Network Working Group
Internet-Draft
Intended status: Informational
Expires: March 3, 2012

L. Cheng
M. Sun
ZTE Corporation
August 31, 2011

Auto-Configuration Extention in Virtual Aggregation
draft-cheng-grow-va-auto-extensions-00

Abstract

Auto-Configuration in Virtual Aggregation as specified in [I-D.ietf-grow-va-auto] requires configuration of a "VP-range list" in ASBRs connected to transit and peer ISPs. These ASBRs simply tag some routes whose prefix falls within the VP-Range with a "can-suppress" tag to indicate whether these routes should be FIB installed. This draft specified an extended auto-configuration mechanism in Virtual Aggregation to support the configuration of both "VP-List" and "popular prefixes". Specifically, based on this mechanism, the ratio of lost packets when VP routes fail could be minimized.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on March 3, 2012.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents

carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction

Virtual Aggregation specified in [I-D.ietf-grow-va] requires configuration of a static "VP-List" on all routers. "VP-List" allows routers to know which prefixes may or may not be FIB installed. Auto-configuration mechanism [I-D.ietf-grow-va-auto] provides an optional method for routers to do routes decision with less configuration.

Auto-configuration is an optional alternative to the VP-list that requires far less configuration. However, further concentrates should be focused on some scenarios where packets transmission maybe seriously influenced based on this mechanism. Furthermore, this mechanism could also be extended to provide more excellent service.

This draft specifies Auto-Configuration Extension Operation, which includes the following two aspects:

- VP routes to be specified particularly. Based on current auto-configuration, tagging routers must not tag VP routes with can-suppress tag. If the ISP has a policy of FIB-installing customer routes, then routes received from customers should also not be tagged. Consequently, there may be three kinds of routes are non-tagged in the AS: routes whose prefix out scope of VP-Range, VP routes and customer routes. According to these tagging rules, non-tagging routers will not be able to identify VP routes. As a result, in the case where all VP routes for a given VP are withdrawn, non-tagging routers would not be able to FIB-install sub-prefixes within the VP. This will influence the normal transmission of data packets seriously.

- Extensions to realize popular prefixes auto-configuration. As specified in [I-D.ietf-grow-va], deployment of Visual Aggregation will cause path stretch. To minimize the latency and load associated with the longer path, ISP could measure traffic volume over time and install the high volume prefixes. These prefixes which are within a VP, but still be FIB installed are called popular prefixes. Furthermore, popular prefixes could also be consisted of policy-based prefixes and static list prefixes. Customer prefixes could be considered as one kind of policy- based prefixes. Consequently, tagging routers could also be configured with a popular prefixes list, and realize popular prefixes auto-configuration by not tag routes whose prefix falls within this list.

1.1. Requirements Language

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 [RFC2119].

2. Terminology

This draft uses terms defined in [I-D.ietf-grow-va]. This section defines some new terms used in this document.

Tagging router: ASBRs which are configured with "VP range" and "Popular-Prefix list". These routers tag routes with different tags based on route type. Typically, all ASBRs that connect to one or more transit provider ISPs must be configured as tagging routers. ASBRs that connect to one or more peer ISPs should be configured as tagging routers. ASBRs that connect to customer networks should not be configured as tagging routers.

Non-tagging router: The VA routers in AS which are not tagging routers.

Popular-Prefix list: List of popular prefixes.

Suppress tag: Tags used by tagging routers to tag routes. Routes with this tag may not be FIB installed by routers. This tag could be attached to a route as a Non-transitive Extended Communities Attribute.

Install tag: Tags used by tagging routers to tag routes. Router with this tag must be FIB installed by any router. This tag could be attached to a route as a Non-transitive Extended Communities Attribute.

3. Specification

3.1. Routes Classification and Routes Tagging

With this extended auto-configuration approach, every tagging router will be configured with the same "VP-range list" and "popular prefix list".

"VP-range list" consists of the ranges of IP address that are collectively covered by all VPs in the AS [I-D.ietf-grow-va-auto]. "Popular-Prefix list" is a list of popular prefixes. These popular prefixes are all regular prefixes, and could be selected by ISPs

individually based on their requirements.

With the extended auto-configuration approach, ASBRs which are tagging routers first classify all routes into three types based on the "VP range" and "Popular-Prefix list" configured in them:

Type1: VP routes which MUST be FIB installed by any router;
Type2: Routes whose prefix falls within "Popular-Prefix list", and routes whose prefix is not fall within "VP range";
Type3: Routes whose prefix falls within "VP range", and meantime are out scope of Type 1 and Type 2 routes.

Tagging routers tag routes explicitly according to route types.

1. All VP routes (type 1) MUST be tagged with a "install tag".
2. All routes falls within Type 2 SHOULD NOT be tagged.
3. All routes falls within Type 3 MAYBE tagged with a "suppress tag".

3.2. Routes Installation

Routers install or suppress FIB entries according to the following rules.

1. Routes with "install tag" MUST be FIB-installed.
2. Routes without any tag SHOULD be FIB-installed.
3. Routes with "suppress tag" MAY be FIB-suppressed.
4. APRs MUST FIB-install routes for sub-prefixes that fall within the APRs!_ VPs, whether or not the route is tagged.

Note: tagging routers conceptually follow these rules after tagging (or not tagging) the route.
Note: these rules apply only to the route used by the routers as the best route.

3.3. Implementation

An instance of mechanism operation is depicted in this subsection. Consider the scenario depicted in Figure. 1.

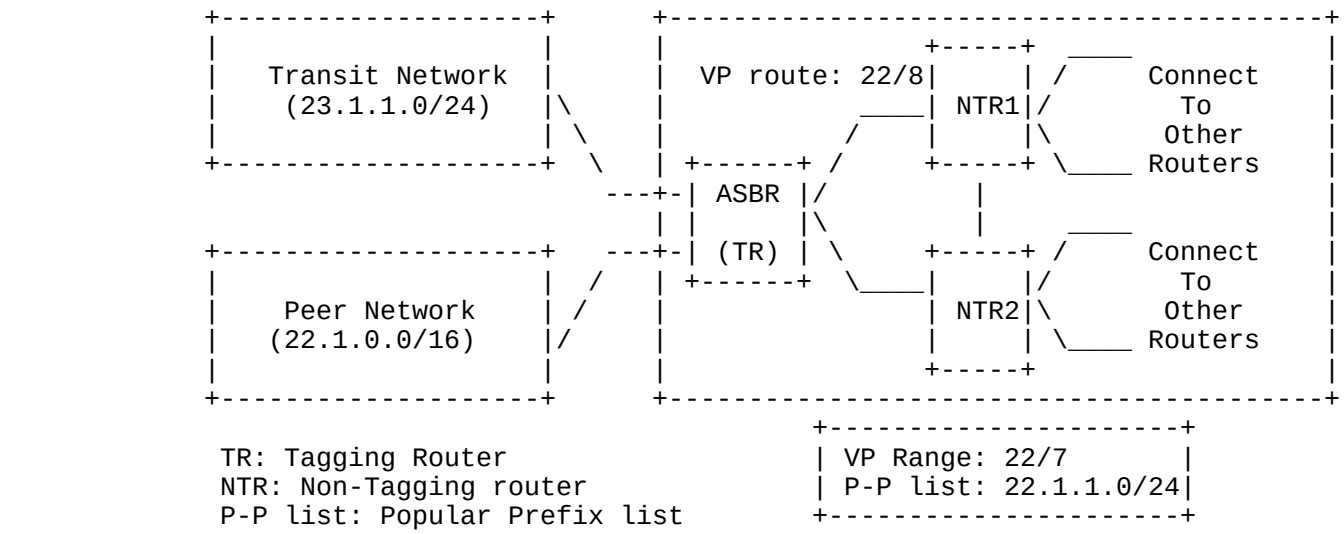


Figure. 1

In this situation, an ASBR connected to transit network (23.1.1.0/24) and peer network (22.1.0.0/16) is selected to be a TR (tagging router). TR is configured with a VP range 22/7 and a popular prefix list contains 23.1.1.0/24. The NTR1 (Non-Tagging Router 1) is an APR (Aggregate Point Router) who announces a VP route 22/8.

To describe operation of different elements, assume that following routes will be received by TR, NTR1 and NTR2.

Routes	Prefix
1	22/8
2	22.1.1.1/32
3	22.1.0.1/32
4	23.1.1.1/32

TR Operation:

- For route with prefix 22/8, as this route is a VP route announced by NTR1, TR will tag it with a "install tag".
- The prefix 22.1.1.1/32 falls within the popular prefix list, TR will not tag it, although the route’s prefix falls within the VP range.

- o TR perceives that prefix 22.1.0.1/32 falls within the VP range and is not a popular prefix. This route will be tagged with a "suppress tag".
- o For route with prefix 23.1.1.1/32, as the prefix falls within VP range 22/7, this route will also be tagged with a "suppress tag".

NTR Operation:

- o For route with prefix 22/8, as this route is tagged with "install tag", all NTRs must FIB install it.
- o For route with prefix 22.1.1.1/32, as this route is not tagged, NTRs should FIB install it. Especially for NTR1, as the prefix falls within the VP (22/8) it announced, it must FIB install the route.

It should be noticed that in this scenario, NTR1 is an APR, and NTR2 is a non-APR. These two kinds of routers will implement different operation upon some suppress tagged routes.

- o According to NTR2, as the router is a non-APR, all routes with "suppress tag" should not be installed. As a result, NTR2 will not FIB install 22.1.0.1/32 and 23.1.1.1/32.

- o Now consider the operation of NTR1.
  - * FOr the route 22.1.0.1/32 with "suppress tag", NTR1 perceives that prefix falls within 22/8, and will FIB install the route.
  - * According to route 23.1.1.1/32 with "suppress tag", NTR1 doesn’t have to FIB install it, as NTR1 is not the APR for this route.

4. Operation under Special Scenario

Based on analysis in section 1, when VP routes are not tagged specially, VP routes failing will influence the packets transmission seriously.

From perspective of non-tagging routers, VP routes could be identified through the "install tag" based on extended auto-configuration mechanism. This section assumes a special scenario that all VP routes for a given VP are withdrawn. Proper operation of tagging routers and non-tagging routers is described as following.

4.1. Non-tagging Routers Operation

When the non-tagging routers find that all VP routes for a given VP withdrawn, they will immediately look up the routing table in the RIB, select and FIB install the suppress tagged routes whose prefixes

fall within the withdrawn VP.

4.2. Tagging Routers Operation

When the tagging routers find that all VP routes for a given VP are withdrawn, they will implement the following operation:

- Look up the routing table in the RIB, select and FIB install the suppress tagged routes whose prefixes fall within the withdrawn VP;
- Record VP information of the withdrawn VP routes;
- When there is an invalid record for a given VP, all routes received by the tagging routers whose prefixs falls within this VP should not be tagged.

According to existence of invalid VP records, once receiving a VP route, tagging routers will compare the VP prefix received with the VP prefixes recorded. If the received prefix match an invalid prefix record, they will implement the following operation:

- Delete the invalid VP prefix record;
- Tag received VP route with "install tag";
- Tag all Type 3 routes whose prefix falls within this VP with "suppress tag".

4.3. Implementation

Consider the implementation usecase depicted in Figure. 1. Assume that all VP routes for the VP 22/8 are withdrawn.

According to this problem, NTRs (include APRs) check their routing table in RIB, and find out suppress tagged routes whose prefixes fall within the VP 22/8, such as routes with 22.1.0.1/32. NTRs will FIB install these routes, and forward packets based on them.

According to this problem, TRs will also FIB install the suppressed tagged routes whose prefixes fall within the VP 22/8. Furthermore, TRs will record the prefix information of VP 22/8. During the period that VP 22/8 is withdrawn, all routes whose prefix falls within the VP will not be tagged by TRs, indicates that routes such like 22.1.1.1/32 and 22.1.0.1/32 should be FIB installed.

Every TR maintains a record of the withdrawn VP 22/8. When the TR receives a new VP route with prefix 22/8, it will consider this situation as "VP Recovery". Withdrawn record of 22/8 will be deleted, and the new VP route will be tagged with "install tag". Furthermore, the Type 3 routes whose prefixes fall within the 22/8 such as 22.1.0.1/32 will be tagged with "suppress tag".

5. IANA Considerations

IANA is requested to assign, from the registry "BGP Assigned non-transitive extended communities", values TBD for "must install" and "can suppress".

Registry Name: BGP Assigned non-transitive extended communities

Name	Type Value
must install	TBD
can suppress	TBD

6. Security Considerations

TBD

7. References

7.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC4360] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", RFC 4360, February 2006.

7.2. Informative References

[I-D.ietf-grow-va]
Francis, P., Xu, X., Ballani, H., Jen, D., Raszuk, R., and L. Zhang, "FIB Suppression with Virtual Aggregation", draft-ietf-grow-va-05 (work in progress), June 2011.

[I-D.ietf-grow-va-auto]
Francis, P., Xu, X., Ballani, H., Jen, D., Raszuk, R., and L. Zhang, "Auto-Configuration in Virtual Aggregation", draft-ietf-grow-va-auto-04 (work in progress), June 2011.

[I-D.ietf-idr-reserved-extended-communities]
Decraene, B. and B. Decraene, "Assigned BGP extended communities", draft-ietf-idr-reserved-extended-communities-01 (work in progress), May 2011.

Authors' Addresses

Li Cheng
ZTE Corporation
Zijinghua Road No.68
NanJing, Yuhuatai District 210012
P.R.China

Email: cheng.li2@zte.com.cn

Mo Sun
ZTE Corporation
Zijinghua Road No.68
NanJing, Yuhuatai District 210012
P.R.China

Phone: +86-025-52871474
Email: sun.mo@zte.com.cn

GROW Working Group
Internet-Draft
Intended status: Informational
Expires: December 09, 2013

N. Hilliard
INEX
E. Jasinska
Microsoft Corporation
R. Raszuk
NTT I3
N. Bakker
AMS-IX B.V.

June 07, 2013

Internet Exchange Route Server Operations
draft-hilliard-ix-bgp-route-server-operations-03

Abstract

The popularity of Internet exchange points (IXPs) brings new challenges to interconnecting networks. While bilateral eBGP sessions between exchange participants were historically the most common means of exchanging reachability information over an IXP, the overhead associated with this interconnection method causes serious operational and administrative scaling problems for IXP participants.

Multilateral interconnection using Internet route servers can dramatically reduce the administrative and operational overhead of IXP participation and these systems used by many IXP participants as a preferred means of exchanging routing information.

This document describes operational considerations for multilateral interconnections at IXPs.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on December 09, 2013.

Copyright Notice

Copyright (c) 2013 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction

Internet exchange points (IXPs) provide IP data interconnection facilities for their participants, typically using shared Layer-2

networking media such as Ethernet. The Border Gateway Protocol (BGP) [RFC4271] is normally used to facilitate exchange of network reachability information over these media.

As bilateral interconnection between IXP participants requires operational and administrative overhead, BGP route servers [I-D.ietf-idr-ix-bgp-route-server] are often deployed by IXP operators to provide a simple and convenient means of interconnecting IXP participants with each other. A route server redistributes prefixes received from its BGP clients to other clients according to a pre-specified policy, and it can be viewed as similar to an eBGP equivalent of an iBGP [RFC4456] route reflector.

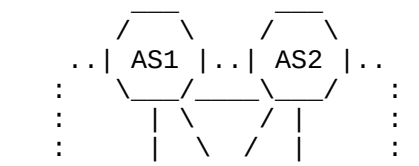
Route servers at IXPs require careful management and it is important for route server operators to thoroughly understand both how they work and what their limitations are. In this document, we discuss several issues of operational relevance to route server operators and provide recommendations to help route server operators provision a reliable interconnection service.

1.1. Notational Conventions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

2. Bilateral BGP Sessions

Bilateral interconnection is a method of interconnecting routers using individual BGP sessions between each participant router on an IXP, in order to exchange reachability information. If an IXP participant wishes to implement an open interconnection policy - i.e. a policy of interconnecting with as many other IXP participants as possible - it is necessary for the participant to liaise with each of their intended interconnection partners. Interconnection can then be implemented bilaterally by configuring a BGP session on both participants' routers to exchange network reachability information. If each exchange participant interconnects with each other participant, a full mesh of BGP sessions is needed, as shown in Figure 1.



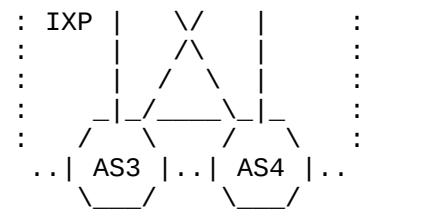


Figure 1: Full-Mesh Interconnection at an IXP

Figure 1 depicts an IXP platform with four connected routers, administered by four separate exchange participants, each of them with a locally unique autonomous system number: AS1, AS2, AS3 and AS4. Each of these four participants wishes to exchange traffic with all other participants; this is accomplished by configuring a full mesh of BGP sessions on each router connected to the exchange, resulting in 6 BGP sessions across the IXP fabric.

The number of BGP sessions at an exchange has an upper bound of n*(n-1)/2, where n is the number of routers at the exchange. As many exchanges have large numbers of participating networks, the amount of administrative and operation overhead required to implement an open interconnection scales quadratically. New participants to an IXP require significant initial resourcing in order to gain value from their IXP connection, while existing exchange participants need to commit ongoing resources in order to benefit from interconnecting with these new participants.

3. Multilateral Interconnection

Multilateral interconnection is implemented using a route server configured to use BGP to distribute network layer reachability information (NLRI) among all client routers. The route server preserves the BGP NEXT_HOP attribute from all received NLRI UPDATE messages, and passes these messages with unchanged NEXT_HOP to its route server clients, according to its configured routing policy, as described in [I-D.ietf-idr-ix-bgp-route-server]. Using this method of exchanging NLRI messages, an IXP participant router can receive an aggregated list of prefixes from all other route server clients using a single BGP session to the route server instead of depending on BGP sessions with each other router at the exchange. This reduces the overall number of BGP sessions at an Internet exchange from n*(n-1)/2 to n, where n is the number of routers at the exchange.

Although a route server uses BGP to exchange reachability information with each of its clients, it does not forward traffic itself and is therefore not a router.

In practical terms, this allows dense interconnection between IXP participants with low administrative overhead and significantly simpler and smaller router configurations. In particular, new IXP participants benefit from immediate and extensive interconnection, while existing route server participants receive reachability information from these new participants without necessarily having to modify their configurations.

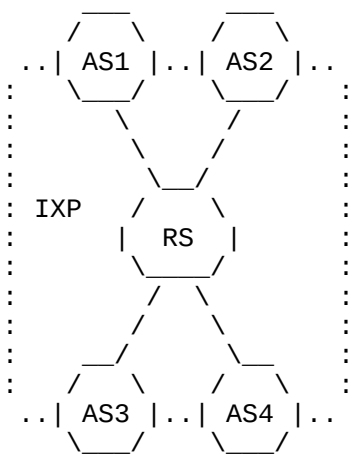


Figure 2: IXP-based Interconnection with Route Server

As illustrated in Figure 2, each router on the IXP fabric requires only a single BGP session to the route server, from which it can receive reachability information for all other routers on the IXP which also connect to the route server.

4. Operational Considerations for Route Server Installations

4.1. Path Hiding

"Path hiding" is a term used in [I-D.ietf-idr-ix-bgp-route-server] to describe the process whereby a route server may mask individual paths by applying conflicting routing policies to its Loc-RIB. When this happens, route server clients receive incomplete information from the route server about network reachability.

There are several approaches which may be used to mitigate against the effect of path hiding; these are described in [I-D.ietf-idr-ix-bgp-route-server]. However, the only method which does not require explicit support from the route server client is for the route server itself to maintain a individual Loc-RIB for each client which is the subject of conflicting routing policies.

4.2. Route Server Scaling

While deployment of multiple Loc-RIBs on the route server presents a simple way to avoid the path hiding problem noted in Section 4.1, this approach requires significantly more computing resources on the route server than where a single Loc-RIB is deployed for all clients. As the [RFC4271] BGP decision process must be applied to all Loc-RIBs deployed on the route server, both CPU and memory requirements on the host computer scale approximately according to O(P * N), where P is the total number of unique paths received by the route server and N is the number of route server clients which require a unique Loc-RIB. As this is a super-linear scaling relationship, large route servers may derive benefit from deploying per-client Loc-RIBs only where they are required.

Regardless of any Loc-RIB optimization technique is implemented, the route server's control plane bandwidth requirements will scale according to O(P * N), where P is the total number of unique paths received by the route server and N is the total number of route server clients. In the case where P_avg (the arithmetic mean number of unique paths received per route server client) remains roughly constant even as the number of connected clients increases, this relationship can be rewritten as O((P_avg * N) * N) or O(N^2). This quadratic upper bound on the network traffic requirements indicates that the route server model will not scale to arbitrarily large sizes.

This scaling analysis presents problems in three key areas: route processor CPU overhead associated with BGP decision process calculations, the memory requirements for handling many different BGP path entries, and the network traffic bandwidth required to distribute these prefixes from the route server to each route server client.

4.2.1. Tackling Scaling Issues

The network traffic scaling issue presents significant difficulties with no clear solution - ultimately, each client must receive a UPDATE for each unique prefix received by the route server. However, there are several potential methods for dealing with the CPU and memory resource requirements of route servers.

4.2.1.1. View Merging and Decomposition

View merging and decomposition, outlined in [RS-ARCH], describes a method of optimising memory and CPU requirements where multiple route server clients are subject to exactly the same routing policies. In this situation, the multiple Loc-RIB views required by each client are merged into a single view.

There are several variations of this approach. If the route server operator has prior knowledge of interconnection relationships between route server clients, then the operator may configure separate Loc-RIBs only for route server clients with unique outbound routing policies. As this approach requires prior knowledge of interconnection relationships, the route server operator must depend on each client sharing their interconnection policies, either in a internal provisioning database controlled by the operator, or else in an external data store such as an Internet Routing Registry Database.

Conversely, the route server implementation itself may implement internal view decomposition by creating virtual Loc-RIBs based on a single in-memory master Loc-RIB, with delta differences for each prefix subject to different routing policies. This allows a more granular and flexible approach to the problem of Loc-RIB scaling, at the expense of requiring a more complex in-memory Loc-RIB structure.

Whatever method of view merging and decomposition is chosen on a route server, pathological edge cases can be created whereby they will scale no better than fully non-optimised per-client Loc-RIBs. However, as most route server clients connect to a route server for the purposes of reducing overhead, rather than implementing complex per-client routing policies, edge cases tend not to arise in practice.

4.2.1.2. Destination Splitting

Destination splitting, also described in [RS-ARCH], describes a method for route server clients to connect to multiple route servers and to send non-overlapping sets of prefixes to each route server. As each route server computes the best path for its own set of prefixes, the quadratic scaling requirement operates on multiple smaller sets of prefixes. This reduces the overall computational and memory requirements for managing multiple Loc-RIBs and performing the best-path calculation on each. In order for this method to perform well, destination splitting would require significant co-ordination between the route server operator and each route server client. In practice, this level of close co-ordination between IXP operators and their participants tends not to occur, suggesting that the approach is unlikely to be of any real use on production IXPs.

4.2.1.3. NEXT_HOP Resolution

As route servers are usually deployed at IXPs which use flat layer 2 networks, recursive resolution of the NEXT_HOP attribute is generally not required, and can be replaced by a simple check to ensure that the NEXT_HOP value for each prefix is a network address on the IXP LAN's IP address range.

4.3. Prefix Leakage Mitigation

Prefix leakage occurs when a BGP client unintentionally distributes NLRI UPDATE messages to one or more neighboring BGP routers. Prefix leakage of this form to a route server can cause serious connectivity problems at an IXP if each route server client is configured to accept all prefix UPDATE messages from the route server. It is therefore RECOMMENDED when deploying route servers that, due to the potential for collateral damage caused by NLRI leakage, route server operators deploy prefix leakage mitigation measures in order to prevent unintentional prefix announcements or else limit the scale of any such leak. Although not foolproof, per-client inbound prefix limits can restrict the damage caused by prefix leakage in many cases. Per-client inbound prefix filtering on the route server is a more deterministic and usually more reliable means of preventing prefix leakage, but requires more administrative resources to maintain properly.

If a route server operator implements per-client inbound prefix filtering, then it is RECOMMENDED that the operator also builds in mechanisms to automatically compare the Adj-RIB-In received from each client with the inbound prefix lists configured for those clients. Naturally, it is the responsibility of the route server client to ensure that their stated prefix list is compatible with what they announce to an IXP route server. However, many network operators do not carefully manage their published routing policies and it is not uncommon to see significant variation between the two sets of prefixes. Route server operator visibility into this discrepancy can provide significant advantages to both operator and client.

4.4. Route Server Redundancy

As the purpose of an IXP route server implementation is to provide a reliable reachability brokerage service, it is RECOMMENDED that exchange operators who implement route server systems provision multiple route servers on each shared Layer-2 domain. There is no requirement to use the same BGP implementation or operating system for each route server on the IXP fabric; however, it is RECOMMENDED that where an operator provisions more than a single server on the same shared Layer-2 domain, each route server implementation be

configured equivalently and in such a manner that the path reachability information from each system is identical.

4.5. AS_PATH Consistency Check

[RFC4271] requires that every BGP speaker which advertises a route to another external BGP speaker prepends its own AS number as the last element of the AS_PATH sequence. Therefore the leftmost AS in an AS_PATH attribute should be equal to the autonomous system number of the BGP speaker which sent the UPDATE message.

As [I-D.ietf-idr-ix-bgp-route-server] suggests that route servers should not modify the AS_PATH attribute, a consistency check on the AS_PATH of an UPDATE received by a route server client would normally fail. It is therefore RECOMMENDED that route server clients disable the AS_PATH consistency check towards the route server.

4.6. Export Routing Policies

Policy filtering is commonly implemented on route servers to provide prefix distribution control mechanisms for route server clients. A route server "export" policy is a policy which affects prefixes sent from the route server to a route server client. Several different strategies are commonly used for implementing route server export policies.

4.6.1. BGP Communities

Prefixes sent to the route server are tagged with specific [RFC1997] or [RFC4360] BGP community attributes, based on pre-defined values agreed between the operator and all client. Based on these community tags, prefixes may be propagated to all other clients, a subset of clients, or none. This mechanism allows route server clients to instruct the route server to implement per-client export routing policies.

As both standard and extended BGP communities values are restricted to 6 octets, the route server operator should take care to ensure that the predefined BGP community values mechanism used on their route server is compatible with [RFC4893] 4-octet autonomous system numbers.

4.6.2. Internet Routing Registry

Internet Routing Registry databases (IRRDBs) may be used by route server operators to implement construct per-client routing policies. [RFC2622] Routing Policy Specification Language (RPSL) provides an comprehensive grammar for describing interconnection relationships,

and several toolsets exist which can be used to translate RPSL policy description into route server configurations.

4.6.3. Client-accessible Databases

Should the route server operator not wish to use either BGP community tags or the public IRRDBs for implementing client export policies, they may implement their own routing policy database system for managing their clients' requirements. A database of this form SHOULD allow a route server client operator to update their routing policy and provide a mechanism for allowing the client to specify whether they wish to exchange all their prefixes with any other route server client. Optionally, the implementation may allow a client to specify unique routing policies for individual prefixes over which they have routing policy control.

4.7. Layer 2 Reachability Problems

Layer 2 reachability problems on an IXP can cause serious operational problems for IXP participants which depend on route servers for interconnection. Ethernet switch forwarding bugs have occasionally been observed to cause non-commutative reachability. For example, given a route server and two IXP participants, A and B, if the two participants can reach the route server but cannot reach each other, then traffic between the participants may be dropped until such time as the layer 2 forwarding problem is resolved. This situation does not tend to occur in bilateral interconnection arrangements, as the routing control path between the two hosts is usually (but not always, due to IXP inter-switch connectivity load balancing algorithms) the same as the data path between them.

Problems of this form can be dealt with using [RFC5881] bidirectional forwarding detection. However, as this is a bilateral protocol configured between routers, and as there is currently no means for automatic configuration of BFD between route server clients, BFD does not currently provide an optimal means of handling the problem.

5. Security Considerations

On route server installations which do not employ path hiding mitigation techniques, the path hiding problem outlined in section Section 4.1 can be used in certain circumstances to proactively block third party prefix announcements from other route server clients.

6. IANA Considerations

There are no IANA considerations.

7. Acknowledgments

The authors would like to thank Chris Hall, Ryan Bickhart and Steven Bakker for their valuable input.

In addition, the authors would like to acknowledge the developers of BIRD, OpenBGPD and Quagga, whose open source BGP implementations include route server capabilities which are compliant with this document.

8. References

8.1. Normative References

[I-D.ietf-idr-ix-bgp-route-server] Jasinska, E., Hilliard, N., Raszuk, R., and N. Bakker, "Internet Exchange Route Server", draft-ietf-idr-ix-bgp-route-server-02 (work in progress), February 2013.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

8.2. Informative References

[RFC1997] Chandrasekeran, R., Traina, P., and T. Li, "BGP Communities Attribute", RFC 1997, August 1996.

[RFC2622] Alaettinoglu, C., Villamizar, C., Gerich, E., Kessens, D., Meyer, D., Bates, T., Karrenberg, D., and M. Terpstra, "Routing Policy Specification Language (RPSL)", RFC 2622, June 1999.

[RFC4271] Rekhter, Y., Li, T., and S. Hares, "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, January 2006.

[RFC4360] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", RFC 4360, February 2006.

[RFC4456] Bates, T., Chen, E., and R. Chandra, "BGP Route Reflection: An Alternative to Full Mesh Internal BGP (IBGP)", RFC 4456, April 2006.

[RFC4893] Vohra, Q. and E. Chen, "BGP Support for Four-octet AS Number Space", RFC 4893, May 2007.

[RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.

[RFC5881] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD) for IPv4 and IPv6 (Single Hop)", RFC 5881, June 2010.

[RS-ARCH] Govindan, R., Alaettinoglu, C., Varadhan, K., and D. Estrin, "A Route Server Architecture for Inter-Domain Routing", 1995, <http://www.cs.usc.edu/research/95-603.ps.Z>.

Authors' Addresses

Nick Hilliard
INEX
4027 Kingswood Road
Dublin 24
IE

Email: nick@inex.ie

Elisa Jasinska
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052
US

Email: ejas@microsoft.com

Robert Raszuk
NTT I3
101 S Ellsworth Avenue Suite 350
San Mateo, CA 94401
US

Email: robert@raszuk.net

Niels Bakker
AMS-IX B.V.
Westeinde 12
Amsterdam, NH 1017 ZN
NL

Email: niels.bakker@ams-ix.net

GROW Working Group
Internet-Draft
Intended status: Informational
Expires: March 18, 2012

R. Raszuk, Ed.
NTT MCL
R. Fernando
K. Patel
Cisco Systems
D. McPherson
Verisign
K. Kumaki
KDDI Corporation

September 15, 2011

Distribution of diverse BGP paths.
draft-ietf-grow-diverse-bgp-path-dist-05

Abstract

The BGP4 protocol specifies the selection and propagation of a single best path for each prefix. As defined today BGP has no mechanisms to distribute paths other then best path between its speakers. This behaviour results in number of disadvantages for new applications and services.

This document presents an alternative mechanism for solving the problem based on the concept of parallel route reflector planes. Such planes can be build in parallel or they can co-exit on the current route reflection platforms. Document also compares existing solutions and proposed ideas that enable distribution of more paths than just the best path.

This proposal does not specify any changes to the BGP protocol definition. It does not require upgrades to provider edge or core routers nor does it need network wide upgrades. The authors believe that the GROW WG would be the best place for this work.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference

material or to cite them other than as "work in progress."

This Internet-Draft will expire on March 18, 2012.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction

Current BGP4 [RFC4271] protocol specification allows for the selection and propagation of only one best path for each prefix. The BGP protocol as defined today has no mechanism to distribute other then best path between its speakers. This behaviour results in a number of problems in the deployment of new applications and services.

This document presents an alternative mechanism for solving the problem based on the concept of parallel route reflector planes. It also compares existing solutions and proposed ideas that enable distribution of more paths than just the best path. The parallel route reflector planes solution brings very significant benefits at a negligible capex and opex deployment price as compared to the alternative techniques and is being considered by a number of network operators for deployment in their networks.

This proposal does not specify any changes to the BGP protocol definition. It does not require upgrades to provider edge or core routers nor does it need network wide upgrades. The only upgrade required is the new functionality on the new or current route reflectors. The authors believe that the GROW WG would be the best place for this work.

2. History

The need to disseminate more paths than just the best path is primarily driven by three requirements. First is the problem of BGP oscillations [I-D.ietf-idr-route-oscillation]. The second is the desire for reduction of time of reachability restoration in the event of network or network element's failure. Third requirement is to enhance BGP load balancing capabilities. Those reasons have lead to the proposal of BGP add-paths [I-D.ietf-idr-add-paths].

2.1. BGP Add-Paths Proposal

As it has been proven that distribution of only the best path of a route is not sufficient to meet the needs of continuously growing number of services carried over BGP the add-paths proposal was submitted in 2002 to enable BGP to distribute more then one path. This is achieved by including as a part of the NLRI an additional four octet value called the Path Identifier.

The implication of this change on a BGP implementation is that it must now maintain per path, instead of per prefix, peer advertisement state to track which of the peers each path was advertised to. This

new requirement has its own memory and processing cost. Suffice to say that by the end of 2009 none of the commercial BGP implementation could claimed to support the new add-path behaviour in production code, in major part due to this resource overhead.

An important observation is that distribution of more than one best path by Autonomous System Border Routers (ASBRs) with multiple EBGP peers attached to it where no "next hop self" is set may result in bestpath selection inconsistency within the autonomous system. Therefore it is also required to attach in the form of a new attribute the possible tie breakers and propagate those within the domain. The example of such attribute for the purpose of fast connectivity restoration to address that very case of ASBR injecting multiple external paths into the IBGP mesh has been presented and discussed in Fast Connectivity Restoration Using BGP Add-paths [I-D.ietf-idr-add-paths] document. Based on the additionally propagated information also best path selection is recommended to be modified to make sure that best and backup path selection within the domain stays consistent. More discussion on this particular point will be contained in the deployment considerations section below. In the proposed solution in this document we observe that in order to address most of the applications just use of best external advertisement is required. For ASBRs which are peering to multiple upstream ASs setting "next hop self" is recommended.

The add paths protocol extensions have to be implemented by all the routers within an AS in order for the system to work correctly. It remains quite a research topic to analyze benefits or risk associated with partial add-paths deployments. The risk becomes even greater in networks not using some form of edge to edge encapsulation.

The required code modifications include enhancements such as the Fast Connectivity Restoration Using BGP Add-path [I-D.pmohapat-idr-fast-conn-restore]. The deployment of such technology in an entire service provider network requires software and perhaps sometimes in the cases of End-of-Engineering or End-of-Life equipment even hardware upgrades. Such operation may or may not be economically feasible. Even if add-path functionality was available today on all commercial routing equipment and across all vendors, experience indicates that to achieve 100% deployment coverage within any medium or large global network may easily take years.

While it needs to be clearly acknowledged that the add-path mechanism provides the most general way to address the problem of distributing many paths between BGP speakers, this document provides a much easier to deploy solution that requires no modification to the BGP protocol where only a few additional paths may be required. The alternative

method presented is capable of addressing critical service provider requirements for disseminating more than a single path across an AS with a significantly lower deployment cost.

3. Goals

The proposal described in this document is not intended to compete with add-paths. Instead if deployed it is to be used as a very easy method to accommodate the majority of applications which may require presence of alternative BGP exit points.

It is presented to network operators as a possible choice and provides those operators who need additional paths today an alternative from the need to transition to a full mesh.

It is intended as a way to buy more time allowing for a smoother and gradual migration where router upgrades will be required for perhaps different reasons. It will also allow the time required where standard RP/RE memory size can easily accommodate the associated overhead with other techniques without any compromises.

4. Multi plane route reflection

The idea contained in the proposal assumes the use of route reflection within the network. Other techniques as described in the following sections already provide means for distribution of alternate paths today.

Let's observe today's picture of simple route reflected domain:

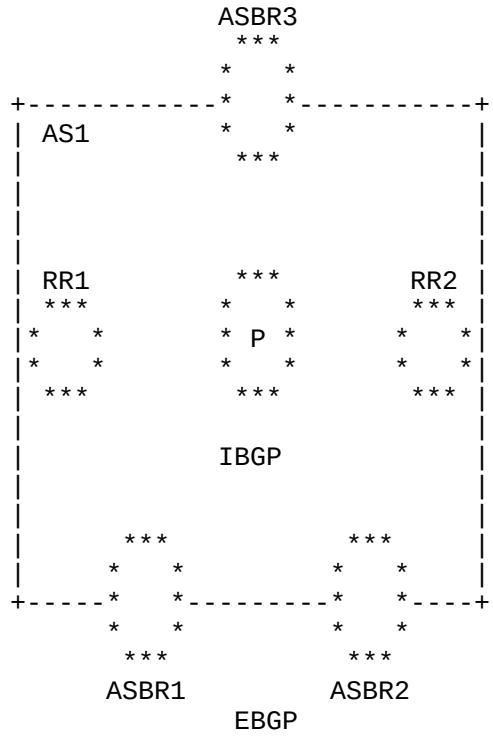


Figure1: Simple route reflection

Figure 1 shows an AS that is connected via EBGP peering at ASBR1 and ASBR2 to an upstream AS or set of ASes. For a given destination "D" ASBR1 and ASBR2 will each have an external path P1 and P2 respectively. The AS network uses two route reflectors RR1 and RR2 for redundancy reasons. The route reflectors propagate the single BGP best path for each route to all clients. All ASBRs are clients of RR1 and RR2.

Below are the possible cases of the path information that ASBR3 may receive from route reflectors RR1 and RR2:

1. When best path tie breaker is the IGP distance: When paths P1 and P2 are considered to be equally good best path candidates the selection will depend on the distance of the path next-hops from the route reflector making the decision. Depending on the positioning of the route reflectors in the IGP topology they may choose the same best path or a different one. In such a case

ASBR3 may receive either the same path or different paths from each of the route reflectors.

2. When best path tie breaker is Multi-Exit-Discriminator or Local Preference: In this case only one path from preferred exit point ASBR will be available to RRs since the other peering ASBR will consider the IBGP path as best and will not announce (or if already announced will withdraw) its own external path. The exception here is the use of BGP Best-External proposal which will allow stated ASBR to still propagate to the RRs its own external path. Unfortunately RRs will not be able to distribute it any further to other clients as only the overall best path will be reflected.

The proposed solution is based on the use of additional route reflectors or new functionality enabled on the existing route reflectors that instead of distributing the best path for each route will distribute an alternative path other then best. The best path (main) reflector plane distributes the best path for each route as it does today. The second plane distributes the second best path for each route and so on. Distribution of N paths for each route can be achieved by using N reflector planes.

As diverse-path functionality may be enabled on a per peer basis one of the deployment model can be realized to continue advertisement of overall best path from both route reflectors while in addition new session can be provisioned to get additional path. That will allow the non interupted use of best path even if one of the RRs goes down provided that the overall best path is still a valid one.

Each plane of route reflectors is a logical entity and may or may not be co-located with the existing best path route reflectors. Adding a route reflector plane to a network may be as easy as enabling a logical router partition, new BGP process or just a new configuration knob on an existing route reflector and configuring an additional IBGP session from the current clients if required. There are no code changes required on the route reflector clients for this mechanism to work. It is easy to observe that the installation of one or more additional route reflector control planes is much cheaper and an easier than the need of upgrading 100s of route reflector clients in the entire network to support different bgp protocol encoding.

Diverse path route reflectors need the new ability to calculate and propagate the Nth best path instead of the overall best path. An implementation is encouraged to enable this new functionality on a per neighbor basis.

While this is an implementation detail, the code to calculate Nth

best path is also required by other BGP solutions. For example in the application of fast connectivity restoration BGP must calculate a backup path for installation into the RIB and FIB ahead of the actual failure.

To address the problem of external paths not being available to route reflectors due to local preference or MED factors it is recommended that ASBRs enable the best-external functionality in order to always inject their external paths to the route reflectors.

4.1. Co-located best and backup path RRs

To simplify the description let's assume that we only use two route reflector planes (N=2). When co-located the additional 2nd best path reflectors are connected to the network at the same points from the perspective of the IGP as the existing best path RRs. Let's also assume that best-external is enabled on all ASBRs.

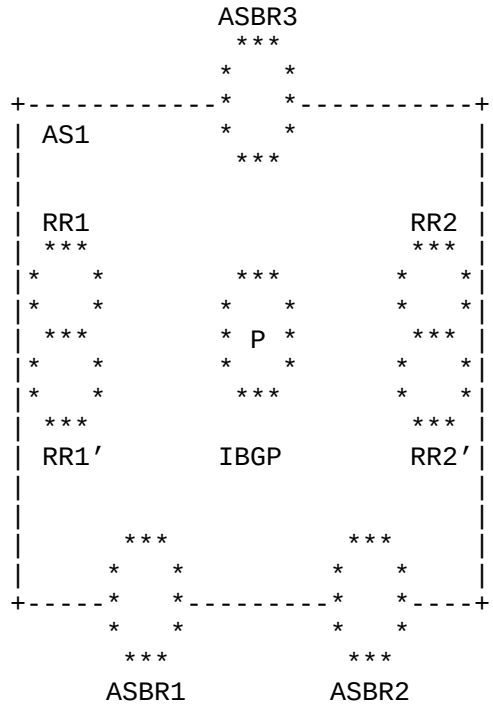


Figure2: Co-located 2nd best RR plane

The following is a list of configuration changes required to enable the 2nd best path route reflector plane:

1. Unless same RR1/RR2 platform is being used adding RR1' and RR2' either as logical or physical new control plane RRs in the same IGP points as RR1 and RR2 respectively.

2. Enabling best-external on ASBRs

3. Enabling RR1' and RR2' for 2nd plane route reflection. Alternatively instructing existing RR1 and RR2 to calculate also 2nd best path.

4. Unless one of the existing RRs is turned to advertise only diverse path to it's current clients configuring new ASBRs-RR' IBGP sessions

The expected behaviour is that under any BGP condition the ASBR3 and P routers will receive both paths P1 and P2 for destination D. The availability of both paths will allow them to implement a number of new services as listed in the applications section below.

As an alternative to fully meshing all RRs and RRs' an operator who has a large number of reflectors deployed today may choose to peer newly introduced RRs' to a hierarchical RR' which would be an IBGP interconnect point within the 2nd plane as well as between planes.

One of the deployment model of this scenario can be achieved by simple upgrade of the existing route reflectors without the need to deploy any new logical or physical platforms. Such upgrade would allow route reflectors to service both upgraded to add-paths peers as well as those peers which can not be immediately upgraded while in the same time allowing to distribute more then single best path. The obvious protocol benefit of using existing RRs to distribute towards their clients best and diverse bgp paths over different IBGP session is the automatic assurance that such client would always get different paths with their next hop being different.

The way to accomplish this would be to create a separate IBGP session for each N-th BGP path. Such session should be preferably terminated at a different loopback address of the route reflector. At the BGP OPEN stage of each such session a different bgp_router_id may be used. Correspondingly route reflector should also allow its clients to use the same bgp_router_id on each such session.

4.2. Randomly located best and backup path RRs

Now let’s consider a deployment case where an operator wishes to enable a 2nd RR’ plane using only a single additional router in a different network location to his current route reflectors. This model would be of particular use in networks where some form of end-to-end encapsulation (IP or MPLS) is enabled between provider edge routers.

Note that this model of operation assumes that the present best path route reflectors are only control plane devices. If the route reflector is in the data forwarding path then the implementation must be able to clearly separate the Nth best-path selection from the selection of the paths to be used for data forwarding. The basic premise of this mode of deployment assumes that all reflector planes have the same information to choose from which includes the same set of BGP paths. It also requires the ability to ignore the step of comparison of the IGP metric to reach the bgp next hop during best-path calculation.

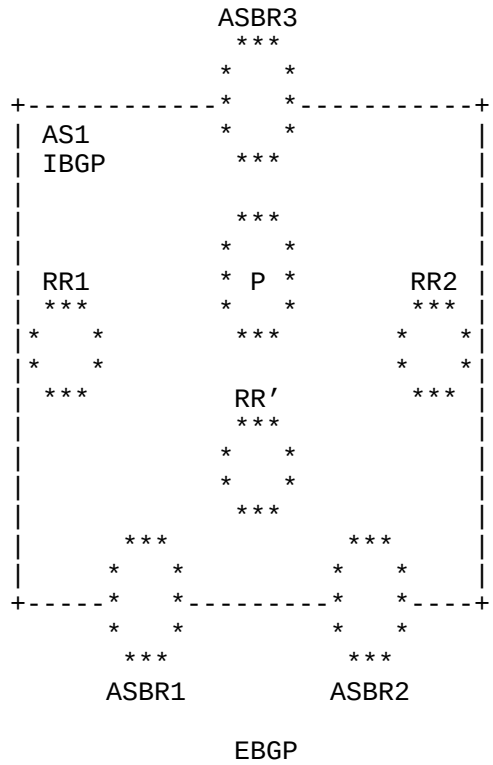


Figure3: Experimental deployment of 2nd best RR

The following is a list of configuration changes required to enable the 2nd best path route reflector RR' as a single platform or to enable one of the existing control plane RRs for diverse-path functionality:

1. If needed adding RR' logical or physical as new route reflector anywhere in the network

2. Enabling best-external on ASBRs

3. Disabling IGP metric check in BGP best path on all route reflectors.

4. Enabling RR' or any of the existing RR for 2nd plane path calculation

5. If required fully meshing newly added RRs' with the all other reflectors in both planes. That condition does not apply if the

newly added RR'(s) already have peering to all ASBRs/PEs.

6. Unless one of the existing RRs is turned to advertise only diverse path to it's current clients configuring new ASBRs-RR' IBGP sessions

In this scenario the operator has the flexibility to introduce the new additional route reflector functionality on any existing or new hardware in the network. Any of the existing routers that are not already members of the best path route reflector plane can be easily configured to serve the 2nd plane either via using a logical / virtual router partition or by having their bgp implementation compliant to this specification.

Even if the IGP metric is not taken into consideration when comparing paths during the bestpath calculation, an implementation still has to consider paths with unreachable nexthops as invalid. It is worth pointing out that some implementations today already allow for configuration which results in no IGP metric comparison during the best path calculation.

The additional planes of route reflectors do not need to be fully redundant as the primary one does. If we are preparing for a single network failure event, a failure of a non backed up N-th best-path route reflector would not result in an connectivity outage of the actual data plane. The reason is that this would at most affect the presence of a backup path (not an active one) on same parts of the network. If the operator chooses to build the N-th best path plane redundantly by installing not one, but two or more route reflectors serving each additional plane the additional robustness will be achieved.

As a result of this solution ASBR3 and other ASBRs peering to RR' will be receiving the 2nd best path.

Similarly to section 4.1 as an alternative to fully meshing all RRs & RRs' an operator who may have a large number of reflectors already deployed today may choose to peer newly introduced RRs' to a hierarchical RR' which would be an IBGP interconnect point between planes.

4.3. Multi plane route servers for Internet Exchanges

Another group of devices where the proposed multi-plane architecture may be of particular applicability are EBGP route servers used at many of internet exchange points.

In such cases 100s of ISPs are interconnected on a common LAN.

Instead of having 100s of direct EBGP sessions on each exchange client, a single peering is created to the transparent route server. The route server can only propagate a single best path. Mandating the upgrade for 100s of different service providers in order to implement add-path may be much more difficult as compared to asking them for provisioning one new EBGP session to an Nth best-path route server plane. That will allow to distribute more then single best BGP path from a given route server to such IX peer.

The solution proposed in this document fits very well with the requirement of having broader EBGP path diversity among the members of any Internet Exchange Point.

5. Discussion on current models of IBGP route distribution

In today's networks BGP4 operates as specified in [RFC4271]

There are a number of technology choices for intra-AS BGP route distribution:

1. Full mesh
2. Confederations
3. Route reflectors

5.1. Full Mesh

A full mesh, the most basic iBGP architecture, exists when all the BGP speaking routers within the AS peer directly with all other BGP speaking routers within the AS, irrespective of where a given router resides within the AS (e.g., P router, PE router, etc..).

While this is the simplest intra-domain path distribution method, historically there have been a number of challenges in realizing such an IBGP full mesh in a large scale network. While some of these challenges are no longer applicable today some may still apply, to include the following:

1. Number of TCP sessions: The number of IBGP sessions on a single router in a full mesh topology of a large scale service provider can easily reach 100s. While on hardware and software used in the late 70s, 80s and 90s such numbers could be of concern, today customer requirements for the number of BGP sessions per box are reaching 1000s. This is already an order of magnitude more then the potential number of IBGP sessions. Advancement in hardware and software used in production routers mean that running a full

mesh of IBGP sessions should not be dismissed due to the resulting number of TCP sessions alone.

2. Provisioning: When operating and troubleshooting large networks one of the top-most requirements is to keep the design as simple as possible. When the autonomous systems network is composed of hundreds of nodes it becomes very difficult to manually provision a full mesh of IBGP sessions. Adding or removing a router requires reconfiguration of all the other routers in the AS. While this is a real concern today there is already work in progress in the IETF to define IBGP peering automation through an IBGP Auto Discovery [I-D.raszuk-idr-ibgp-auto-mesh] mechanism.

3. Number of paths: Another concern when deploying a full IBGP mesh is the number of BGP paths for each route that have to be stored at every node. This number is very tightly related to the number of external peerings of an AS, the use of local preference or multi-exit-discriminator techniques and the presence of best-external [I-D.ietf-idr-best-external] advertisement configuration. If we make a rough assumption that the BGP4 path data structure consumes about 80-100 bytes the resulting control plane memory requirement for 500,000 IPv4 routes with one additional external path is 38-48 MB while for 1 million IPv4 routes it grows linearly to 76-95 MB. It is not possible to reach a general conclusion if this condition is negligible or if it is a show stopper for a full mesh deployment without direct reference to a given network.

To summarize, a full mesh IBGP peering can offer natural dissemination of multiple external paths among BGP speakers. When realized with the help of IBGP Auto Discovery peering automation this seems like a viable deployment especially in medium and small scale networks.

5.2. Confederations

For the purpose of this document let’s observe that confederations [RFC5065] can be viewed as a hierarchical full mesh model.

Within each sub-AS BGP speakers are fully meshed and as discussed in section 2.1 all full mesh characteristics (number of TCP sessions, provisioning and potential concern over number of paths still apply in the sub-AS scale).

In addition to the direct peering of all BGP speakers within each sub-AS, all sub-AS border routers must also be fully meshed with each other. Sub-AS border routers configured with best-external functionality can inject additional exit paths within a sub-AS.

To summarize, it is technically sound to use confederations with the combination of best-external to achieve distribution of more than a single best path per route in a large autonomous systems.

In topologies where route reflectors are deployed within the confederation sub-ASes the technique describe here does apply.

5.3. Route reflectors

The main motivation behind the use of route reflectors [RFC4456] is the avoidance of the full mesh session management problem described above. Route reflectors, for good or for bad, are the most common solution today for interconnecting BGP speakers within an internal routing domain.

Route reflector peerings follow the advertisement rules defined by the BGP4 protocol. As a result only a single best path per prefix is sent to client BGP peers. That is the main reason why many current networks are exposed to a phenomenon called BGP path starvation which essentially results in inability to deliver a number of applications discussed later.

The route reflection equivalent when interconnecting BGP speakers between domains is popularly called the Route Server and is globally deployed today in many internet exchange points.

6. Deployment considerations

The diverse BGP path dissemination proposal allows the distribution of more paths than just the best-path to route reflector or route server clients of today’s BGP4 implementations.

From the client’s point of view receiving additional paths via separate IBGP sessions terminated at the new router reflector plane is functionally equivalent to constructing a full mesh peering without the problems that such a full mesh would come with set of problems as discussed in earlier section.

By precisely defining the number of reflector planes, network operators have full control over the number of redundant paths in the network. This number can be defined to address the needs of the service(s) being deployed.

The Nth plane route reflectors should be acting as control plane network entities. While they can be provisioned on the current production routers selected Nth best BGP paths should not be used directly in the date plane with the exception of such paths being BGP

multipath eligible and such functionality is enabled. On RRs being in the data plane unless multipath is enabled 2nd best path is expected to be a backup path and should be installed as such into local RIB/FIB.

The proposed architecture deployed along with the BGP best-external functionality covers all three cases where the classic BGP route reflection paradigm would fail to distribute alternate exit points paths.

1. ASBRs advertising their single best external paths with no local-preference or multi-exit-discriminator present.

2. ASBRs advertising their single best external paths with local-preference or multi-exit-discriminator present and with BGP best-external functionality enabled.

3. ASBRs with multiple external paths.

Let’s discuss the 3rd above case in more detail. This describes the scenario of a single ASBR connected to multiple EBGP peers. In practice this peering scenario is quite common. It is mostly due to the geographic location of EBGP peers and the diversity of those peers (for example peering to multiple tier 1 ISPs etc...). It is not designed for failure recovery scenarios as single failure of the ASBR would simultaneously result in loss of connectivity to all of the peers. In most medium and large geographically distributed networks there is always another ASBR or multiple ASBRs providing peering backups, typically in other geographically diverse locations in the network.

When an operator uses ASBRs with multiple peerings setting next hop self will effectively allow to locally repair the atomic failure of any external peer without any compromise to the data plane. The most common reason for not setting next hop self is traditionally the associated drawback of loosing ability to signal the external failures of peering ASBRs or links to those ASBRs by fast IGP flooding. Such potential drawback can be easily avoided by using different peering address from the address used for next hop mapping as well as removing such next hop from IGP at the last possible BGP path failure.

Herein one may correctly observe that in the case of setting next hop self on an ASBR, attributes of other external paths such ASBR is peering with may be different from the attributes of its best external path. Therefore, not injecting all of those external paths with their corresponding attribute can not be compared to equivalent paths for the same prefix coming from different ASBRs.

While such observation in principle is correct one should put things in perspective of the overall goal which is to provide data plane connectivity upon a single failure with minimal interruption/packet loss. During such transient conditions, using even potentially suboptimal exit points is reasonable, so long as forwarding information loops are not introduced. In the mean time BGP control plane will on its own re-advertise newly elected best external path, route reflector planes will calculate their Nth best paths and propagate to its clients. The result is that after seconds even if potential sub-optimality were encountered it will be quickly and naturally healed.

7. Summary of benefits

The diverse BGP path dissemination proposal provides the following benefits when compared to the alternatives:

1. No modifications to BGP4 protocol.

2. No requirement for upgrades to edge and core routers. Backward compatible with the existing BGP deployments.

3. Can be easily enabled by introduction of a new route reflector, route server plane dedicated to the selection and distribution of Nth best-path or just by new configuration of the upgraded current route reflector(s).

4. Does not require major modification to BGP implementations in the entire network which will result in an unnecessary increase of memory and CPU consumption due to the shift from today’s per prefix to a per path advertisement state tracking.

5. Can be safely deployed gradually on a RR cluster basis.

6. The proposed solution is equally applicable to any BGP address family as described in Multiprotocol Extensions for BGP-4 RFC4760 [RFC4760]. In particular it can be used "as is" without any modifications to both IPv4 and IPv6 address families.

8. Applications

This section lists the most common applications which require presence of redundant BGP paths:

1. Fast connectivity restoration where backup paths with alternate exit points would be pre-installed as well as pre-resolved in the FIB of routers. That would allow for a local action upon reception of a critical event notification of network / node failure. This failure recovery mechaism based on the presence of backup paths is also suitable for gracefully addressing scheduled maintenane requirements as described in [I-D.decraene-bgp-graceful-shutdown-requirements].

2. Multi-path load balancing for both IBGP and EBGP.

3. BGP control plane churn reduction both intra-domain and inter-domain.

An important point to observe is that all of the above intra-domain applications based on the use of reflector planes but are also applicable in the inter-domain Internet exchange point examples. As discussed in section 4.3 an internet exchange can conceptually deploy shadow route server planes each responsible for distribution of an Nth best path to its EBGP peers. In practice it may just equal to new short configuration and establishment of new BGP sessions to IX peers.

9. Security considerations

The new mechanism for diverse BGP path dissemination proposed in this document does not introduce any new security concerns as compared to base BGP4 specification [RFC4271].

10. IANA Considerations

The new mechanism for diverse BGP path dissemination does not require any new allocations from IANA.

11. Contributors

The following people contributed significantly to the content of the document:

Selma Yilmaz
Cisco Systems
170 West Tasman Drive
San Jose, CA 95134
US
Email: seyilmaz@cisco.com

Satish Mynam
Cisco Systems
170 West Tasman Drive
San Jose, CA 95134
US
Email: mynam@cisco.com

Isidor Kouvelas
Cisco Systems
170 West Tasman Drive
San Jose, CA 95134
US
Email: kouvelas@cisco.com

12. Acknowledgments

The authors would like to thank Bruno Decraene, Bart Peirens, Eric Rosen, Jim Uttaro, Renwei Li and George Wes for their valuable input.

The authors would also like to express special thank you to number of operators who helped to optimize the provided solution to be as close as possible to their daily operational practices. Especially many thx goes to Ted Seely, Shan Amante, Benson Schliesser and Seiichi Kawamura.

13. References

13.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC4271] Rekhter, Y., Li, T., and S. Hares, "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, January 2006.

[RFC4760] Bates, T., Chandra, R., Katz, D., and Y. Rekhter, "Multiprotocol Extensions for BGP-4", RFC 4760,

January 2007.

[RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.

13.2. Informative References

[I-D.decraene-bgp-graceful-shutdown-requirements]
Decraene, B., Francois, P., pelsser, c., Ahmad, Z., and A. Armengol, "Requirements for the graceful shutdown of BGP sessions", draft-decraene-bgp-graceful-shutdown-requirements-01 (work in progress), March 2009.

[I-D.ietf-idr-add-paths]
Walton, D., Chen, E., Retana, A., and J. Scudder, "Advertisement of Multiple Paths in BGP", draft-ietf-idr-add-paths-05 (work in progress), July 2011.

[I-D.ietf-idr-best-external]
Marques, P., Fernando, R., Chen, E., Mohapatra, P., and H. Gredler, "Advertisement of the best external route in BGP", draft-ietf-idr-best-external-04 (work in progress), April 2011.

[I-D.ietf-idr-route-oscillation]
McPherson, D., "BGP Persistent Route Oscillation Condition", draft-ietf-idr-route-oscillation-01 (work in progress), February 2002.

[I-D.pmohapat-idr-fast-conn-restore]
Mohapatra, P., Fernando, R., Filsfils, C., and R. Raszuk, "Fast Connectivity Restoration Using BGP Add-path", draft-pmohapat-idr-fast-conn-restore-01 (work in progress), March 2011.

[I-D.raszuk-idr-ibgp-auto-mesh]
Raszuk, R., "IBGP Auto Mesh", draft-raszuk-idr-ibgp-auto-mesh-00 (work in progress), June 2003.

[RFC4456] Bates, T., Chen, E., and R. Chandra, "BGP Route Reflection: An Alternative to Full Mesh Internal BGP (IBGP)", RFC 4456, April 2006.

[RFC5065] Traina, P., McPherson, D., and J. Scudder, "Autonomous System Confederations for BGP", RFC 5065, August 2007.

Authors' Addresses

Robert Raszuk (editor)
NTT MCL
101 S Ellsworth Avenue Suite 350
San Mateo, CA 94401
US

Email: robert@raszuk.net

Rex Fernando
Cisco Systems
170 West Tasman Drive
San Jose, CA 95134
US

Email: rex@cisco.com

Keyur Patel
Cisco Systems
170 West Tasman Drive
San Jose, CA 95134
US

Email: keyupate@cisco.com

Danny McPherson
Verisign
21345 Ridgetop Circle
Dulles, VA 20166
US

Email: dmcpherson@verisign.com

Kenji Kumaki
KDDI Corporation
Garden Air Tower
Iidabashi, Chiyoda-ku, Tokyo 102-8460
Japan

Email: ke-kumaki@kddi.com

Network Working Group
Internet-Draft
Intended status: Informational
Expires: January 2, 2012

P. Francis
MPI-SWS
X. Xu
Huawei
H. Ballani
Cornell U.
D. Jen
UCLA
R. Raszuk
Cisco
L. Zhang
UCLA

July 1, 2011

Auto-Configuration in Virtual Aggregation
draft-ietf-grow-va-auto-04.txt

Abstract

Virtual Aggregation as specified in [I-D.ietf-grow-va] requires configuration of a static "VP-List" on all routers. The VP-List allows routers to know which prefixes may or may not be FIB-installed. This draft specified an optional method of determining this that requires far less configuration. Specifically, it requires the configuration of a "VP-Range" in ASBRs connected to transit and peer ISPs. A Non-transitive Extended Communities Attribute is used to convey to other routers that a given route can be FIB-suppressed.

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on January 2, 2012.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction

As the current VA specification stands ([I-D.ietf-grow-va]), routers have to know which prefixes they must FIB-install and which they need not FIB-install. The VP-List tells them this: they must FIB-install routes to Virtual Prefixes (VP), and they need not FIB-install routes to prefixes that fall within VPs for which they are not an Aggregation Point Router (APR). The same VP-List must be installed in every router.

This draft specifies an optional alternative to the VP-List that requires far less configuration. Specifically, a list of one or more "VP-Ranges" is configured in ASBRs --- typically ASBRs that do not connect to customer networks. These ASBRs then simply tag routes as to whether the route can be suppressed. This is simpler than the current configured VP-List approach in two regards. First, fewer routers need to be configured. Second, the VP-Range is simpler than the VP-List. In most cases, once an ISP is past its initial VA roll-out phase, the VP-Range consists of a single 0/0 entry.

This draft uses terms defined in [I-D.ietf-grow-va].

1.1. Requirements notation

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 [RFC2119].

2. Specification

With the "VP-Range" approach to determining suppressability, certain ASBRs are designated as "tagging routers". Tagging routers explicitly tag routes with a Non-transitive Extended Communities Attribute that indicates whether the route can be FIB-suppressed. All ASBRs that connect to one or more transit provider ISPs MUST be tagging routers. ASBRs that connect to one or more peer ISPs SHOULD be tagging routers. ASBRs that connect to customer networks SHOULD NOT be tagging routers.

Tagging routers are configured with a "VP-Range" list. This consists of the ranges of IP address that are collectively covered by all VPs in the AS. In a mature deployment of VA, the range would amount to all IP addresses, in which case the VP-Range is simply 0/0. Early in VA deployment, when an ISP is still in the testing or roll-out phase, the VP-Range may consist of multiple entries.

Tagging routers SHOULD tag any route whose prefix falls within the

VP-Range with a "can-suppress" tag, with the following exceptions:

1. Tagging routers MUST NOT tag VP routes with can-suppress (where a VP route is that route to the VP that the router originates in its role as an APR).
2. If the ISP has a policy of FIB-installing customer routes, then routes received from customers SHOULD NOT be tagged with can-suppress.

The can-suppress tag itself is an Extended Communities Attribute [RFC4360] to be assigned by IANA from the "well-known" pool define in [I-D.ietf-idr-reserved-extended-communities]. The Transitive Bit MUST be set to value 1 (the community is non-transitive across ASes).

Routers install or suppress FIB entries according to the following rules. Note that tagging routers conceptually follow these rules after tagging (or not tagging) the route. Note also that these rules apply only to the route used by the router as the best route. In other words, if a router receives two routes for the same prefix, and one route is tagged can-suppress and the other is not, the router follows these rules only with respect to the route that it selects as the best route.

1. Routes without the can-suppress tag MUST be FIB-installed.
2. APRs MUST FIB-install routes for sub-prefixes that fall within the APR’s VPs, whether or not the route is tagged can-suppress.
3. Otherwise, routers MAY FIB-suppress routes tagged as can-suppress.

3. IANA Considerations

IANA is requested to assign, from the registry "BGP Assigned non-transitive extended communities", a value TBD for "VA can suppress":

Registry Name: BGP Assigned non-transitive extended communities

Name	Type Value
VA can suppress	TBD

4. Security Considerations

As of this writing, there are no known new security threats introduced by this draft.

5. Acknowledgements

The authors would like to thank Wes George and Bruno Decraene for their reviews and suggestions.

6. References

6.1. Normative References

[I-D.ietf-grow-va]
Francis, P., Xu, X., Ballani, H., Jen, D., Raszuk, R., and L. Zhang, "FIB Suppression with Virtual Aggregation", draft-ietf-grow-va-04 (work in progress), Oct 2009.

[I-D.ietf-idr-reserved-extended-communities]
Decraene, B. and P. Francois, "Assigned BGP extended communities", draft-ietf-idr-reserved-extended-communities-01 (work in progress), May 2011.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC4360] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", RFC 4360, February 2006.

6.2. Informative References

Authors' Addresses

Paul Francis
Max Planck Institute for Software Systems
Gottlieb-Daimler-Strasse
Kaiserslautern 67633
Germany

Phone: +49 631 930 39600
Email: francis@mpi-sws.org

Xiaohu Xu
Huawei Technologies
No.3 Xinxi Rd., Shang-Di Information Industry Base, Hai-Dian District
Beijing, Beijing 100085
P.R.China

Phone: +86 10 82836073
Email: xuxh@huawei.com

Hitesh Ballani
Cornell University
4130 Upson Hall
Ithaca, NY 14853
US

Phone: +1 607 279 6780
Email: hitesh@cs.cornell.edu

Dan Jen
UCLA
4805 Boelter Hall
Los Angeles, CA 90095
US

Phone:
Email: jenster@cs.ucla.edu

Robert Raszuk
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134
USA

Phone:
Email: raszuk@cisco.com

Lixia Zhang
UCLA
3713 Boelter Hall
Los Angeles, CA 90095
US

Phone:
Email: lixia@cs.ucla.edu

Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: December 29, 2011

S. Tsuchiya, Ed.
Cisco Systems
S. Kawamura
NEC BIGLOBE, Ltd.
R. Bush
C. Pelsser
Internet Initiative Japan, Inc.
June 27, 2011

Route Flap Damping Deployment Status Survey
draft-shishio-grow-isp-rfd-implement-survey-02

Abstract

BGP Route Flap Damping [RFC2439] is a mechanism that targets route stability. It penalyzes routes that flap with the aim of reducing CPU load on the routers.

But it has side-effects. Thus, in 2006, RIPE recommended not to use Route Flap Damping (see [RIPE-378]).

Now, some researchers propose to turn RFD, with less aggressive parameters, back on [draft-ymbk-rfd-usable].

This document describes results of a survey conducted amoung service provider on their use of BGP Route Flap Damping.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on December 29, 2011.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the

document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Survey Purpose

RIPE published some recommendations such as [RIPE-178],[RIPE-210],[RIPE-229] and [RIPE-378].

The purpose of this survey is to understand the current usage and requirements of Route Flap Damping [RFC2439] among service providers.

2. Survey's target and period

2.1. For Japan

Target: Japan Network Operator Group janog@janog.gr.jp

Period: Jan 28,2011 - Feb 12,2011

2.2. For Global

Target: All operators who has answered the survey https://www.surveymonkey.com/s/rfd-survey.

We posted this document to the following mailing list.

North American Network Operators Group nanog@nanog.org
RIPE Routing Working Group routing-wg@ripe.net
Asia Pacific OperatorS Forum apops@apops.net
Africa Network Operators Group afnog@afnog.org
South Asian Network Operators Group sanog@sanog.org
Latin America and Caribbean Region Network Operators Group lacnog@lacnic.net

Period:Mar 7,2011 - May 25,2011

3. Survey Results

3.1. Q1.Which is the best description of your job role?

3.1.1. Japan

This question did not exist for Japan version.

3.1.2. Global

BGP operator:27
Researcher:1
Engineer of vendor:3
Engineer of Network/System Integrator:13
Student:0
Other:0

3.2. Q2.Do you use Route Flap Damping ?

Answer	Japan	Global	Total Number	Percentage[%]
YES	5	8	13	20.6
NO	8	36	49	77.8
Skipped Q2.	1	0	1	1.6

3.3. Q3.If you select No on Q2,why?

Answer	Japan	Global	Total Number	Percentage[%]
Do not have the need	3	7	10	19.6
Did not know about the feature	2	3	5	9.8
No benefits expected	3	7	10	19.6
Customers would complain	1	4	5	9.8
Because I read [RIPE-378]	2	13	15	29.4
Other	3	3	6	11.8

1 person answered Q3,even if he selected "Yes" on Q2.

3.4. Q4.If you select Yes on Q2,what parameter do you use?

Answer	Japan	Global	Total Number	Percentage[%]
Default parameters	3	3	6	40.0
[RIPE-178]	0	1	1	6.7
[RIPE-210]	0	0	0	0.0
[RIPE-229]	0	1	1	6.7
Other	3	4	7	46.7

1 person answered Q4, even if he selected "No" on Q2.

3.5. Q5.Do you know Randy Bush et. al's report ''Route Flap Damping Considered Usable?''

Answer	Japan	Global	Total Number	Percentage[%]
YES	12	21	33	52.4
NO	7	22	29	46.0
Skipped Q5.	0	1	1	1.6

One person skipped Q2, but answered Q5.

3.6. Q6.IOS's max-penalty is currently limited to 20K. Do you need this limitation to be relaxed to over 50K?

Answer	Japan	Global	Total Number	Percentage[%]
YES	10	14	24	38.1
NO	9	23	32	50.8
Skipped Q6.	0	7	7	11.1

3.7. Q7.According to [draft-ymbk-rfd-usable],Suppress Threshold should be set to 6K.Do you think the default value on implementations should be changed to 6K?''

Answer	Japan	Global	Total Number	Percentage[%]
YES	N/A	17	17	38.6
NO	N/A	18	18	40.9
Skipped Q7.	N/A	9	9	20.5

This question did not exist for Japan version.

3.8. Q8.If you have any comments, please fill this box.

Free format

3.8.1. Japan

-Our peer seems to have damping enabled, and our prefix gets damped sometimes.

-We do not enable damping because we think that customers want a non-damped route.

-From the perspective of a downstream ISP, if our upstream told us that an outage occurred because a route was damped, I may call and ask "is it written in the agreement that you will do this?"

-We use damping pretty heavily

-I had RFD turned on until this morning when I discovered our router has CSCtd26215 issues. I would like to turn on a "useful" RFD.

3.8.2. Global

-Statistical reports from big Service Providers may better visualize the situation.

-best current practices is nice, but always needs to be adjusted to reflect local network settings.

-We used RFD in the past and came to the conclusion that we do not want to use RFD any more. We still have it configured to be able to get Flap statistics out of our Cisco boxes, but no prefixes get dampended

-We recently removed all RFD from the configs due to the information read on the topic among the preso’s on the NANOG Archive.

-after seeing this survey, I read the draft; sounds promising; would be nice to see vendors start to implement it.

-Q3, other: Juniper RFD is broken, default values count penalty for both update and withdrawal, and they would not fix that. No clear motivation for us, has caused outage when our customers (with primiary and backup connection to us) had a flapping link.

-Strong desire to see the path vector penalized rather than the prefix.

4. Summary of data

From the survey we see that there are many service providers with RFD disabled. The reason varies among providers, but it is clear that there are those who wish that RFD was made useful. [draft-ymbk-rfd-usable] describes how to improve RFD with minor changes to some parameters. From the comments in the survey, the most significant fear of enabling RFD is its impact on customers.

5. Acknowledgements

We thank the 63 respondant to this survey.

6. IANA Considerations

This document has no actions for IANA.

7. Security Considerations

This document has no security considerations.

8. References

8.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC2439] Villamizar, C., Chandra, R., and R. Govindan, "BGP Route Flap Damping", RFC 2439, November 1998.

8.2. Informative References

[I-D.ymbk-rfd-usable]
Pelsser, C., Bush, R., Patel, K., Mohapatra, P., and O. Maennel, "Making Route Flap Damping Usable", draft-ymbk-rfd-usable-00 (work in progress), March 2011.

[RIPE-178]
Barber, T., Doran, S., Panigl, C., and J. Schmitz, ""RIPE Routing-WG Recommendation for coor-dinated route-flap damping parameters"", Feb 1998, <ftp://ftp.ripe.net/ripe/docs/ripe-178.txt>.

[RIPE-210]
Barber, T., Doran, S., Karrenberg, D., Panigl, C., and J. Schmitz, ""RIPE Routing-WG Recommendation for coordinated route-flap damping parameters"", May 2000, <ftp://ftp.ripe.net/ripe/docs/ripe-210.txt>.

[RIPE-229]
Panigl, C., Schmitz, J., Smith, P., and C. Vistoli, ""RIPE Routing-WG Recommendations for Coordinated Route-flap Damping Parameters"", Oct 2001,

<ftp://ftp.ripe.net/ripe/docs/ripe-229.txt>.

[RIPE-378]
Smith, P. and C. Panigl, ""RIPE Routing Working Group Recommendations On Route-flap Damping"", May 2006, <http://www.ripe.net/ripe/docs/ripe-378>.

[Route Flap Damping Considered Usable?]
Pelsser, C., Maennel, O., Patel, K., and R. Bush, ""Route Flap Damping Considered Useable"", Nov 2011, <http://ripe61.ripe.net/presentations/222-101117.ripe-rfd.pdf>.

Appendix A. Additional Stuff

This becomes an Appendix.

Authors' Addresses

Shishio Tsuchiya (editor)
Cisco Systems
Shinjuku Mitsui Building, 2-1-1, Nishi-Shinjuku
Shinjuku-Ku, Tokyo 163-0409
Japan

Phone: +81 3 6434 6543
Email: shtsuchi@cisco.com

Seiichi Kawamura
NEC BIGLOBE, Ltd.
14-22, Shibaura 4-chome
Minatoku, Tokyo 108-8558
JAPAN

Phone: +81 3 3798 6085
Email: kawamucho@mesh.ad.jp

Randy Bush
Internet Initiative Japan, Inc.
5147 Crystal Springs
Bainbridge Island, Washington 98110
US

Phone: +1 206 780 0431 x1
Email: randy@psg.com

Cristel Pelsser
Internet Initiative Japan, Inc.
Jinbocho Mitsui Buiding, 1-105
Kanda-Jinbocho, Chiyoda-kun 101-0051
JP

Phone: +81 3 5205 6464
Email: cristel@iij.ad.jp