



INOG:B - MARCH 7 2017

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 peak
- 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 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

BGP COMMUNITIES - INFORMATIVE

Community: 0:8218 13237:45049 13237:46081

BGP COMMUNITIES - INFORMATIVE

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

...

BGP COMMUNITIES - INFORMATIVE

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

...

BGP COMMUNITIES - SIGNALLING

\$ 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"



RFC8092

BGP Large Communities

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

RIR	RIR Pool	Unadv	Adv	16-bit	Unadv	Adv	32-bit	Unadv	Adv
AFRINIC	931	392	979	265	277	736	666	115	243
APNIC	2583	4895	7518	480	2804	5254	2103	2091	2264
ARIN	2996	9265	17264	1177	8454	15897	1819	811	1367
RIPE NCC	6630	6609	23137	2706	4855	18163	3924	1754	4974
LACNIC	806	935	5170	390	547	2489	416	388	2681
TOTAL	13946	22096	54068	5018	16937	42539	8928	5159	11529

Source: <http://www.potaroo.net/tools/asn32/> - As of March 6 2017

Problem: You can't squeeze 64-bits into 32-bits

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

+	-----	+
	COUNT(*)	
+	-----	+
	13	
+	-----	+

Or, about 15% of members.

This is a serious operational issue - example:

- 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 \Rightarrow 4 BGP sessions each $\Rightarrow \sim(n-1)^2$ sessions
 - 90 IXP members \Rightarrow 89 BGP sessions each $\Rightarrow \sim 7,921$ sessions
- Route servers solve the scaling problem.

Route Servers:



IXP full mesh peering relationships



IXP route server peering relationships

This is a serious operational issue - example:

- 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 \Rightarrow 4 BGP sessions each $\Rightarrow \sim(n-1)^2$ sessions
 - 90 IXP members \Rightarrow 89 BGP sessions each $\Rightarrow \sim 7,921$ 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

Internet Engineering Task Force (IETF)
Request for Comments: 8092
Category: Standards Track
ISSN: 2070-1721

J. Heitz, Ed.
Cisco
J. Snijders, Ed.
NTT
K. Patel
Arrcus
I. Bagdonas
Equinix
N. Hilliard
INEX
February 2017

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.

RFC8092

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-advertise, 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:

Description	Community
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

Implementation on INEX Route Servers - RFC8092:

Description	Community
Prevent announcement of a prefix to a peer	43760:0:peer-as
Announce a route to a certain peer	43760:1:peer-as
Prevent announcement of a prefix to all peers	43760:0:0
Announce a route to all peers	43760:1:0

Major Milestones Towards an RFC Standard

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

BGP Speaker Implementation Status

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

Visit <http://largebgpcommunities.net/implementations/> for the Latest Status

Tools & Ecosystem Implementation Status

Implementation	Software	Status	Details
DE-CIX	pbgpp	✓ Done!	PR16
FreeBSD	tcpdump	✓ Done!	PR213423
Marco d'Itri	zebra-dump-parser	✓ Done!	PR3
OpenBSD	tcpdump	✓ Done!	OpenBSD 6.1 (patch)
pmacct.net	pmacct	✓ Done!	PR61
RIPE NCC	bgpdump	✓ Done!	Issue 41 (commit)
tcpdump.org	tcpdump	✓ Done!	PR543 (commit)
Yoshiyuki Yamauchi	mrtparse	✓ Done!	PR13
Wireshark	Dissector	✓ Done!	18172 (patch)

Visit <http://largebgpcommunities.net/implementations/> for the Latest Status

Try It Yourself! Beacon Prefixes

The following are announced with AS path 2914_15572\$

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

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 First Network Operator to Deploy in Production

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**

THANK YOU

Any Questions?

barry.odonovan@inex.ie

[@comepeerwithme](#) / [@barryo79](#)

<https://www.inex.ie/>

<http://largebgpcommunities.net/>

