Any route that gets installed in the
OSPF routing table must be translated into a VPN-IPv4 route and then
redistributed into BGP.  BGP will then distribute these routes to the
other PE routers.

2.  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 RFC 2119.

3.  Information Loss and Loops

A PE, say PE1, may learn a route to a particular VPN-IPv4 address
prefix via BGP.  This may cause it to generate a summary LSA or an
AS-external LSA in which it reports that address prefix.  This LSA
may then be distributed to a particular CE, say CE1.  The LSA may

then be distributed throughout a particular OSPF area, reaching
another CE, say CE2.  CE2 may then distribute the LSA to another PE,
say PE2.

As stated in the previous section, PE2 must run the OSPF routing
computation to determine whether a particular address prefix,
reported in an LSA from CE2, is reachable from CE2 via a path that
does not include any PE router.  Unfortunately, there is no standard
way to do this.  The OSPFv2 LSAs do not necessarily carry the
information needed to enable PE2 to determine that the path to
address prefix X in a particular LSA from CE2 is actually a path that
includes, say PE1.  If PE2 then leaks X into BGP as a VPN-IPv4 route,
then PE2 is violating one of the constraints for loop-freedom in BGP;
viz., that routes learned from a particular BGP domain are not
redistributed back into that BGP domain.  This could cause a routing
loop to be created.

It is therefore necessary to have a means by which an LSA may carry
the information that a particular address prefix has been learned
from a PE router.  Any PE router that receives an LSA with this
information would omit the information in this LSA from its OSPF
routing computation, and thus it would not leak the information back
into BGP.

When a PE generates an AS-external LSA, it could use a distinct tag
value to indicate that the LSA is carrying information about an
address prefix for whom the path includes a PE router.  However, this
method is not available in the case where the PE generates a Summary
LSA.  Per [VPN-OSPF], each PE router must function as an OSPF area 0
router.  If the PE-CE link is an area 0 link, then it is possible for
the PE to receive, over that link, a summary LSA that originated at
another PE router.  Thus, we need some way of marking a summary LSA
to indicate that it is carrying information about a path via a PE
router.

4.  Using the LSA Options to Prevent Loops

The high-order bit of the LSA Options field (a previously unused bit)
is used to solve the problem described in the previous section.  We
refer to this bit as the DN bit.  When a type 3, 5, or 7 LSA is sent
from a PE to a CE, the DN bit MUST be set.  The DN bit MUST be clear
in all other LSA types.

```
              +-------------------------------------+
              | DN | * | DC | EA | N/P | MC | E | * |
              +-------------------------------------+
```

                    Options Field with DN Bit
                     (RFC 2328, Section A.2)

   When the PE receives, from a CE router, a type 3, 5, or 7 LSA with
   the DN bit set, the information from that LSA MUST NOT be used during
   the OSPF route calculation.  As a result, the LSA is not translated
   into a BGP route.  The DN bit MUST be ignored in all other LSA types.

   This prevents routes learned via BGP from being redistributed to BGP.
   (This restriction is analogous to the usual OSPF restriction that
   inter-area routes that are learned from area 0 are not passed back to
   area 0.)

   Note that the DN bit has no other effect on LSA handling.  In
   particular, an LSA with the DN bit set will be put in the topological
   database, aged, flooded, etc., just as if DN were not set.

5.  Security Considerations

   An attacker may cause the DN bit to be set, in an LSA traveling from
   CE to PE, when the DN bit should really be clear.  This may cause the
   address prefixes mentioned in that LSA to be unreachable from other
   sites of the VPN.  Similarly, an attacker may cause the DN bit to be
   clear, in an LSA traveling in either direction, when the DN bit
   should really be set.  This may cause routing loops for traffic that
   is destined to the address prefixes mentioned in that LSA.

   These possibilities may be eliminated by using cryptographic
   authentication as specified in Section D of [OSPFv2].

6.  Acknowledgements

   The idea of using the high-order options bit for this purpose is due
   to Derek Yeung.  Thanks to Yakov Rekhter for his contribution to this
   work.  We also wish to thank Acee Lindem for his helpful comments.

7.  Normative References

   [OSPFv2]    Postel, J., "Suggested Telnet Protocol Changes", RFC 328,
               April 1972.

   [VPN]       Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private
               Networks (VPNs)", RFC 4364, February 2006.

   [VPN-OSPF] Rosen, E., Psenak, P., and P. Pillay-Esnault, "OSPF as the
               Provider/Customer Edge Protocol for BGP/MPLS IP Virtual
               Private Networks (VPNs)", RFC 4577, June 2006.

Authors' Addresses

   Eric C. Rosen
   Cisco Systems, Inc.
   1414 Massachusetts Avenue
   Boxborough, MA 01719

   EMail: erosen@cisco.com


   Peter Psenak
   Cisco Systems
   BA Business Center, 9th Floor
   Plynarenska 1
   Bratislava 82109
   Slovakia

   EMail: ppsenak@cisco.com


   Padma Pillay-Esnault
   Cisco Systems
   3750 Cisco Way
   San Jose, CA 95134

   EMail: ppe@cisco.com