Internet Engineering Task Force
Internet-Draft
Updates: 4271 (if approved)
Intended status: Standards Track
Expires: 1 January 2026

E. Chen
Palo Alto Networks
J. Yuan
Individual Contributor

30 June 2025

Deterministic Route Redistribution into BGP
draft-chen-bgp-redist-07.txt

Abstract

In this document we present several examples of non-deterministic routing behavior involving route redistribution into BGP. To eliminate such non-deterministic behavior, we propose an enhancement to BGP route selection that would take into account the administrative distance under certain conditions. Additionally, We recommend lowering the LOCAL_PREF value in implementation for the redistributed backup route when appropriate.

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 BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

Status of This Memo

Table of Contents

1. Introduction

A routing protocol usually downloads its best (or active) route to the routing table, also known as Routing Information Base (RIB), which in turn selects the best (or active) route to program the forwarding table.

When comparing routes from different routing protocols, RIB typically uses the "administrative distance" [ADMIN-DIS] [CISCO-AD] [JUNIPER-AD] (abbreviated as "admin-distance" hereafter) as the tie breaker. The convention is that a route with a lower admin-distance is more preferred, and that is assumed in this document when specific admin-distance values are given as examples. The admin-distance associated with a route in RIB is commonly used to implement various routing schemes such as designating primary and backup routes in a network.

On the other hand, the route selection in BGP [RFC4271] involves comparing the LOCAL_PREF, AS_PATH and other BGP attributes. The bestpath in BGP usually becomes the candidate for downloading to the RIB, and for advertising to BGP neighbors.

It is common to redistribute routes from other routing protocols (such as "static routing" [STATIC-R]) into BGP for route propagation. This topic is briefly discussed in Sect. 9.4 [RFC4271]. A redistributed route is usually assigned the same LOCAL_PREF value as the one for IBGP routes, and has an empty AS_PATH attribute.

The interaction between RIB and BGP follows these general rules:

* A local route may be redistributed into BGP only if it is active in RIB based on the admin-distance.

* Only the bestpath in BGP is downloaded to RIB.

Currently the admin-distance does not play any role in BGP route selection as specified in [RFC4271]. Due to the lack of such correlation between RIB and BGP, when a backup route (based on the admin-distance) is redistributed into BGP as shown in the next section, routing may converge to different paths depending on the order of path arrival. Such non- deterministic routing behavior is clearly detrimental to network design and operations.

To eliminate the non-deterministic routing behavior involving route redistribution into BGP, we propose an enhancement to BGP route selection that would take into account the admin-distance under certain conditions. Additionally, We recommend lowering the LOCAL_PREF value in implementation for the redistributed backup route when appropriate.

The proposed enhancement and recommendation are backward compatible, and can be deployed on a per-router basis.

While static routing is used as examples in the document, the proposed enhancement and recommendation also apply when a route is redistributed from other routing protocols into BGP.

2. The Problem

In this section several examples are presented to illustrate the non-deterministic routing behavior involving route redistribution into BGP.

2.1. On a Single Router

Consider an example in which there are two paths for the same destination on a single router. As shown in the following table, the primary path A is received from an external BGP neighbor, and the backup path B is a static route and is configured for redistribution into BGP.

Path	Type	Admin_Distance	LOCAL_PREF	AS_PATH
A	EBGP	20	100	65535
B	Static	150	100	--

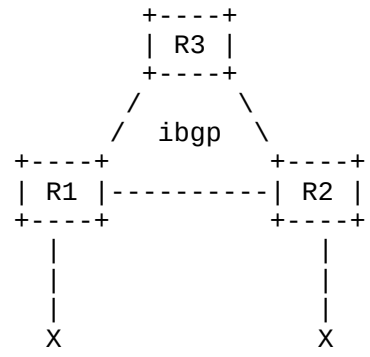
Depending on the order of path arrival, the path that arrives first would be selected as the bestpath in both RIB and BGP.

More specifically, if Path A is received in BGP and is downloaded to RIB first, it would remain as the best in RIB (due to the admin-distance) even after Path B is installed in RIB later. In this case, Path A would be selected as the best in both RIB and BGP.

If Path B appears in RIB and is redistributed into BGP first, it would remain as the best in BGP (due to its local origin or a shorter AS-PATH) even after Path A is received in BGP later. In this case, Path B would be selected as the best in both RIB and BGP.

2.2. In a Network

Consider the following example in which Routers R1, R2 and R3 are part of a provider's network and IBGP sessions are maintained among them. There are two customer connections, a primary connection on R1 and a backup connection on R2. The customer route X is statically routed on both R1 and R2, and is redistributed into BGP. On R2, the backup path for X is configured with a less preferred admin-distance than the one for IBGP paths.



While R1 consistently selects the local static route as the best one, the route selection on R2 would be non-deterministic. As shown in the following figure, there are potentially two BGP paths A and B for X on R2, with Path A learned from R1 and Path B locally redistributed.

Path	Type	Admin_Distance	LOCAL_PREF	AS_PATH
A	IBGP	200	100	--
B	Static	230	100	--

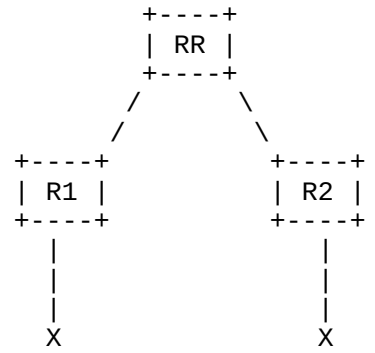
Depending on the order of arrival of these two paths, the path that arrives first would be selected as the bestpath in both RIB and BGP.

More specifically, if Path A is received in BGP and is downloaded to RIB first, it would remain as the best in RIB (due to the admin-distance) even after Path B is installed in RIB later. In this case, Path A would be selected as the best in both RIB and BGP.

If Path B appears in RIB and is redistributed into BGP first, it would remain as the best in BGP (due to its local origin or a lower IGP metric) even when Path A is received in BGP later. In this case, Path B would be selected as the best in both RIB and BGP.

The non-deterministic route selection on R2 could lead other routers, such as R3, to converge to different paths, resulting in unpredictable routing behavior in the network and inconsistency from the intended routing design.

A network using BGP route reflection [RFC4456] without multiple-paths [RFC7911], or BGP confederation [RFC5065], may experience additional cases of network-wide "non-deterministic" routing behavior. For example in the following figure, when both R1 and R2 advertise their respective local routes to the route reflector (RR) simultaneously, the RR would use the "IGP metric" to choose the bestpath between the two IBGP paths. As a result the network may or may not converge to the primary path.



3. The Proposed Solution

To eliminate the non-deterministic routing behavior involving route redistribution into BGP, we propose an enhancement to BGP route selection that would take into account the admin-distance under certain conditions. Additionally, We recommend lowering the LOCAL_PREF value in implementation for the redistributed backup route in the absence of configuration adjustment on any BGP attributes that could influence route selection in a network.

3.1. Enhancement to BGP Route Selection

To make it deterministic on a single router regarding the route being sourced and advertised to the network, we propose that the following procedure be added prior to the step that compares the degrees of preference of routes and identifies the route that has the highest degree of preference, as described in Sect. 9.1.1 [RFC4271] for BGP route selection:

> When comparing a locally redistributed route with another route that is either locally aggregated or is received from a BGP neighbor, favor the one with a more preferred admin-distance. The admin-distance for a BGP route is obtained as follows:
>
> > For a locally redistributed route, it is inherited from the route being redistributed from RIB.
> >
> > For a non-redistributed route, it is of the same value as the admin-distance assigned to the route for the purpose of RIB installation (regardless of whether it is actually installed in RIB).

As the admin-distance is not propagated by BGP, comparing the admin-distance may not be sufficient when IBGP paths are involved. To influence route selection in a network, the LOCAL_PREF or other relevant BGP attributes should also be adjusted as described in the next section.

3.2. Setting the LOCAL_PREF Value

To designate a non-BGP route as a backup route in the network, it should be assigned a less preferred admin-distance than the value for IBGP routes. When such a route is redistributed into BGP, it should be treated as a backup route in the whole network by using one or more of the BGP attributes to influence route selection, namely, LOCAL_PREF, AS_PATH, MULTI_EXIT_DISC, and ORIGIN [RFC4271]. Configuration can be used to achieve the intended outcome.

In the absence of configuration that adjusts any of the aforementioned BGP attributes for the redistributed backup route, we RECOMMEND that a lower LOCAL_PREF value (e.g., half of the default LOCAL_PREF value for IBGP routes) be assigned in implementation.

The adjustment of the LOCAL_PREF for the redistributed backup route would ensure the intended primary/backup roles are maintained for the routes involved. Furthermore, the adjustment would eliminate the non-deterministic behavior in the common deployment with a pair of primary/backup connections in a network, as described earlier in this document.

However, in scenarios where route redistribution is a part of a more complex deployment involving more than just a pair of primary/backup connections in a network, additional configuration adjustments will be necessary and can be built following the general guidelines discussed in this document.

4. IANA Considerations

This document has no request for IANA.

5. Security Considerations

The solution proposed in this document does not change the underlying security or confidentiality issues inherent in the existing BGP [RFC4271].

6. Acknowledgments

We would like to thank Naiming Shen, Acee Lindem, Robert Raszuk, Jeffrey Haas and Jakob Heitz for inputs and discussions.

We would also like to thank Jeffrey Haas for making us aware of the procedure that compares the "Route Preference" in the BGP Path Selection by Juniper Networks [JUNIPER-BGP]. In essence the enhancement to the BGP route selection proposed in this document is similar to that procedure.

7. References

7.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <https://www.rfc-editor.org/info/rfc4271>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

7.2. Informative References

[ADMIN-DIS] Wikipedia, "Administrative Distance", <https://en.wikipedia.org/wiki/Administrative_distance>.

[CISCO-AD] Cisco Systems, "Describe Administrative Distance", <https://www.cisco.com/c/en/us/support/docs/ip/border-gateway-protocol-bgp/15986-admin-distance.html>.

[JUNIPER-AD] Juniper Networks, "Understanding Route Preference Values (Administrative Distance)", <https://www.juniper.net/documentation/us/en/software/junos/routing-overview/bgp/topics/concept/routing-protocols-default-route-preference-values.html>.

[JUNIPER-BGP] Juniper Networks, "Understanding BGP Path Selection", <https://www.juniper.net/documentation/us/en/software/junos/bgp/topics/topic-map/bgp-overview.html#id-10119586>.

[RFC4456] Bates, T., Chen, E., and R. Chandra, "BGP Route Reflection: An Alternative to Full Mesh Internal BGP (IBGP)", RFC 4456, DOI 10.17487/RFC4456, April 2006, <https://www.rfc-editor.org/info/rfc4456>.

[RFC5065] Traina, P., McPherson, D., and J. Scudder, "Autonomous System Confederations for BGP", RFC 5065, DOI 10.17487/RFC5065, August 2007, <https://www.rfc-editor.org/info/rfc5065>.

[RFC7911] Walton, D., Retana, A., Chen, E., and J. Scudder, "Advertisement of Multiple Paths in BGP", RFC 7911, DOI 10.17487/RFC7911, July 2016, <https://www.rfc-editor.org/info/rfc7911>.

[STATIC-R] Wikipedia, "Static routing", <https://en.wikipedia.org/wiki/Static_routing>.

Authors' Addresses

Enke Chen
Palo Alto Networks
United States of America
Email: enchen@paloaltonetworks.com

Jenny Yuan
Individual Contributor
United States of America
Email: jennyyuan1000@yahoo.com