Abstract

This document defines protocol extensions and procedures for BGP PE-CE router iteration in RFC2547bis [4] networks. These have the objective of making the usage of the RFC2547bis VPN transparent to the customer network, as far as routing information is concerned.
Table of Contents

1. Introduction .............................................. 3
2. IP VPN network as a Route Server ......................... 4
3. Path attributes ............................................ 5
4. Carrying internal BGP routes ............................... 6
5. Next-hop handling ......................................... 7
6. Exchanging routes between different VPN customer networks .. 8
7. Security considerations ................................... 10
8. IANA considerations ....................................... 11
9. Acknowledgments ........................................... 12
10. Normative References .................................... 12
Authors’ Addresses ........................................... 13
Intellectual Property and Copyright Statements ................ 14
1. Introduction

In current deployments, when BGP is used as the PE-CE routing protocol, these peering sessions are typically configured as an external peering between the VPN provider AS and the customer network AS. At each External BGP boundary, Path Attributes [1] are modified as per standard BGP rules. This includes prepending the AS_PATH attribute with the autonomous system of the originating customer CE and the autonomous system(s) of the provider edge router(s).

In order for such routes not to be rejected by AS_PATH loop detection, a PE router advertising a route received from a remote PE, often remaps the customer network autonomous-system number to its own. Otherwise the customer network can use different autonomous-system numbers at different sites or configure their CE routers to accept routes containing their own AS number.

While this technique works well in situations where there are no BGP routing exchanges between the client network and other networks, it does have drawbacks for customer networks that use BGP internally for purposes other than interaction between CE and PE routers.

In order to make the usage of RFC2547bis VPN services as transparent as possible to any external interaction, it is desirable to define a mechanism by which PE-CE routers can exchange BGP routes by means other than external BGP.

One can consider a RFC2547bis VPN as a provider-managed backbone service interconnecting several customer-managed sites. While this model is not universal it does constitute a good starting point.

Independently of the presence of VPN service, networks which use an hierarchical design are typically modeled such that the top-level core or backbone participates in a full iBGP mesh which distributes routing information between sites via BGP route reflection [2] or confederations [3]. This will be our service model definition.
2. IP VPN network as a Route Server

In a typical backbone/area hierarchical design, routers that attach an area (or site) to the core, use BGP route reflection (or confederations) to distribute routes between the top-level core iBGP mesh and the local area iBGP cluster.

To provide equivalent functionality in a network using a provider provisioned backbone, one can consider the VPN network as the equivalent of an Internal BGP Route Server which multiplexes information from N VPN attachment points.

A route learned by any of the PEs in the IP VPN network, is available to all other PEs that import the Route Target used to identify the customer network. This is conceptually equivalent to a centralized route server.

In a PE router, PE received routes are not advertised back to other PEs. It is this split horizon technique that prevents routing loops in an IP VPN environment.

When a route is advertised from PE to CE, if its is advertised as an iBGP route, the CE will not advertise it further unless it is itself configured as a Route Reflector (or has an external BGP session). This is a consequence of the default BGP behavior of not advertising iBGP routes back to iBGP peers.

A PE router can also act as a route reflector to local CE routers. Reflection can also be used hierarchically in order to avoid direct communication between the PE and non-directly connected CEs that may exist in the site.

This Route Server model can also be used to support a confederation style abstraction to CE devices. We choose not to describe in detail the procedures for that mode of operation, at this point. Confederations are considered to be less common than route reflection in enterprise environments.
3. Path attributes

--> push path attributes --> vrf-export --> 2547
VRF route PE-PE route advertisement

<-- pop path attributes <-- vrf-import <--

The diagram above shows the BGP path attribute stack processing in relation to existing 2547 route processing procedures. BGP path attributes received from a customer network are pushed into the stack, before adding the Export Route Targets to the BGP path attributes. Conversely, the stack is popped after the Import Target processing step that identifies the VRF table in which a PE received route is accepted.

When a PE received route is imported into a VRF, its IGP metric, as far as BGP path selection is concerned, should be the metric to the remote PE address, expressed in terms of the service provider metric domain.

For the purposes of VRF route selection performed at the PE, between routes received from local CEs and remote PEs, VPN network IGP metrics should always be considered higher (thus least preferred) than local site metrics.

When backdoor links are present, this would tend to direct the traffic between two sites through the backdoor link for BGP routes originated by a remote site. However BGP already has policy mechanisms to address this type of situations such as the LOCAL_PREF attribute.

When a given CE is connected to more than one PE, it will not advertise the route that it receives from a PE to another PE unless configured as a route reflector, due to the standard BGP route advertisement rules.

When a CE reflects a PE received route to another PE, the fact that the original attributes of a route are preserved across the VPN network prevents the formation of routing loops due to mutual redistribution between the two networks.
4. Carrying internal BGP routes

In order to carry the original BGP attributes of a route received from a CE, this document defines a new BGP path attribute:

ATTR_SET (type code 128)

ATTR_SET is an optional transitive attribute that carries a set of BGP path attributes. An attribute set (ATTR_SET) can include any BGP attribute that can occur in a BGP UPDATE message, except the MP_REACH and MP_UNREACH attributes.

This attribute is used by a PE router to store the original set of BGP attributes it receives from a CE. When a PE router advertises a PE-received route to a CE, it will use the path attributes carried in the ATTR_SET attribute.

In other words, the BGP Path Attributes are "pushed" into this stack like attribute when the route is received by the VPN network and "popped" when the route is advertised in the PE to CE direction.

Using this mechanism isolates the customer network from the attributes used in the VPN network and vice versa. Attributes as the route reflection cluster list attribute are segregated such that customer network cluster identifiers won’t be considered by the VPN network route reflectors and vice-versa.

The autonomous system number present in the ATTR_SET attribute is designed to prevent a route originating in a given autonomous-system iBGP to be leaked into a different autonomous-system, without proper AS_PATH manipulation. It should contain the autonomous system of the customer network that originates the given set of attributes.

The NEXT_HOP attribute SHOULD NOT be included in an ATTR_SET.
5. Next-hop handling

When RFC2547bis VPNs are not in use, the NEXT_HOP attribute in iBGP routes carries the address of the border router advertising the route into the domain.

An important component of BGP route selection is the IGP distance to the NEXT_HOP of the route.

When a VPN service is used to provide interconnection between different sites, since the VPN network runs a different IGP domain, metrics between the VPN and customer networks are not comparable.

However, the most important component of a metric is the inter-area metric, which is known to the VPN network. The intra-area metric is typically negligible.

The use of route reflection, for instance, requires metrics to be configured so that inter-cluster/area metrics are always greater than intra-cluster metrics.

The approach taken by this document is to rewrite the NEXT_HOP attribute at the PE-CE boundary. PE routers take into account the PE-PE IGP distance calculated by the VPN network IGP, when selecting between routes advertised from different PEs.

An advantage of the proposed method is that the customer network can run independent IGPs at each site.
6. Exchanging routes between different VPN customer networks

A given VPN customer network SHOULD use internal or external BGP sessions consistently for peering sessions where the same autonomous system is used.

In scenarios such as what is commonly referred to an "extranet" VPN, routes MAY be advertised to both internal and external VPN attachments, belonging to different autonomous systems.

Consider the example given above where (PE1, CE1) and (PE2, CE3) sessions are iBGP. In RFC2547 VPNs, a route received from CE1 above may be distributed to the VRFs corresponding to the attachment points for CE2 and 3.

The desired result, in such a scenario is to present the internal peer (CE3) with a BGP advertisement that contains the same BGP Path Attributes received from CE1 and to the external peer (CE2) a BGP advertisement that would correspond to a situation where AS 1 and 2 have a external BGP session between them.

In order to achieve this goal the following set of rules apply:

When advertising an iBGP originated route to iBGP, a PE router MUST check that the autonomous-system contained in the ATTR_SET attribute matches the autonomous system of the CE to which the route is being advertised.

In case the autonomous-systems do match, the route is advertised with the attributes contained in the ATTR_SET attribute. Otherwise, in the case of an autonomous-system mismatch, the set of attributes to be advertised to the CE in question shall be constructed as follows:

1. The path attributes are set to the attributes contained in the ATTR_SET attribute.

2. Internal BGP specific attributes are discarded (LOCAL_PREF, ORIGINATOR, CLUSTER_LIST, etc).
3. The autonomous-system contained in the ATTR_SET attribute is prepended to the as-path following the rules that would apply to an external BGP peering between the source and destination ASes.

4. Internal BGP specific attributes corresponding to the configuration of destination AS (LOCAL_PREF) are added.

When advertising an iBGP originated route to eBGP, a PE router shall apply steps 1 to 3 defined above and subsequently prepend its own autonomous-system number to the AS_PATH attribute (i.e. both the originator and VPN network as numbers are prepended).

When advertising an eBGP originated route to iBGP, a PE router MUST prepend its own as number before adding iBGP only as-path attributes (LOCAL_PREF).

In all cases where an iBGP originating route is processed, attributes present on the VPN route other than the NEXT_HOP attribute are ignored, both from the point of view of route selection in the VRF Adj-RIB-in and route advertisement to a CE router.
7. Security considerations

It is worthwhile to consider the security implications of this proposal from two independent perspectives: the IP VPN provider and the IP VPN customer.

From a IP VPN provider perspective, this mechanism will assure separation between the BGP path attributes advertised by the customer CE router and the BGP attributes used within the provider network, thus potentially improving security.

Although this behavior is largely implementation dependent, currently it is possible for a CE device to inject BGP attributes (extended communities, for example) that have semantics on the IP VPN provider network, unless explicitly disabled by configuration in the PE.

With the rules specified for the ATTR_SET path attribute, any attribute that has been received from a CE is pushed into the stack before the route is advertised out to other PEs.

From the perspective of the VPN customer network, it is our opinion that there is no change to the security profile of PE-CE interaction. While having an iBGP session allows the PE to specify additional attributes not allowed on an eBGP session (e.g. local-pref), this does not significantly change the fact that the VPN customer must trust its service provider to provide it correct routing information.
8. IANA considerations

This document defines a new BGP path attribute which is part of a registry space managed by IANA. We request that IANA update its registry with the value specified above for the ATTR_SET path attribute.
9. Acknowledgments

The authors would like to acknowledge the contributions of Yakov Rekhter, Luyuan Fang and Jan Novak.

10. Normative References


Authors' Addresses

Pedro Marques
Juniper Networks
1194 N. Mathilda Ave.
Sunnyvale, CA  94089
US

Email: roque@juniper.net

Robert Raszuk
Cisco Systems, Inc.
170 West Tasman Dr
San Jose, CA  95134
US

Email: rraszuk@cisco.com

Luca Martini
Cisco Systems, Inc.
9155 East Nichols Avenue, Suite 400
Englewood, CO  80112
US

Email: lmartini@cisco.com

Dan Tappan
Cisco Systems, Inc.
300 Beaver Brook Rd.
Boxborough, MA  01719
US

Email: tappan@cisco.com
Intellectual Property Statement

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Disclaimer of Validity

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Copyright Statement

Copyright (C) The Internet Society (2005). This document is subject to the rights, licenses and restrictions contained in BCP 78, and except as set forth therein, the authors retain all their rights.

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.