Internet Engineering Task Force (IETF)
Request for Comments: 6774
Category: Informational
ISSN: 2070-1721

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

November 2012

Distribution of Diverse BGP Paths

Abstract

The BGP4 protocol specifies the selection and propagation of a single best path for each prefix. As defined and widely deployed today, BGP has no mechanisms to distribute alternate paths that are not considered best path between its speakers. This behavior results in a number of disadvantages for new applications and services.

The main objective of this document is to observe that by simply adding a new session between a route reflector and its client, the Nth best path can be distributed. This 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 a software upgrade of provider edge (PE) routers acting as route reflector clients.

Status of This Memo

This document is not an Internet Standards Track specification; it is published for informational purposes.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Not all documents approved by the IESG are a candidate for any level of Internet Standard; see Section 2 of RFC 5741.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc6774.

Table of Contents

1. Introduction

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

This document presents a mechanism for solving the problem based on the conceptual creation of parallel route-reflector planes. It also compares existing solutions and proposes 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 (full BGP mesh or add-paths [ADD-PATHS]) 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.

2. History

The need to disseminate more paths than just the best path is primarily driven by three issues. The first is the problem of BGP oscillations [RFC3345]. The second is the desire for faster reachability restoration in the event of failure of the network link or network element. The third is a need to enhance BGP load-balancing capabilities. These issues have led to the proposal of BGP add-paths [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 the continuously growing number of services carried over BGP, the add-paths proposal was submitted in 2002 to enable BGP to distribute more than one path. This is achieved by including an additional four-octet value called the "Path Identifier" as a part of the Network Layer Reachability Information (NLRI).

The implication of this change on a BGP implementation is that it must now maintain a per-path, instead of per-prefix, peer advertisement state to track to which of the peers a given path was advertised. This new requirement comes with its own memory and processing cost.

An important observation is that distribution of more than one best path by the Autonomous System Border Routers (ASBRs) with multiple External BGP (EBGP) peers attached where no "next-hop self" is set may result in inconsistent best-path selection within the autonomous system. Therefore, it is also required to attach the possible tiebreakers in the form of a new attribute and propagate those within

the domain. The example of such an attribute for the purpose of fast connectivity restoration to address that very case of ASBR injecting multiple external paths into the Internal BGP (IBGP) mesh has been presented and discussed in "Advertisement of Multiple Paths in BGP" [ADD-PATHS]. Based on the additionally propagated information, 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 is contained in Section 6, "Deployment Considerations". In the proposed solution in this document, we observe that to address most of the applications, just use of the best external advertisement is required. For ASBRs that are peering to multiple upstream domains, setting "next-hop self" is recommended.

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

The required code modifications can offer the foundation for enhancements, such as the "Fast Connectivity Restoration Using BGP Add-path" [FAST-CONN]. The deployment of such technology in an entire service-provider network requires software, and perhaps sometimes, in the case of End-of-Engineering or End-of-Life equipment, even hardware upgrades. Such an 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 it may easily take years to achieve 100% deployment coverage within any medium or large global network.

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 solution that is much easier to deploy and 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. That, in light of the number of general network scaling concerns documented in RFC 4984 [RFC4984], "Report from the IAB Workshop on Routing and Addressing", may provide a significant advantage.

3. Goals

The proposal described in this document is not intended to compete with add-paths. It provides an interim solution until add-paths are standardized and implemented and until support for that function can be deployed across the network.

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. The Nth best path describes a set of N paths with different BGP next hops with no implication of ordering or preference among said N paths.

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 so that 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.

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

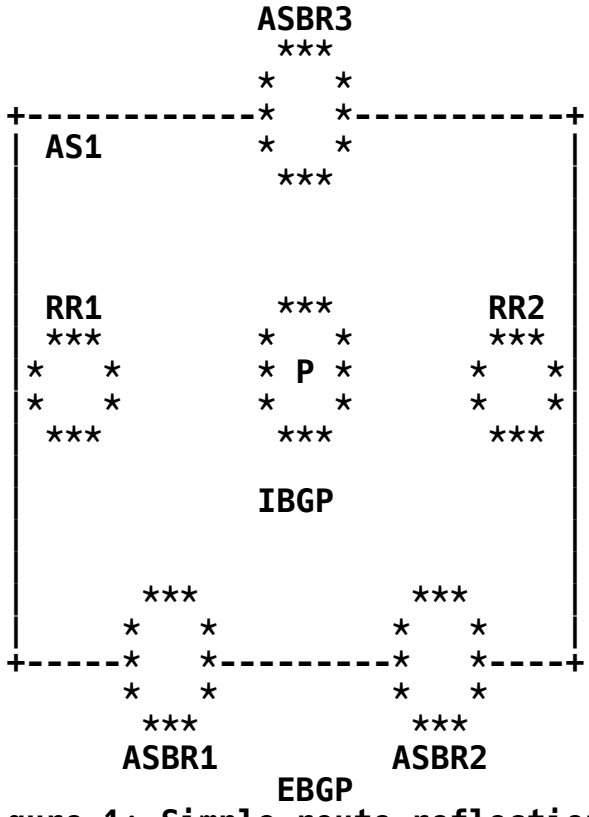

Figure 1: Simple route reflection

Abbreviations used:
- RR - Route Reflector
- P - Core router

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 may 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.

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

1. When the best-path tiebreaker 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's 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 the best-path tiebreaker is MULTI_EXIT_DISC (MED) or LOCAL_PREF: In this case, only one path from the 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 the BGP Best-External proposal [EXT-PATH], which will allow a stated ASBR to still propagate to the RRs on 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.

There is no requirement of path ordering. The "Nth best path" really describes set of N paths with different BGP next hops.

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 than 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 models can be realized to continue advertisement of the overall best path from both route reflectors, while in addition a new session can be provisioned to get an additional path. This will allow the uninterrupted use of the best path, even if one of the RRs goes down, provided that the overall best path is still a valid one.

Each plane of the 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 is easier than upgrading hundreds 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 the 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 Routing Information Base (RIB) and Forwarding Information Base (FIB) ahead of the actual failure.

To address the problem of external paths not being available to route reflectors due to LOCAL_PREF or MED factors, it is recommended that ASBRs enable [EXT-PATH] 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 second-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 functionality is enabled on all ASBRs.

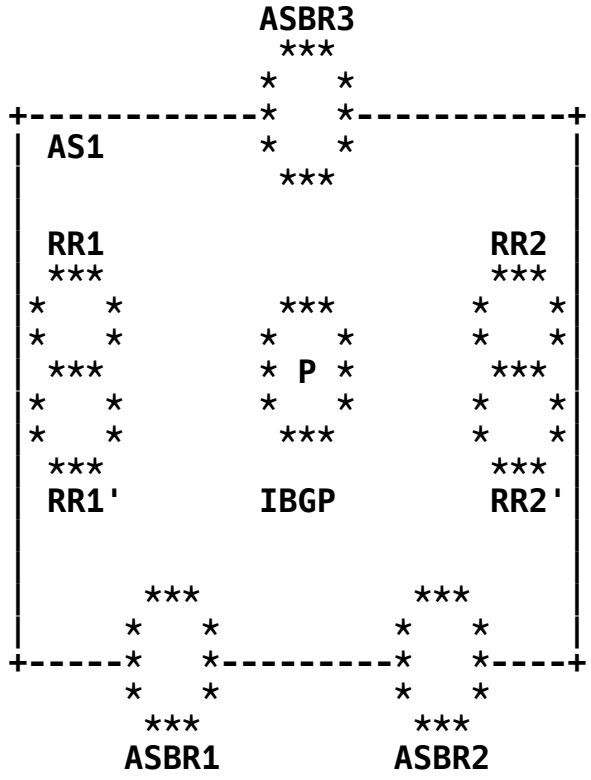

Figure 2: Co-located Second-Best-Path RR Plane

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

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

2. Enabling best-external functionality on ASBRs.

3. Enabling RR1' and RR2' for second plane route reflection. Alternatively, instructing existing RR1 and RR2 to calculate the second-best path also.

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

The expected behavior 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 Section 8 ("Applications").

As an alternative to fully meshing all RRs and RRs', an operator that 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 second plane as well as between planes.

One deployment model of this scenario can be achieved by simply upgrading the existing route reflectors without deploying any new logical or physical platforms. Such an upgrade would allow route reflectors to service both peers that have upgraded to add-paths, as well as those peers that cannot be immediately upgraded while at the same time allowing distribution of more than a single best path. The obvious protocol benefit of using existing RRs to distribute towards their clients' best and diverse BGP paths over different IBGP sessions is the automatic assurance that such a 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 Nth BGP path. Such a 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, the 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 in which an operator wishes to enable a second RR' plane using only a single additional router in a different network location from his current route reflectors. This model would be of particular use in networks in which 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.

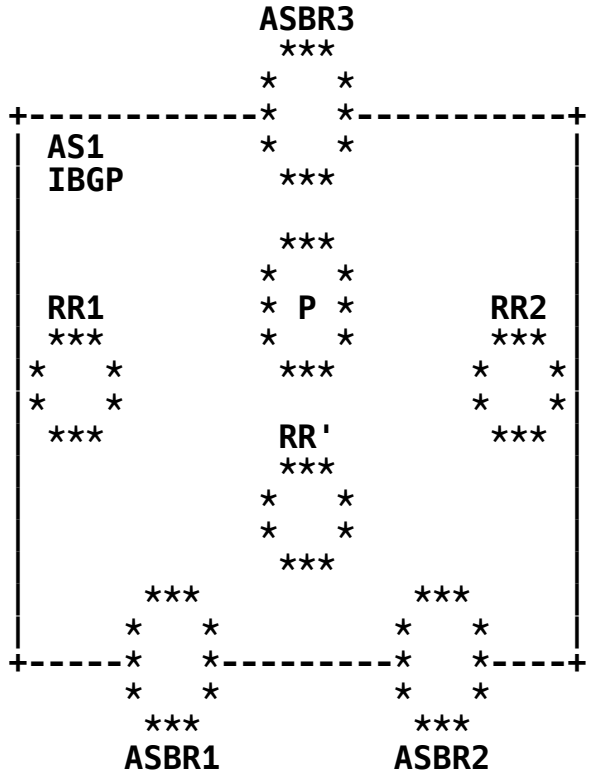

Figure 3: Experimental Deployment of Second-Best-Path RR Plane

The following is a list of configuration changes required to enable the second-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 a new route reflector anywhere in the network.

2. Enabling best-external functionality 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 second plane path calculation.

5. If required, fully meshing newly added RRs' with all the other reflectors in both planes. This condition does not apply if the newly added RR'(s) already have peering to all ASBRs/PEs.

6. Configure new BGP sessions between ASBRs and RRs (unless one of the existing RRs is set to advertise only diverse path to its current clients).

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 existing routers that are not already members of the best-path route-reflector plane can be easily configured to serve the second plane either by 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 best-path calculation, an implementation still has to consider paths with unreachable next hops invalid. It is worth pointing out that some implementations today already allow for configuration that 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 plane does. If we are preparing for a single network failure event, a failure of a non-backed-up Nth best-path route reflector would not result in a 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 the same parts of the network. If the operator chooses to create the Nth 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 second best path.

Similarly to Section 4.1, as an alternative to fully meshing all RRs and diverse path RRs', operators may choose to peer newly introduced RRs' to a hierarchical RR', which would be an IBGP interconnect point between planes.

It is recommended that an implementation advertise the overall best path over the Nth diverse-path session if there is no other BGP path with a different next hop present. This is equivalent to today's case where the client is connected to more than one RR.

4.3. Multi-Plane Route Servers for Internet Exchanges

Another group of devices in which the proposed multi-plane architecture may be of particular applicability is the EBGP route servers used at many Internet exchange points.

In such cases, hundreds of ISPs are interconnected on a common LAN. Instead of having hundreds 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 hundreds of different service providers in order to implement add-path may be much more difficult as compared to asking them to provision one new EBGP session to an Nth best path route server plane. This allows the distribution of more than the single best BGP path from a given route server to such an Internet exchange point (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 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, the following (as well as others) may still apply:

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 hundreds. Such numbers could be a concern on hardware and software used in the late 70s, 80s, and 90s. Today, customer requirements for the number of BGP sessions per box are reaching thousands. This is already an order of magnitude more than the potential number of IBGP sessions. Advancements in the

hardware and software used in production routers means 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 topmost requirements is to keep the design as simple as possible. When the autonomous system's 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 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 mechanism [AUTO-MESH].

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_PREF or MED techniques, and the presence of best-external [EXT-PATH] 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 (diverse) 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 described here applies.

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. This is the main reason many current networks are exposed to a phenomenon called BGP path starvation, which essentially results in the inability to deliver a number of applications discussed later.

When interconnecting BGP speakers between domains, the route reflection equivalent is popularly called the "Route Server" and is globally deployed today in many Internet exchange points.

6. Deployment Considerations

Distribution of the diverse-BGP-paths proposal allows the dissemination of more paths than just the best path to the route-reflector or route-server clients of today's BGP4 implementations. As a deployment recommendation, it needs to be mentioned that fast connectivity restoration as well as a majority of intra-domain BGP-level load balancing needs can be accommodated with only two paths (overall best and second best). Therefore, as a deployment recommendation, this document suggests use of N=2 with diverse-path.

From the client's point of view, receiving additional paths via separate IBGP sessions terminated at the new route-reflector plane is functionally equivalent to constructing a full-mesh peering without the problems such a full mesh would come with, 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 act 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. Regarding RRs being in the data plane unless multipath is enabled, the second best path is expected to be a backup path and should be installed as such into the local RIB/FIB.

The use of the term "planes" in this document is more of a conceptual nature. In practice, all paths are still kept in the single table where normal best path is calculated. This means that tools like the looking glass should not observe any changes or impact when diverse-path has been enabled.

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 (diverse) paths. These are

1. ASBRs advertising their single best-external paths with no LOCAL_PREF or MED present.

2. ASBRs advertising their single best-external paths with LOCAL_PREF or MED present and with BGP best-external functionality enabled.

3. ASBRs with multiple external paths.

This section focuses on discussion of case 3 above 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 local repair of the atomic failure of any external peer without any compromise to the data plane. Traditionally, the most common reason for not setting next-hop self is the associated drawback of losing the ability to signal the external failures of peering ASBRs or links to those ASBRs by fast IGP flooding. Such a potential drawback can be easily avoided by using a different peering address from the address used for next-hop mapping and removing the next-hop from the 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 that the 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 attributes cannot 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, the BGP control plane will on its own re-advertise the newly elected best external path, and route-reflector planes will calculate their Nth best paths and propagate them to its clients. The result is that after seconds, even if potential suboptimality were encountered, it will be quickly and naturally healed.

7. Summary of Benefits

Distribution of the diverse-BGP-paths proposal provides the following benefits when compared to the alternatives:

1. No modifications to the BGP4 protocol.

2. No requirement for upgrades to edge and core routers (as required in [ADD-PATHS]). It is backward compatible with the existing BGP deployments.

3. Can be easily enabled by the introduction of a new route reflector, a 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 would 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 an RR cluster basis.

6. The proposed solution is equally applicable to any BGP address family as described in "Multiprotocol Extensions for BGP-4" [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 that require the presence of redundant BGP paths:

1. Fast connectivity restoration in which backup paths with alternate exit points would be pre-installed as well as pre-resolved in the FIB of routers. This allows for a local action upon reception of a critical event notification of network/node failure. This failure recovery mechanism that is based on the presence of backup paths is also suitable for gracefully addressing scheduled maintenance requirements as described in [BGP-SHUTDOWN].

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

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

An important point to observe is that all of the above intra-domain applications are 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 be equal to a 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 the base BGP4 specification [RFC4271] and especially when compared against full-IBGP-mesh topology.

In addition, the authors observe that all BGP security issues as described in [RFC4272] apply to the additional BGP session or sessions as recommended by this specification. Therefore, all recommended mitigation techniques to BGP security are applicable here.

10. 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
Juniper Networks
1194 N. Mathilda Ave
Sunnyvale, CA 94089
US
Email: smynam@juniper.net

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

11. Acknowledgments

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

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

12. References

12.1. Normative References

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

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

12.2. Informative References

[ADD-PATHS] Walton, D., Chen, E., Retana, A., and J. Scudder, "Advertisement of Multiple Paths in BGP", Work in Progress, June 2012.

[AUTO-MESH] Raszuk, R., "IBGP Auto Mesh", Work in Progress, January 2004.

[BGP-SHUTDOWN] Decraene, B., Francois, P., Pelsser, C., Ahmad, Z., and A. Armengol, "Requirements for the Graceful Shutdown of BGP Sessions", Work in Progress, September 2009.

[EXT-PATH] Marques, P., Fernando, R., Chen, E., Mohapatra, P., and H. Gredler, "Advertisement of the Best External Route in BGP", Work in Progress, January 2012.

[FAST-CONN] Mohapatra, P., Fernando, R., Filsfils, C., and R. Raszuk, "Fast Connectivity Restoration Using BGP Add-path", Work in Progress), October 2011.

[RFC3345] McPherson, D., Gill, V., Walton, D., and A. Retana, "Border Gateway Protocol (BGP) Persistent Route Oscillation Condition", RFC 3345, August 2002.

[RFC4272] Murphy, S., "BGP Security Vulnerabilities Analysis", RFC 4272, January 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
United States

EMail: robert@raszuk.net

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

EMail: rex@cisco.com

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

EMail: keyupate@cisco.com

Danny McPherson
Verisign, Inc.
12061 Bluemont Way
Reston, VA 20190
United States

EMail: dmcpherson@verisign.com

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

EMail: ke-kumaki@kddi.com