L3VPN WG
draft-ietf-l3vpn-bgpvpn-auto-09.txt
INFORMATIONAL
Expiration Date: November 2007

Hamid Ould-Brahim
Nortel Networks

Eric C. Rosen
Cisco Systems

Yakov Rekhter
Juniper Networks

(Editors)

April 2007

Using BGP as an Auto-Discovery Mechanism for VR-based Layer-3 VPNs

Status of this Memo

Abstract

In any provider-based VPN scheme, the Provider Edge (PE) devices attached to a common VPN must exchange certain information as a prerequisite to establish VPN-specific connectivity. The main purpose of an auto-discovery mechanism is to enable a PE to dynamically discover the set of remote PEs having VPN members in

common. The auto-discovery mechanism proceeds by having a PE advertises to other PEs, at a minimum, its own IP address and the list of VPN members configured on that PE. Once that information is received the remote PEs will then identify the list of VPN sites members of the same VPN, and use the information carried within the discovery mechanism to establish VPN connectivity. This draft defines a BGP based auto-discovery mechanism for Virtual Router-based layer-3 VPNs. This mechanism is based on the approach used by BGP/MPLS-IP-VPN for distributing VPN routing information within the service provider(s).

Changes from 07 version (DELETE THIS WHEN IT BECOMES RFC)

- Updated the IANA section to reflect the review from IANA
- Nits from Harald's feedback.

1. Introduction

In any provider-based VPN scheme, the Provider Edge (PE) devices attached to a common VPN must exchange certain information as a prerequisite to establish VPN-specific connectivity. An auto-discovery mechanism allows a PE to dynamically discover the set of remote PEs having VPN members in common. The auto-discovery mechanism proceeds by having a PE advertises to other PEs, at a minimum, its own IP address and the list of VPN sites configured on that PE. Once that information is received the remote PEs will then identify the list of VPN sites member of the same VPN with the advertising PE, and use the information carried within the discovery mechanism to establish VPN connectivity.

The purpose of this draft is to define a BGP based auto-discovery mechanism for VR-based VPNs [VPN-VR] solution. This mechanism is based on the approach used by [BGP/MPLS-IP-VPN] for distributing VPN routing information within the service provider(s).

Virtual router (VR) addresses must be exchanged, along with the information needed to enable the PEs to determine which VRs are in the same VPN ("membership"), and which of those VRs are to have VPN connectivity ("topology"). Once the VRs are reachable through the tunnels, routes ("reachability") are then exchanged by running existing routing protocols per VPN basis.

The BGP-4 multiprotocol extensions are used to carry various information about VR-based VPNs. VPN-specific information associated with the NLRI is encoded either as attributes of the NLRI, or as part of the NLRI itself, or both.

2. Provider-Provisioned VPN Reference Model

When using BGP as an auto-discovery mechanism, VR-based l3vpns are using a network reference model as illustrated in figure 1.

```
                  PE                          PE
            +---------------+            +---------------+
+--------+  | +-----------+ |            | +-----------+ | +--------+
|  VPN-A |  | |  VPN-A    | |            | |  VPN-A    | | |  VPN-A |
|  Sites |--| |Database /| |  BGP route | | Database/| |-|  sites |
+--------+  | |Processing| |<----------->| |Processing| | +--------+
            | +-----------+ | Distribution| +-----------+ |
            |               |            |               |
+--------+  | +-----------+ |            | +-----------+ | +--------+
| VPN-B  |  | |  VPN-B    | |  --------  | |   VPN-B   | | |  VPN-B |
| Sites  |--| |Database /| |-(Backbones)-| | Database/| |-|  sites |
+--------+  | |Processing| |  --------  | |Processing| | +--------+
            | +-----------+ |            | +-----------+ |
            |               |            |               |
+--------+  | +-----------+ |            | +-----------+ | +--------+
| VPN-C  |  | |  VPN-C    | |            | |   VPN-C   | | |  VPN-C |
| Sites  |--| |Database /| |            | | Database/| |-|  sites |
+--------+  | |Processing| |            | |Processing| | +--------+
            | +-----------+ |            | +-----------+ |
            +---------------+            +---------------+
```

Figure 1: Network based VPN Reference Model

It is assumed that the PEs can use BGP to distribute information to each other. This may be via direct IBGP peering, via direct EBGP peering, via multihop BGP peering, through intermediaries such as Route Reflectors, through a chain of intermediate BGP connections, etc.

3. Carrying VR-based VPN information in BGP

The BGP-4 multiprotocol extensions are used to carry various information about VPNs. VPN-specific information associated with the NLRI is encoded either as attributes of the NLRI, or as part of the NLRI itself, or both. The addressing information in the NLRI field is ALWAYS within the VPN address space, and therefore MUST be unique within the VPN. The address specified in the BGP next hop attribute, on the other hand, is in the service provider addressing space.

The NLRI is a VPN-IP address or a labeled VPN-IP address. The NLRI address prefix is an address of one of the virtual routers configured on the PE. That address is used by the VRs to establish routing adjacencies and tunnel to each other [VPN-VR].

4. Interpretation of VPN Information in the VR Model

4.1 Membership Discovery

The VPN-ID format as defined in [RFC-2685] is used to identify a VPN. All virtual routers that are members of a specific VPN share the same VPN-ID. A VPN-ID is carried in the NLRI to make addresses of VRs globally unique. Making these addresses globally unique is necessary if one uses BGP for VRs' auto-discovery.

4.2 Encoding of the VPN-ID in the NLRI

For the virtual router model, the VPN-ID is carried within the route distinguisher (RD) field. In order to hold the 7-bytes VPN-ID, the first byte of RD type field is used to indicate the existence of the VPN-ID format. A value of 0x80 in the first byte of RD's type field indicates that the RD field is carrying the VPN-ID format. In this case, the type field range 0x8000-0x80ff will be reserved for the virtual router case.

4.3 VPN-ID Extended Community

A new extended community is used to carry the VPN-ID format. This attribute is transitive across the Autonomous system boundary. The type field of the VPN-ID extended community is of regular type to be assigned by IANA [BGP-COMM]. The remaining 7 bytes hold the VPN-ID value field as per [RFC-2685]. The BGP UPDATE message will carry information for a single VPN. It is the VPN-ID Extended Community, or more precisely route filtering based on the Extended Community that allows one VR to find out about other VRs in the same VPN.

4.4 VPN Topology Information

A new extended community is used to indicate different VPN topology values. This attribute is transitive across the Autonomous system boundary. The value of the type field for extended type is assigned by IANA. The first two bytes of the value field (of the remaining 6 bytes) are reserved. The actual topology values are carried within the remaining four bytes. The following topology values are defined:

Value	Topology Type
1	"Hub"
2	"Spoke"
3	"Mesh"

Arbitrary values can also be used to allow specific topologies to be constructed.

In a hub and spoke topology, spoke VRs (i.e., PE having VRs as spokes within the VPN) will advertise their BGP information with VPN topology extended community with value of "2". Spoke VRs will only be allowed to connect to hub VRs and therefore spoke VR-based PEs will just import VPN information from BGP that is set of "1". Hub sites can connect to both hub and spoke sites (i.e., Hub VRs can import VPN topology of both values "1", "2", or "3". In a mesh topology, mesh sites connect to each other, each VR will advertise VPN topology information of "3".

Furthermore, in the presence of both hub and spoke and mesh topologies within the same VPN, mesh sites can as well connect to hub sites and vice versa.

5. Tunnel Discovery

Layer-3 VPNs must be implemented through some form of tunneling mechanism, where the packet formats and/or the addressing used within the VPN can be unrelated to that used to route the tunneled packets across the backbone. There are numerous tunneling mechanisms that can be used by a network based VPN (e.g., IP/IP [RFC-2003], GRE tunnels [RFC-1701], IPSec [RFC-2401], and MPLS tunnels [RFC-3031]). Each of these tunnels allows for opaque transport of frames as packet payload across the backbone, with forwarding disjoint from the address fields of the encapsulated packets. A provider edge router may terminate multiple types of tunnels and forward packets between these tunnels and other network interfaces in different ways. BGP can be used to carry tunnel endpoint addresses between edge routers.

The BGP next hop will carry the service provider tunnel endpoint address. As an example, if IPSec is used as tunneling mechanism, the IPSec tunnel remote address will be discovered through BGP, and the actual tunnel establishment is achieved through IPSec signaling protocol.

When MPLS tunneling is used, the label carried in the NLRI field is associated with an address of a VR, where the address is carried in the NLRI and is encoded as a VPN-IP address.

The auto-discovery mechanism should convey minimum information for the tunnels to be setup. The means of distributing multiplexors must be defined either via some sort of tunnel-protocol-specific signaling mechanism, or via additional information carried by the auto-discovery protocol. That information may or may not be used directly within the specific signaling protocol. On one end of the spectrum, the combination of IP address (such as BGP next hop and IP address carried within the NLRI) and the label and/or VPN-ID provides sufficient information for a PE to setup per VPN tunnels or shared tunnels per set of VPNs. On another end of the spectrum additional specific tunnel related information can be carried within the discovery process if needed.

6. Scalability Considerations

In this section, we briefly summarize the main characteristics of our model with respect to scalability.

Recall that the Service Provider network consists of (a) PE routers, (b) BGP Route Reflectors, (c) P routers (which are neither PE routers nor Route Reflectors), and, in the case of multi-provider VPNs, (d) ASBRs.

A PE router, unless it is a Route Reflector should not retain VPN-related information unless it has at least one VPN with an Import Target identical to one of the VPN-related information Route Target attributes. Inbound filtering should be used to cause such information to be discarded. If a new Import Target is later added to one of the PE's VPNs (a "VPN Join" operation), it must then acquire the VPN-related information it may previously have discarded.

This can be done using the refresh mechanism described in [BGP-RFSH].

The outbound route filtering mechanism of [BGP-ORF], [BGP-CONS] can also be used to advantage to make the filtering more dynamic.

Similarly, if a particular Import Target is no longer present in any of a PE's VPNs (as a result of one or more "VPN Prune" operations), the PE may discard all VPN-related information which, as a result, no longer have any of the PE's VPN's Import Targets as one of their Route Target Attributes.

Note that VPN Join and Prune operations are non-disruptive, and do not require any BGP connections to be brought down, as long as the refresh mechanism of [BGP-RFSH] is used.

As a result of these distribution rules, no one PE ever needs to maintain all routes for all VPNs; this is an important scalability consideration.

Route reflectors can be partitioned among VPNs so that each partition carries routes for only a subset of the VPNs supported by the Service Provider. Thus no single route reflector is required to maintain VPN-related information for all VPNs.

For inter-provider VPNs, if multi-hop EBGP is used, then the ASBRs need not maintain and distribute VPN-related information at all.

P routers do not maintain any VPN-related information. In order to properly forward VPN traffic, the P routers need only maintain routes to the PE routers and the ASBRs.

As a result, no single component within the Service Provider network has to maintain all the VPN-related information for all the VPNs. So the total capacity of the network to support increasing numbers of VPNs is not limited by the capacity of any individual component.

An important consideration to remember is that one may have any number of INDEPENDENT BGP systems carrying VPN-related information. This is unlike the case of the Internet, where the Internet BGP system must carry all the Internet routes. Thus one significant (but perhaps subtle) distinction between the use of BGP for the Internet routing and the use of BGP for distributing VPN-related information, as described in this document is that the former is not amenable to partition, while the latter is.

7. Security Considerations

This document describes a BGP-based auto-discovery mechanism which enables a PE router that attaches to a particular VPN to discover the set of other PE routers that attach to the same VPN. Each PE router that is attached to a given VPN uses BGP to advertise that fact. Other PE routers which attach to the same VPN receive these BGP advertisements. This allows that set of PE routers to discover each other. Note that a PE will not always receive these advertisements directly from the remote PEs; the advertisements may be received from "intermediate" BGP speakers.

It is of critical importance that a particular PE should not be "discovered" to be attached to a particular VPN unless that PE really is attached to that VPN, and indeed is properly authorized to be attached to that VPN. If any arbitrary node on the Internet could start sending these BGP advertisements, and if those advertisements were able to reach the PE routers, and if the PE routers accepted those advertisements, then anyone could add any site to any VPN. Thus the auto-discovery procedures described here presuppose that a particular PE trusts its BGP peers to be who they appear to be, and further that it can trusts those peers to be properly securing their local attachments. (That is, a PE must trust that its peers are attached to, and are authorized to be attached to, the VPNs to which they claim to be attached.).

If a particular remote PE is a BGP peer of the local PE, then the BGP authentication procedures of RFC 2385 can be used to ensure that the remote PE is who it claims to be, i.e., that it is a PE that is trusted.

If a particular remote PE is not a BGP peer of the local PE, then the information it is advertising is being distributed to the local

PE through a chain of BGP speakers. The local PE must trust that its peers only accept information from peers that they trust in turn, and this trust relation must be transitive. BGP does not provide a way to determine that any particular piece of received information originated from a BGP speaker that was authorized to advertise that particular piece of information. Hence the procedures of this document should be used only in environments where adequate trust relationships exist among the BGP speakers.

Some of the VPN schemes which may use the procedures of this document can be made robust to failures of these trust relationships. That is, it may be possible to keep the VPNs secure even if the auto-discovery procedures are not secure. For example, a VPN based on the VR model can use IPsec tunnels for transmitting data and routing control packets between PE routers. An illegitimate PE router which is discovered via BGP will not have the shared secret which makes it possible to set up the IPsec tunnel, and so will not be able to join the VPN. Similarly, [IP-GRE] describes procedures for using IPsec tunnels to secure VPNs based on the [BGP/MPLS-IP-VPN] model. The details for using IPsec to secure a particular sort of VPN depend on that sort of VPN and so are out of scope of the current document.

8. IANA Considerations

IANA has assigned new extended community <TBD> for Topology values for VR-based L3VPN solution.

IANA has assigned new extended community <TBD> for carrying VPN-ID format based on RFC2685 format.

IANA has assigned new SAFI number <TBD> for indicating that the NLRI is carrying information for VR for labeled prefixes.

SAFI number "140" for indicating that the NLRI is carrying information for VR for non-labeled prefixes.

9. Use of BGP Capability Advertisement

A BGP speaker that uses VPN information as described in this document with multiprotocol extensions should use the Capability Advertisement procedures [RFC-3392] to determine whether the speaker could use Multiprotocol Extensions with a particular peer.

10. Acknowledgement

The authors would like to acknowledge Benson Schliesser, and Thomas Narten for the constructive and fruitful comments.

11. Normative References

[BGP-COMM] Ramachandra, Tappan, et al., "BGP Extended Communities Attribute", RFC4360.

[BGP-MP] Bates, Chandra, Katz, and Rekhter, "Multiprotocol Extensions for BGP4", February 1998, RFC 2283.

[RFC-3107] Rekhter Y, Rosen E., "Carrying Label Information in BGP4", January 2000, RFC3107.

[BGP/MPLS-IP-VPN] Rosen E., et al, "BGP/MPLS VPNs", RFC4364.

[RFC-2685] Fox B., et al, "Virtual Private Networks Identifier", RFC 2685, September 1999.

[RFC-3392] Chandra, R., et al., "Capabilities Advertisement with BGP-4", RFC3392, May 2002.

[VPN-VR] Knight, P., Ould-Brahim H., Gleeson, B., "Network based IP VPN Architecture using Virtual Routers", Work in progress.

12. Informative References

[RFC-1701] Hanks, S., Li, T., Farinacci, D. and P. Traina, "Generic Routing Encapsulation (GRE)", RFC 1701, October 1994.

[RFC-2003] Perkins, C., "IP Encapsulation within IP", RFC 2003, October 1996.

[RFC-2026] Bradner, S., "The Internet Standards Process -- Revision 3", RFC 2026, October 1996.

[RFC-2401] Kent S., Atkinson R., "Security Architecture for the Internet Protocol", RFC 2401, November 1998.

[RFC-2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", RFC 2119, March 1997.

[IP-GRE] Rosen, E., et al., "Use of PE-PE GRE or IP in BGP/MPLS IP Virtual Private Networks", draft-ietf-l3vpn-gre-ip-2547-03.txt, October 2004, Work in Progress.

[BGP-RFSH] Chen, A., "Route Refresh Capability for BGP-4", RFC 2918, September 2000.

[BGP-ORF] Chen, E., and Rekhter, Y., "Cooperative Route Filtering

Capability for BGP-4", draft-ietf-idr-route-filter-11.txt,
December 2004, Work in Progress.

[BGP-CONS] Marques, P., et al., "Constrained VPN route distribution" draft-ietf-l3vpn-rt-constrain-01.txt, September 2004, work in progress

13. Annex: Auto-Discovery in VR and MPLS-IP-VPN Interworking Scenarios

Two interwoking scenarios are considered when the network is using both virtual routers and BGP/MPLS-IP-VPN. The first scenario is a CE-PE relationship between a PE (implementing BGP/MPLS-IP-VPN), and a VR appearing as a CE to the PE. The connection between the VR, and the PE can be either direct connectivity, or through a tunnel (e.g., IPSec).

The second scenario is when a PE is implementing both architectures. In this particular case, a single BGP session configured on the service provider network can be used to advertise either BGP/MPLS-IP-VPN VPN information or the virtual router related VPN information. From the VR and the BGP/MPLS-IP-VPN point of view there is complete separation from data path and addressing schemes. However the PE's interfaces are shared between both architectures.

A PE implementing only BGP/MPLS-IP-VPN will not import routes from a BGP UPDATE message containing the VPN-ID extended community. On the other hand, a PE implementing the virtual router architecture will not import routes from a BGP UPDATE message containing the route target extended community attribute.

The granularity at which the information is either BGP/MPLS-IP-VPN related or VR-related is per BGP UPDATE message. Different SAFI numbers are used to indicate that the message carried in BGP multiprotocol extension attributes is to be handled by the VR or BGP/MPLS-IP-VPN architectures.

14. Contributors

Bryan Gleeson
Nokia
313 Fairchild Drive
Mountain View CA 94043 USA
bryan.gleeson/at/nokia.com

Peter Ashwood-Smith
Nortel Networks
P.O. Box 3511 Station C,
Ottawa, ON K1Y 4H7, Canada
Phone: +1 613 763 4534

Email: petera@nortelnetworks.com

Luyuan Fang
AT&T
200 Laurel Avenue
Middletown, NJ 07748
Email: Luyuanfang@att.com
Phone: +1 (732) 420 1920

Jeremy De Clercq
Alcatel
Francis Wellesplein 1
B-2018 Antwerpen, Belgium
Phone: +32 3 240 47 52
Email: jeremy.de_clercq@alcatel.be

Riad Hartani
Caspian Networks
170 Baytech Drive
San Jose, CA 95143
Phone: 408 382 5216
Email: riad@caspiannetworks.com

Tissa Senevirathne
Force10 Networks
1440 McCarthy Blvd,
Milpitas, CA 95035.
Phone: 408-965-5103
Email: tsenevir@hotmail.com

15. Author' Addresses

Hamid Ould-Brahim
Nortel Networks
P O Box 3511 Station C
Ottawa, ON K1Y 4H7, Canada
Email: hbrahim@nortelnetworks.com

Eric C. Rosen
Cisco Systems, Inc.
1414 Massachusetts Avenue
Boxborough, MA 01719
E-mail: erosen@cisco.com

Yakov Rekhter
Juniper Networks
1194 N. Mathilda Avenue
Sunnyvale, CA 94089

Email: yakov@juniper.net

Intellectual Property Statement

Copyright Statement