Network Working Group
Internet Draft
Intended status: Standards Track
Expires: November 2008

Hamid Ould Brahim
Don Fedyk
(Nortel)
Yakov Rekhter
(Juniper Networks)

May 14, 2008

BGP-based Auto-Discovery for Layer-1 VPNs

draft-ietf-l1vpn-bgp-auto-discovery-05.txt

Status of this Memo

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.

Abstract

The purpose of this document is to define a BGP-based auto-discovery mechanism for Layer-1 VPNs (L1VPNs). The auto-discovery mechanism for L1VPNs allows the provider network devices to dynamically discover the set of PEs having ports attached to CE members of the same VPN. That information is necessary for completing the signaling phase of L1VPN connections. One main objective of a L1VPN auto-discovery mechanism is to support the "single-end provisioning" model, where addition of a new port to a given L1VPN would involve configuration changes only on the PE that has this port and on the CE that is connected to the PE via this port.

1. Introduction

The purpose of this document is to define a BGP-based auto-discovery mechanism for Layer-1 VPNs (L1VPNs) [L1VPN-FRMK]. The auto-discovery mechanism for L1VPNs allows the provider network devices to dynamically discover the set of PEs having ports attached to CE members of the same VPN. That information is necessary for completing the signaling phase of L1VPN connections. One main objective of a L1VPN auto-discovery mechanism is to support the "single-end provisioning" model, where addition of a new port to a given L1VPN would involve configuration changes only on the PE that has this port and on the CE that is connected to the PE via this port.

The auto-discovery mechanism proceeds by having a PE advertise to other PEs, at a minimum, its own IP address and the list of <private address, provider address> tuples local to that PE. Once that information is received, the remote PEs will identify the list of VPN members they have in common with the advertising PE, and use the information carried within the discovery mechanism to perform address resolution during the signaling phase of Layer-1 VPN connections.

Figure 1 highlights the network reference for using BGP-based auto-discovery mechanism for Layer-1 VPNs. For the purpose of auto-discovery mechanism, BGP is running only on the provider network. The PEs maintain per VPN information tables called Port Information Table (PIT) related to <private address, provider address> information. More information on the PIT tables is described in section 2.

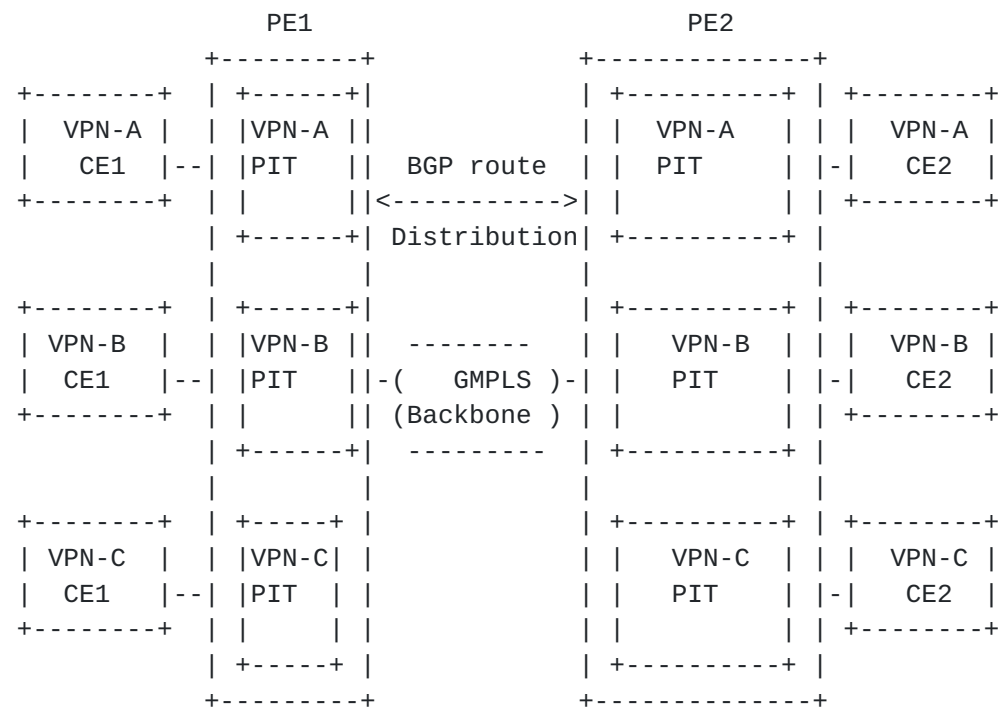

Figure 1 BGP auto-discovery for L1VPN

[L1VPN-FRMK] describes two modes of operation for a L1VPN: the basic mode and the enhanced mode. This document describes an auto-discovery mechanism for the basic mode only.

2. Procedures

In the context of L1VPNs, a CE is connected to a PE via one or more ports, where each port may consist of one or more channels or sub-channels. Each port on a CE that connects the CE to a PE has an identifier that is unique within that L1VPN (but need not be unique across several L1VPNs). We refer to this identifier as the customer port identifier (CPI). Each port on a PE also has an identifier that is unique within the provider network. We refer to this identifier as the provider port identifier (PPI). Note that IP addresses used for CPIs or PPIs could be either IPv4 or IPv6 addresses.

For each L1VPN that has at least one port configured on a PE, the PE maintains a Port Information Table (PIT). A PIT contains a list of <CPI, PPI> tuples for all the ports within its L1VPN. Note that a PIT may also hold routing information (for example when CPIs are learnt using a routing protocol).

A PIT on a given PE is populated with two types of information.

- Information related to the CEs' ports attached to the ports on the PE. This information could be locally configured at the PE or could

be received from the CEs.

- Information received from other PEs through the auto-discovery mechanism.

We refer to the former as local information, and to the latter as remote information. Propagation of local information to other PEs is accomplished by using BGP multiprotocol extensions [RFC4760]. To restrict the flow of this information to only the PITs within a given L1VPN, we use BGP route filtering based on the Route Target Extended Community [BGP-COMM], as follows.

Each PIT on a PE is configured with one or more Route Target Communities, called "export Route Targets", that are used for tagging the local information when it is exported into the provider's BGP. The granularity of such tagging could be as fine as a single <CPI, PPI> pair. In addition, each PIT on a PE is configured (at provisioning time) with one or more Route Target Communities, called "import Route Targets", that restrict the set of routes that could be imported from provider's BGP into the PIT to only the routes that have at least one of these Communities.

When a service provider adds a new L1VPN port to a particular PE (at provisioning time), this port is associated at provisioning time with a PIT on that PE, and this PIT is associated (again at provisioning time) with that L1VPN.

Note that since the protocol used to populate a PIT with remote information is BGP, since BGP works across multiple autonomous systems, it follows that the mechanism described in this document could support L1VPNs that span multiple autonomous systems.

Although multi-AS L1VPNs are currently out of scope for the Basic Mode, the mechanisms defined in this document appear to be easily applicable to a multi-AS scenario should such a need arise in the future. At that time additional work may be required to examine various aspects including security.

3. Carrying L1VPN information in BGP

The <CPI, PPI> mapping is carried using the Multiprotocol Extensions to BGP [RFC4760]. [RFC4760] defines the format of two BGP attributes, MP_REACH_NLRI and MP_UNREACH_NLRI that can be used to announce and withdraw the announcement of reachability information. We introduce a new subsequent address family identifier, called Layer-1 VPN auto-discovery information (to be assigned by the IANA), and also a new NLRI format for carrying the CPI and PPI information.

One or more <PPI, CPI> tuples could be carried in the above mentioned BGP attributes.

The format of the NLRI is described in figure 2.

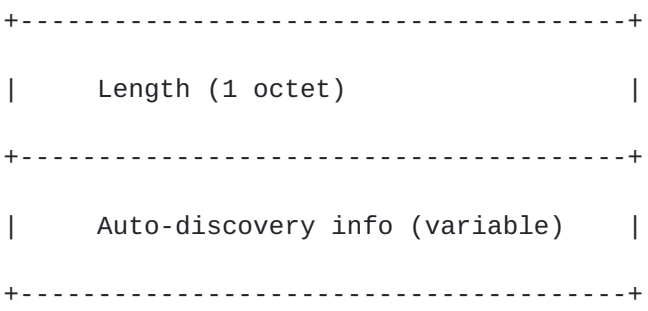

Figure 2 Encoding of the NLRI

Note that the encoding of the auto-discovery information is described in [L1VPN-BM] and note also that if the value of the Length of the Next Hop field (of the MP_REACH_NLRI attribute) is 4, then the Next Hop contains an IPv4 address. If this value is 16, then the Next Hop contains an IPv6 address.

4. Carrying L1VPN Traffic Engineering Information in BGP

In addition to reachability information, the auto-discovery mechanism MAY carry Traffic Engineering information used for the purpose of egress path selection. For example a PE may learn the switching capability and the maximum LSP bandwidth of remote L1VPN interfaces from the remote PEs. This document uses the BGP Traffic Engineering Attribute [BGP-TE-ATTRIBUTE] to carry such information.

5. Scalability

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

A PE router, unless it is a Route Reflector, does not retain L1VPN-related information unless it has at least one VPN with an Import Target identical to one of the VPN-related information Route Target attributes. If a PE does not have a VPN with a matching Import Route Target it MUST then discard received l1VPN information. Inbound filtering MUST 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 previously has discarded.

In this case the refresh mechanism described in [BGP-RFSH] MUST be used. 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 VPN (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 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 L1VPNs; 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 L1VPNs 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.

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.

6. Security Considerations

This document describes a BGP-based auto-discovery mechanism which enables a PE that attaches to a particular L1VPN 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 PEs to discover each other. Note that a PE will not always receive these advertisements directly from the remote PEs; the advertisements can be received from "intermediate" BGP speakers.

It is of critical importance that a particular PE MUST 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 nodes, and if the PE nodes accepted those

advertisements, then anyone could add any site to any L1VPN. 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 trust 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 L1VPNs 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 SHOULD 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 MUST be used only in environments where adequate trust relationships exist among the BGP speakers (such as the case of using the auto-discovery mechanism within a single provider network).

7. IANA Considerations

This document requires assignment of a new SAFI, called Layer-1 VPN auto-discovery information (see Section 3). This assignment has to be done from the Subsequent Address Family Identifier (SAFI) registry using the Standards Action allocation procedures. Suggested value is 69.

8. References

8.1. Normative References

[RFC4760] Bates, Chandra, Katz, and Rekhter, "Multiprotocol Extensions for BGP4", January 2007, RFC 4760.

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

8.2. Informative References

[BGP-TE-ATTRIBUTE] Ould-Brahim, H., Fedyk, D., Rekhter, Y., "Traffic Engineering Attribute", draft-ietf-softwire-bgp-te-attribute-00.txt, work in progress.

[BGP-ORF] Chen, E., and Rekhter, Y., "Outbound Route Filtering Capability for BGP-4", draft-ietf-idr-route-filter-16.txt, Work in Progress.

[BGP-CONS] Marques, P., et al., "Constrained VPN route distribution", RFC4684.

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

[L1VPN-FRMK] Tomonori Takeda, et al., "Framework and Requirements for Layer 1 Virtual Private Networks", RFC4847.

[L1VPN-BM] Fedyk, D., Rekhter, Y. (Eds.), "Layer 1 VPN Basic Mode", draft-ietf-l1vpn-basic-mode, work in progress.

9. Acknowledgment

We would like to thank Adrian Farrel for the useful comments.

10. Authors' Addresses

Hamid Ould-Brahim
Nortel
P O Box 3511 Station C
Ottawa ON K1Y 4H7 Canada
Phone: +1 (613) 763 4730
Email: hbrahim@nortel.com

Yakov Rekhter
Juniper Networks
1194 N. Mathilda Avenue
Sunnyvale, CA 94089
Email: yakov@juniper.net

Don Fedyk
Nortel
600 Technology Park
Billerica, Massachusetts
01821 U.S.A
Phone: +1 (978) 288 3041
Email: dwfedyk@nortel.com