

Deterministic Route Redistribution into BGP

[draft-chen-bgp-redist-00.txt](#)

1. Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of [Section 10 of RFC2026](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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.''

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/1id-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

2. Abstract

In this document we propose an enhancement to the BGP route selection algorithm that would make the interaction of redistributed routes and other BGP routes deterministic, thus facilitating the deployment of various routing requirements. The proposed enhancement is backward compatible.

3. Introduction

A routing protocol usually downloads its best (or active) routes to the RIB, which in turn selects the best routes to program the forwarding engine. When comparing routes from different protocols, RIB typically uses an "admin distance" (or "protocol preference") as the tie breaker. The "admin distance" is a value assigned to each route being downloaded to RIB, and it is a local matter how the values are assigned and compared.

On the other hand, the BGP route selection algorithm (as specified in [1]) involves comparing the LOCAL_PREF, AS_PATH, MED, IGP Metric and other parameters of all the paths involved. The BGP best path usually becomes the candidate for downloading to the RIB, and for advertising to BGP peers.

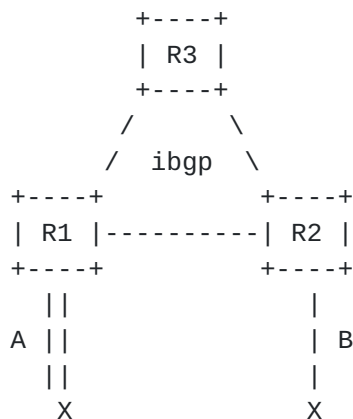
It is common to redistribute routes from other routing protocols (such as RIP and static routing) into BGP in order to implement various routing policies. A redistributed route typically has empty AS_PATH attribute, and zero IGP metric.

As detailed in the following section, the interaction of redistributed routes and other BGP routes are sometimes order dependent, and the BGP best path selected can thus be non-deterministic. Consequently, complicated configurations are sometimes required in order to deploy simple routing requirements (such as primary and backup).

In this document we propose an enhancement to the BGP route selection algorithm that would make the interaction of redistributed routes and other BGP routes deterministic, thus facilitating the deployment of various routing requirements. The proposed enhancement is backward compatible.

4. The Problem

One common routing setup for a multi-homed customer is to treat one connection as the primary, and another as the backup. Consider the following example in which there are two customer connections, the primary path A and the backup path B. Routers R1, R2 and R3 are part of a provider network and IBGP sessions are maintained among them. The customer route X is statically routed on R1 and R2, and is redistributed into BGP. On R2, the backup path for X is configured with a less preferred "admin distance" than any other path for X.



Let us take a look at the BGP table and RIB on R2. There are potentially two BGP paths for X, one locally redistributed due to path B, and one learned from R1. Depending on the order of arrival of these two paths, the routing behavior may differ:

- (1) When the IBGP path from R1 gets into the BGP table first, it would be selected as the best path, and then downloaded to the RIB. Due to the more preferred value of the "admin distance", the IBGP path would be selected as the best path in RIB, and the local path B would serve as a backup and would not be redistributed into BGP (assuming that only the active path in RIB is redistributed into BGP). As a result, R1, R2, R3 and the rest of the network would converge to the primary path on R1.
- (2) When the IBGP path from R1 gets into the BGP table later than the locally redistributed route does, then the two paths in the BGP table need to be compared for route selection. By default the LOCAL_PREF, AS-PATH, and MED are the same. Due to its IGP metric, the locally redistributed route would be selected as the best path, and thus propagated to other IBGP peers. In this scenario the intended backup path B is selected as the primary path on R2, and some portions of the network (like R3) may converge to use the path from R2 as well.

To eliminate the non-deterministic routing behavior in this case, one approach is to configure lower LOCAL_PREF (and modify any other vendor specific route selection criteria preceding the LOCAL_PREF comparison) for the redistributed route. However, like all route-specific BGP configurations, this approach tend to increase the operational complexity and cost.

An alternative solution is proposed in the next section that does not require any route-specific BGP configurations.

5. The Proposed Solution

We propose that the following step be added as the first step in the BGP route selection:

When comparing a locally redistributed route with one that is received from a BGP peer, 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, the "admin distance" is inherited from the route being redistributed from RIB.

For a route received from a BGP peer, the "admin distance" is the same as the "admin distance" assigned to the route for the purpose of RIB downloading (regardless of whether it is actually downloaded).

The proposed revision is backward compatible.

It is noted that the proposed algorithm is generally applicable when a route is redistributed from one protocol into another protocol.

6. Security Considerations

This extension does not introduce any security issues.

7. Intellectual Property Considerations

Redback Networks may have intellectual property rights claimed in regard to some of the specification contained in this document.

8. Acknowledgments

The authors would like to thank Naiming Shen and Acee Lindem for their inputs and review.

9. References

[1] Y. Rekhter, T. Li, and S. Hares, "A Border Gateway Protocol 4 (BGP-4)", [draft-ietf-idr-bgp4-23.txt](#), November 2003.

10. Author Information

Enke Chen
Redback Networks, Inc.
300 Holger Way
San Jose, CA 95134
e-mail: enke@redback.com

Jenny Yuan
Redback Networks, Inc.
300 Holger Way
San Jose, CA 95134
e-mail: jenny@redback.com

