Inter-Domain Routing
Internet-Draft
Intended status: Standards Track
Expires: September 15, 2011

A. Retana
R. White
Cisco Systems, Inc.
March 14, 2011

BGP Custom Decision Process
draft-retana-bgp-custom-decision-01

Abstract

The BGP specification 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, 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.

Table of Contents

1. Introduction

There are a number of metrics available within the BGP decision process [RFC4271] 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.

- 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.

- 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.

- 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.

The BGP specification 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, called the Cost Community, which may be used in tie breaking during the best path selection process. The end result is a custom decision process.

2. 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].

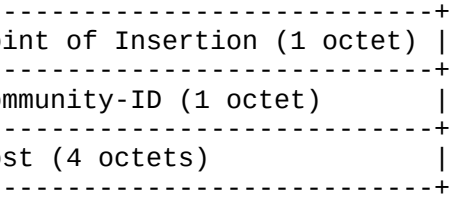
3. The BGP Cost Community

The BGP Cost Community is an Opaque Extended Community [RFC4360]

defined as follows:

Type Field:
The value of the high-order octet of this Opaque Extended Community is 0x03 or 0x43. The value of the low-order octet of the extended type field for this community is 0x01.

Value Field:
The Value field contains three distinct sub-fields, described below:



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

The BGP decision process includes some steps that do not correspond to any path attribute; the following values are defined:

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 neighboring 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 [RFC4456]) 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 from a peer in any part of the local domain. The Point of Insertion may also be specified using the values assigned by IANA (Section 6) or this document.

If a BGP speaker receives a path that contains the Cost Community, it SHOULD consider its value at the Point of Insertion specified, when calculating the best path [RFC4271].

If the Point of Insertion is not valid for the local best path selection implementation, then the Cost Community SHOULD be silently ignored. Paths that do not contain the Cost Community (for a valid, particular Point of Insertion) MUST be 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 MUST be 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.

If the non-transitive version of 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.

5. Deployment Considerations

The mechanisms described in this document may be used to modify the BGP path selection process arbitrarily. 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.

6. IANA Considerations

IANA is asked to assign the type values indicated in Section 3 to the Cost Community in the BGP Opaque Extended Community registry [BGP_EXT].

Section 3 also defines a series of values to be used to indicate steps in the best path selection process that do not map directly to a path attribute. IANA is expected to maintain a registry for the Cost Community Point of Insertion values. Values 1 through 127 are to be assigned using the "Standards Action" policy or the Early Allocation process [RFC4020]. Values 128 through 191 are to be assigned using the "IETF Consensus" policy. Values 192 through 254 are to be assigned using the "First Come First Served" policy. Values 0 and 255 are reserved for future use and SHOULD NOT be used. All the policies mentioned are documented in [RFC5226].

Some of the values in this new registry match the values assigned in the BGP Path Attributes registry [BGP_PAR]. It is RECOMMENDED that an effort be made to assign the same values in both tables when applicable. The table in Appendix A shows the initial allocations for the new Cost Community Point of Insertion registry.

7. Security Considerations

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

8. Acknowledgements

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

9. References

9.1. Normative References

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

[RFC4020] Kompella, K. and A. Zinin, "Early IANA Allocation of Standards Track Code Points", BCP 100, RFC 4020, February 2005.

[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.

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

9.2. Informative References

[BGP_EXT] Internet Assigned Numbers Authority, "BGP Extended Communities", 2010, <http://www.iana.org/assignments/bgp-extended-communities>.

[BGP_PAR] Internet Assigned Numbers Authority, "BGP Parameters", 2010, <http://www.iana.org/assignments/bgp-parameters/>.

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

Appendix A. Cost Community Point of Insertion Registry

The tables below document the initial Cost Community Point of Insertion Registry

Range	Registration Procedure
0	Reserved
1-127	Standards Action
128-191	IETF Consensus
192-254	First Come First Served
255	Reserved

Registration Procedure

Value	Code	Reference
1	ORIGIN	RFC4271
2	AS_PATH	RFC4271
3	Unassigned	
4	MULTI_EXIT_DISC	RFC4271
5	LOCAL_PREF	RFC4271
6-25	Unassigned	
26	AIGP	draft-ietf-idr-aigp
27-127	Unassigned	
128	ABSOLUTE_VALUE	draft-retana-bgp-custom-decision
129	IGP_COST	draft-retana-bgp-custom-decision
130	EXTERNAL_INTERNAL	draft-retana-bgp-custom-decision
131	BGP_ID	draft-retana-bgp-custom-decision

Point of Insertion Codes

Appendix B. Changes from version -00

The changes with respect to version -00 of this draft are as follow:

- Defined a transitive type. (Section 3)
- Updated the IANA Considerations (Section 6) to create a Cost Community Point of Insertion Registry. (Appendix A)
- Miscellaneous Updates: updated format, refreshed references, updated acknowledgements, minor edits.

Authors' Addresses

Alvaro Retana
Cisco Systems, Inc.
7025 Kit Creek Rd.
Research Triangle Park, NC 27709
USA

Phone: +1 919 392 2061
Email: aretana@cisco.com

Russ White
Cisco Systems, Inc.
7025 Kit Creek Rd.
Research Triangle Park, NC 27709
USA

Phone: +1 919 392 3139
Email: russwh@cisco.com