BGP Large Communities Attribute - RFC8092

Barry O’Donovan

Internet Neutral Exchange Association
Company Limited by Guarantee
INEX is the internet peering point for the island of Ireland

- Member owned association, not for profit and founded in 1996
- >90 peering organisations including content and ~98% of eyeballs in the country
- >150Gbps of IP data exchanged at peek
- Dual infrastructure network in 6 points of presence in Dublin, own dark fibre
- Opened INEX Cork in 2016
- Just launched 100Gb ports on Arista switches
- Other switching kit is Extreme Networks x670’s and Brocade (being retired)
- Home of IXP Manager - https://www.ixpmanager.org/
To talk about BGP large communities, we need to know:

- **AS** - autonomous systems
  - Essentially a network such as an Internet Service Provider
- **ASN** - AS number - how such networks are identified
  - Was typically a 16-bit number (e.g. AS2128 is INEX)
  - RFC4893 extended this to 32-bit numbers in 2007
- **BGP** - how networks share routing and reachability information between AS’
- **BGP communities** - extra information that can be attached to a prefix / route
BGP community example:

gw2#show ip bgp 5.10.6.0/23
13237 6833 198726

194.88.240.57 from 194.88.240.57

Origin IGP, localpref 400, valid, external, best

Community: 0:8218 13237:45049 13237:46081
BGP COMMUNITIES

BGP Communities Attribute (RFC1997, August 1996)

- Designed to simplify internet routing policies
- Can be informative or used to signal an action
- 32-bit value displayed as [16-bit ASN : 16-bit value]

Community: 0:8218 13237:45049 13237:46081
Community: 0:8218 13237:45049 13237:46081
Community: 0:8218 13237:45049 13237:46081

$ whois as13237
...
remarks: ** BGP communities for euNetworks transit customers **
remarks: Origin communities on received routes:
remarks: 13237:400cc Tags used for peers
remarks: 13237:440cc Tags used for transit peers
remarks: 13237:450cc Tags used for BGP customer routes
remarks: 13237:470cc Tags used for aggregate routes
remarks: with the following cc = country code
remarks: 31 = NL
remarks: 33 = FR
remarks: 35 = IE
remarks: 42 = CZ
remarks: 43 = AT
remarks: 44 = UK
remarks: 45 = DK
remarks: 49 = DE
...
Community: 0:8218 13237:45049 13237:46081

$ whois as13237
...
remarks: ** BGP communities for euNetworks transit customers **
remarks: Origin communities on received routes:
remarks: 13237:400cc Tags used for peers
remarks: 13237:440cc Tags used for transit peers
remarks: 13237:450cc Tags used for BGP customer routes
remarks: 13237:470cc Tags used for aggregate routes
remarks: with the following cc = country code
remarks: 31 = NL
remarks: 33 = FR
remarks: 35 = IE
remarks: 42 = CZ
remarks: 43 = AT
remarks: 44 = UK
remarks: 45 = DK
remarks: 49 = DE
...
$ whois as13237

remarks:  Prepend communities to modify announcements to peers
remarks:  13237:3801n announcements to AS9033 (ECIX
remarks:  13237:3802n announcements to AS43760 (INEX RS)
...
remarks:  with n = 0,1,2,3 meaning
remarks:  n = 0 do not announce to peer
remarks:  n = 1 prepend "AS13237"
remarks:  n = 2 prepend "AS13237 AS13237"
remarks:  n = 3 prepend "AS13237 AS13237 AS13237"
BGP Large Communities

RFC8092
Problem: You can’t squeeze 64-bits into 32-bits

- 32-bit ASNs became an operational reality with RFC4893 in May 2007
- 32-bit ASNs now widely used as edge and transit ASNs

Source: [http://www.potaroo.net/tools/asn32/](http://www.potaroo.net/tools/asn32/) - As of March 6 2017
Problem: You can’t squeeze 64-bits into 32-bits

```sql
mysql> SELECT COUNT(*) FROM inex.cust WHERE autsys > 65535;
```

```
+----------+
| COUNT(*) |
+----------+
|    13    |
+----------+
```

Or, about 15% of members.
IXPs facilitate multi-lateral BGP peering sessions
- one network can peer with all the other networks at an exchange

Presents a different scaling challenge:
- 5 IXP members  => 4 BGP sessions each  => \(\sim(n-1)^2\) sessions
- 90 IXP members => 89 BGP sessions each => \(\sim7,921\) sessions

Route servers solve the scaling problem.
Route Servers:

IXI full mesh peering relationships

IXI route server peering relationships
IXPs facilitate multi-lateral BGP peering sessions
  • one network can peer with all the other networks at an exchange
  • Presents a different scaling challenge:
    - 5 IXP members => 4 BGP sessions each => ~(n-1)^2 sessions
    - 90 IXP members => 89 BGP sessions each => ~7,921 sessions
  • Route servers solve the scaling problem.
    • But they introduce a network policy control problem.
Well known communities for route server prefix filtering

- Prevent announcement of a prefix to a peer 0:peer-as
- Announce a route to a certain peer 43760:peer-as
- Prevent announcement of a prefix to all peers 0:43760
- Announce a route to all peers 43760:43760 (no-op)

Note that INEX’s route server ASN is: 43760

Problem: members with a 32-bit ASN cannot be filtered with this mechanism and nor can then signal filtering for other 32-bit ASNs.
The Solution: RFC8092

BGP Large Communities Attribute

Abstract

This document describes the BGP Large Communities attribute, an extension to BGP-4. This attribute provides a mechanism to signal opaque information within separate namespaces to aid in routing management. The attribute is suitable for use with all Autonomous System Numbers (ASNs) including four-octet ASNs.
Like RFC 1997 Communities, but Larger
Design Goals

- Simply “larger”. That’s it. No (or little!) room for bike-shedding.
  - Extends RFC1997 communities for 32-bit ASNs
  - Signal an action without losing information about the origin or target
- No *well-known* communities (no-advertize, no-export, blackhole, etc.)
  - RFC1997 well-known communities can still be used
- Easy to implement and adopt
Encoding and Usage

• BGP Large Communities are encoded as 96-bit values
  • [ 32-bit ASN : 32-bit value : 32-bit value ]
• Canonical representation:
  • $me:$action:$you
• Easy to implement and adopt
### Implementation on INEX Route Servers - RFC1997:

<table>
<thead>
<tr>
<th>Description</th>
<th>Community</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevent announcement of a prefix to a peer</td>
<td>0:peer-as</td>
</tr>
<tr>
<td>Announce a route to a certain peer</td>
<td>43760:peer-as</td>
</tr>
<tr>
<td>Prevent announcement of a prefix to all peers</td>
<td>0:43760</td>
</tr>
<tr>
<td>Announce a route to all peers</td>
<td>43760:43760</td>
</tr>
</tbody>
</table>

**RFC8092**
Implementation on INEX Route Servers - RFC8092:

<table>
<thead>
<tr>
<th>Description</th>
<th>Community</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevent announcement of a prefix to a peer</td>
<td>43760:0:peer-as</td>
</tr>
<tr>
<td>Announce a route to a certain peer</td>
<td>43760:1:peer-as</td>
</tr>
<tr>
<td>Prevent announcement of a prefix to all peers</td>
<td>43760:0:0</td>
</tr>
<tr>
<td>Announce a route to all peers</td>
<td>43760:1:0</td>
</tr>
</tbody>
</table>
### Major Milestones Towards an RFC Standard

<table>
<thead>
<tr>
<th>Date</th>
<th>Milestone</th>
</tr>
</thead>
<tbody>
<tr>
<td>September 2, 2016</td>
<td>Published draft-heitz-idr-large-community-03</td>
</tr>
<tr>
<td>September 6, 2016</td>
<td>Requested IDR WG Adoption</td>
</tr>
<tr>
<td>September 24, 2016</td>
<td>IDR Working Group Adoption of draft-ietf-idr-large-community-00</td>
</tr>
<tr>
<td>September 29, 2016</td>
<td>Early IANA BGP Path Attributes Code (30) Allocation</td>
</tr>
<tr>
<td>October 11, 2016</td>
<td>BGP Large Communities Beacon Prefixes Announced</td>
</tr>
<tr>
<td>October 17, 2016</td>
<td>Start of IDR Working Group Last Call</td>
</tr>
<tr>
<td>October 26, 2016</td>
<td>Early IANA BGP Path Attributes Code (32) Allocation</td>
</tr>
<tr>
<td>November 2, 2016</td>
<td>Start of IETF Last Call and IESG Review</td>
</tr>
<tr>
<td>December 1, 2016</td>
<td>Start of IESG Last Call</td>
</tr>
<tr>
<td>December 18, 2016</td>
<td>IESG Ballot Issued</td>
</tr>
<tr>
<td>January 5, 2017</td>
<td>IESG Approved Revision -12 for RFC Publication</td>
</tr>
<tr>
<td>February 16, 2017</td>
<td>RFC 8092 “BGP Large Communities Attribute” Published</td>
</tr>
</tbody>
</table>
## BGP Speaker Implementation Status

<table>
<thead>
<tr>
<th>Implementation</th>
<th>Software</th>
<th>Status</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arista</td>
<td>EOS</td>
<td>Planned</td>
<td>Feature Requested BUG169446</td>
</tr>
<tr>
<td>Cisco</td>
<td>IOS XR</td>
<td>✔️ Done!</td>
<td>beta (perhaps in 6.3.2 for real?)</td>
</tr>
<tr>
<td>cz.nic</td>
<td>BIRD</td>
<td>✔️ Done!</td>
<td>BIRD 1.6.3 (commit)</td>
</tr>
<tr>
<td>ExaBGP</td>
<td>ExaBGP</td>
<td>✔️ Done!</td>
<td>PR482</td>
</tr>
<tr>
<td>FreeRangeRouting</td>
<td>frr</td>
<td>✔️ Done!</td>
<td>Issue 46 (Commit)</td>
</tr>
<tr>
<td>nop.nu</td>
<td>freeRouter</td>
<td>✔️ Done!</td>
<td></td>
</tr>
<tr>
<td>Juniper</td>
<td>Junos OS</td>
<td>Planned</td>
<td>Second Half 2017 (perhaps 17.3R1?)</td>
</tr>
<tr>
<td>MikroTik</td>
<td>RouterOS</td>
<td>Won't Implement Until RFC Feature Requested 2016090522001073</td>
<td></td>
</tr>
<tr>
<td>Nokia</td>
<td>SR OS</td>
<td>Planned</td>
<td>Third Quarter 2017</td>
</tr>
<tr>
<td>OpenBSD</td>
<td>OpenBGPD</td>
<td>✔️ Done!</td>
<td>OpenBSD 6.1 (commit)</td>
</tr>
<tr>
<td>OSGR</td>
<td>GoBGP</td>
<td>✔️ Done!</td>
<td>PR1094</td>
</tr>
<tr>
<td>rtbrick</td>
<td>Fullstack</td>
<td>✔️ Done!</td>
<td>FullStack 17.1</td>
</tr>
<tr>
<td>Quagga</td>
<td>Quagga</td>
<td>✔️ Done!</td>
<td>Quagga 1.2.0 875</td>
</tr>
<tr>
<td>Ubiquiti Networks</td>
<td>EdgeOS</td>
<td>Planned</td>
<td>Internal Enhancement Requested</td>
</tr>
<tr>
<td>VyOS</td>
<td>VyOS</td>
<td>Requested</td>
<td>Feature Requested T143</td>
</tr>
</tbody>
</table>

Visit [http://largebgpcommunities.net/implementations/](http://largebgpcommunities.net/implementations/) for the Latest Status
## Tools & Ecosystem Implementation Status

<table>
<thead>
<tr>
<th>Implementation</th>
<th>Software</th>
<th>Status</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>DE-CIX</td>
<td>pbgpp</td>
<td>✅ Done!</td>
<td>PR16</td>
</tr>
<tr>
<td>FreeBSD</td>
<td>tcpdump</td>
<td>✅ Done!</td>
<td>PR213423</td>
</tr>
<tr>
<td>Marco d’Itri</td>
<td>zebra-dump-parser</td>
<td>✅ Done!</td>
<td>PR3</td>
</tr>
<tr>
<td>OpenBSD</td>
<td>tcpdump</td>
<td>✅ Done!</td>
<td>OpenBSD 6.1 (patch)</td>
</tr>
<tr>
<td>pmacct.net</td>
<td>pmacct</td>
<td>✅ Done!</td>
<td>PR61</td>
</tr>
<tr>
<td>RIPE NCC</td>
<td>bgpdump</td>
<td>✅ Done!</td>
<td>Issue 41 (commit)</td>
</tr>
<tr>
<td>tcpdump.org</td>
<td>tcpdump</td>
<td>✅ Done!</td>
<td>PR543 (commit)</td>
</tr>
<tr>
<td>Yoshiyuki Yamauchi</td>
<td>mrtparse</td>
<td>✅ Done!</td>
<td>PR13</td>
</tr>
<tr>
<td>Wireshark</td>
<td>Dissector</td>
<td>✅ Done!</td>
<td>18172 (patch)</td>
</tr>
</tbody>
</table>

Visit [http://largebgpcommunities.net/implementations/](http://largebgpcommunities.net/implementations/) for the Latest Status
Try It Yourselves!

The following are announced with AS path 2914_15572$

- 192.147.168.0/24
- 2001:67c:208c::/48
- BGP Large Community: (15562:1:1)

Beacon Prefixes

Cisco IOS Output (Without BGP Large Communities Support)

```
route-views>show ip bgp 192.147.168.0
BGP routing table entry for 192.147.168.0/24, version 98399100
Paths: (39 available, best #30, table default)
  Not advertised to any peer
  Refresh Epoch 1
  701 2914 15562
   137.39.3.55 from 137.39.3.55 (137.39.3.55)
    Origin IGP, localpref 100, valid, external
    unknown transitive attribute: flag 0xE0 type 0x20 length 0xC
    value 0000 3CCA 0000 0001 0000 0001
    rx pathid: 0, tx pathid: 0
```

BIRD Output (With BGP Large Communities Support)

```
COLOCLUE1 11:06:17 from 94.142.247.3 [100/-] [AS15562i]
Type: BGP unicast univ
BGP.origin: IGP
BGP.as_path: 8283 2914 15562
BGP.next_hop: 94.142.247.3
BGP.med: 0
BGP.local_pref: 100
BGP.community: (2914,410) (2914,1206) (2914,2203) (8283,1)
BGP.large_community: (15562, 1, 1)
```
INEX Deploys Large BGP Communities in Production

Nov 7, 2016

The Internet Neutral Exchange Association (INEX) is the first network operator in the world to deploy Large BGP Communities in production! Their deployment using BIRD is another important step in its adoption. INEX is a neutral, industry-owned association that provides IP peering facilities in five secure data centers around Dublin, Ireland. The INEX route server routing policy is extended with support to control routing information sent to the route server as follows:

Representation: `inex:action:rsclient`
Any Questions?

barry.odonovan@inex.ie
@comepeerwithme / @barryo79
https://www.inex.ie/
http://largebgpcommunities.net/