

Network Working Group
Internet Draft
Expiration Date: April 2003
File name: [draft-retana-bgp-custom-decision-00.txt](#)

Alvaro Retana
Russ White
Cisco Systems, Inc.
October 2002

BGP Custom Decision Process
draft-retana-bgp-custom-decision-00.txt

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. Internet Drafts may be updated, replaced, or obsoleted by other documents at any time. It is not appropriate to use Internet Drafts as reference material or to cite them other than as a "working draft" or "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>.

Abstract

The BGP specification [[RFC1771](#)] defines a Decision Process for installation of routes into the Loc-RIB. This process takes into account an extensive series of path attributes, which can be manipulated to indicate preference for specific paths. It is cumbersome (if at all possible) for the end user to define policies that will select, after partial comparison, a path based on subjective local (domain and/or node) criteria.

This document defines a new Extended Community [[EXT_COMM](#)], called the Cost Community, which may be used in tie breaking during the best path selection process. The end result is a local custom decision process.

1. Introduction

There are a number of metrics available within the BGP decision process [[RFC1771](#)] which can be used to determine the exit point for traffic, but there is no metric, or combination of metrics, which can be used to break a tie among generally equal paths.

- o LOCAL_PREF: The LOCAL_PREF is an absolute tie breaker near the beginning of the decision process. There is no way to configure the LOCAL_PREF such that the MED, IGP metric, and other metrics are considered before breaking a tie.
- o MED: The MULTI_EXIT_DISC is an indicator of which local entrance point an AS would like a peering AS to use; MED isn't suitable to break the tie between two equal cost paths learned from two peer ASes. MED is also compared before the IGP metric; there is no way to set the MED so a path with a higher IGP metric is preferred over a path with a lower IGP metric.
- o IGP Metric: It is possible, using the IGP metric, to influence individual paths with otherwise equal cost metrics, but only by changing the next hop towards each path, and configuring the IGP costs of reaching each next hop. This method is cumbersome, and prone to confusion and error.

2. Specification of Requirements

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](#)].

3. The BGP Cost Community

The BGP Cost Community is an Opaque Extended Community [[EXT_COMM](#)] defined as follows:

Type Field:

The value of the high-order octet of the extended Type Field is 0x43, which indicates it is non-transitive. The value of the low-order octet of the extended type field for this community is TBD.

Value Field

The Value field contains three distinct sub-fields, described

below:

```

+-----+
| Point of Insertion (1 octet) |
+-----+
| Community-ID (1 octet)      |
+-----+
| Cost (4 octets)              |
+-----+

```

The Point of Insertion sub-field contains the value of the path attribute [[BGP_PAR](#)] *after* which this community MUST be considered during the best path selection process.

The BGP decision process [[RFC1771](#)] includes some steps that do not correspond to any path attribute; the following values are defined:

Value	Meaning
-------	---------

128

ABSOLUTE_VALUE - Indicates that the Cost Community MUST be considered as the first step in determining the Degree of Preference of a path.

129

IGP_COST - Indicates that the Cost Community MUST be considered after the interior (IGP) distance to the next-hop has been compared.

130

EXTERNAL_INTERNAL - Indicates that the Cost Community MUST be considered after the paths advertised by BGP speakers in a neighbouring autonomous system (if any) have been selected.

131

BGP_ID - Indicates that the Cost Community MUST be considered after the BGP Identifier (or ORIGINATOR_ID [[RFC2796](#)]) has been compared.

The Community-ID sub-field contains an identifier to distinguish between multiple instances of the Cost Community.

The Cost sub-field contains a value assigned by the network administrator and that is significant to the local autonomous system. The lower cost MUST be preferred. The default value is 0x7FFFFFFF (half the maximum value).

4. Operation

The network administrator may use the Cost Community to assign a value to a path originated or learned by a peer in any part of the local domain. The Point of Insertion may also be specified using the values defined in the IANA registry [[BGP_PAR](#)] or this document.

If a BGP speaker receives a path that contains the Cost Community, it MUST consider its value at the Point of Insertion specified, when calculating the best path [[RFC1771](#)].

If the Point of Insertion is not valid for the local best path selection implementation, then the Cost Community is silently ignored. Paths that do not contain the Cost Community (for a valid, particular Point of Insertion) are considered to have the default value.

Multiple Cost Communities may indicate the same Point of Insertion. In this case, the Cost Community with the lowest Community-ID is considered first. In other words, all the Cost Communities for a specific Point of Insertion are considered, starting with the one with the lowest Community-ID.

If a range of routes is to be aggregated and the resultant aggregates path attributes do not carry the ATOMIC_AGGREGATE attribute, then the resulting aggregate should have an Extended Communities path attribute which contains the set union of all the Cost Communities from all of the aggregated routes. If multiple Cost Communities for the same Point of Insertion (and with the same Community-ID), then only the ones with the highest Cost SHOULD be included.

The Cost Community is a non-transitive Extended Community [[EXT_COMM](#)]. If a Cost Community is received across an Autonomous System boundary, then the receiver SHOULD strip it off the BGP update, and ignore it when running the selection process.

4.1. Deployment Considerations

The mechanisms described in this document may be used to modify the BGP path selection process arbitrarily, within an AS. It is important that a consistent path selection process be maintained across the local Autonomous System to avoid potential routing loops. In other words, if the Cost Community is used, all the nodes in the AS that may have to consider this new community at any Point of Insertion SHOULD be aware of the mechanisms described in this document.

5. IANA Considerations

The section titled "The BGP Cost Community" defines a series of values to be used to indicate steps in the best path selection process [[RFC1771](#)] that do not map directly to a path attribute. IANA is expected to maintain the registry for these values. Values 128 through 191 are to be assigned by IANA using the "IETF Consensus" policy defined in [RFC2434](#) [[RFC2434](#)]. Values 192 through 254 are for "Private Use" as defined in [RFC2434](#) [[RFC2434](#)].

The Value space defined above matches the one already used and maintained by IANA to assign BGP path attributes [[BGP_PAR](#)]. It is RECOMMENDED that the values specified in this document are added to the current registry [[BGP_PAR](#)] and the process to register new attributes be updated [[RFC2042](#)].

6. Security Considerations

This document introduces no new security concerns to BGP or other specifications referenced in this document.

7. Acknowledgments

We would like to thank Chris Whyte, Khamsa Enaya, John Scudder, Tom Barron, Eric Rosen, Barry Friedman, Gargi Nalawade, Ruchi Kapoor and Chandra Appanna for their comments and suggestions. We would like to also thank Dan Tappan for the Opaque Extended Community type.

8. References

[RFC1771]

Rekhter, Y., and T. Li, "A Border Gateway Protocol 4 (BGP-4)", [RFC 1771](#), March 1995.

[EXT_COMM]

Sangli, S., Tappan, D., and Rekhter, Y., "BGP Extended Communities Attribute", Work in Progress ([draft-ietf-idr-bgp-ext-communities-05.txt](#)), May 2002.

[RFC2119]

Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [RFC 2119](#), March 1997.

[BGP_PAR]

Internet Assigned Numbers Authority, "BGP Parameters",

<http://www.iana.org/assignments/bgp-parameters>.

[RFC2796]

Bates, T., Chandra, R., and Chen, E., "BGP Route Reflection - An Alternative to Full Mesh IBGP", [RFC 2796](#), April 2000.

[RFC2434]

Narten, T., Alvestrand, H., "Guidelines for Writing an IANA Considerations Section in RFCs", [RFC 2434](#), October 1998.

[RFC2042]

Manning, B., "Registering New BGP Attribute Types", [RFC 2042](#), January 1997.

[9. Authors' Addresses](#)

Alvaro Retana
Cisco Systems, Inc.
7025 Kit Creek Rd.
Research Triangle Park, NC 27709
Email: aretana@cisco.com

Russ White
Cisco Systems, Inc.
7025 Kit Creek Rd.
Research Triangle Park, NC 27709
Email: riw@cisco.com

