

# BGP Large Communities for IXPs

Greg Hankins

greg.hankins@nokia.com

Nokia

# IXP Members Use BGP Communities

- [RFC 1997](#) style communities have been available for the past 20 years
  - Encodes a 32-bit value displayed as: “16-bit ASN:16-bit value”
  - Designed to simplify Internet routing policies
  - Signals routing information between networks so that an action can be taken
- Broad support in BGP implementations
- Widely deployed and required by members for Internet routing

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

INEX RFC 1997 Communities Examples

# Needed RFC 1997 Style Communities, but Larger

- We knew we'd run out of 16-bit ASNs eventually and came up with 32-bit ASNs
  - RIRs started allocating 32-bit ASNs by request in 2007, no distinction between 16-bit and 32-bit ASNs now
- However, you can't fit a 32-bit value into a 16-bit field
  - Can't use native 32-bit ASNs with RFC 1997 communities
- Needed an Internet routing communities solution for 32-bit ASNs for almost 10 years
  - Parity and fairness so everyone can use their globally unique ASN



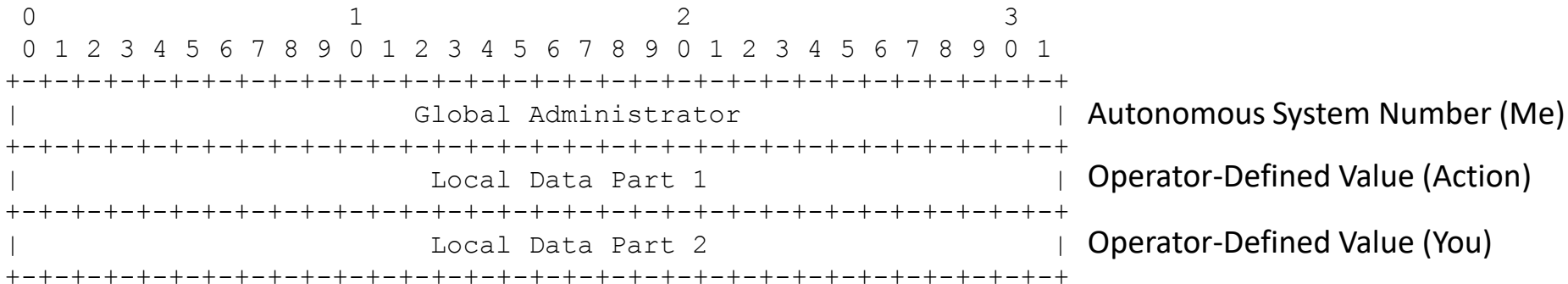
# The Solution: [RFC 8092](#)

## “BGP Large Communities Attribute”

- Idea progressed rapidly from inception in March 2016
- First I-D in September 2016 to RFC publication on February 16, 2017 in just seven months
- Final standard, plus a number of implementation and tools developed as well
- IXPs and their members can test and deploy the new technology now



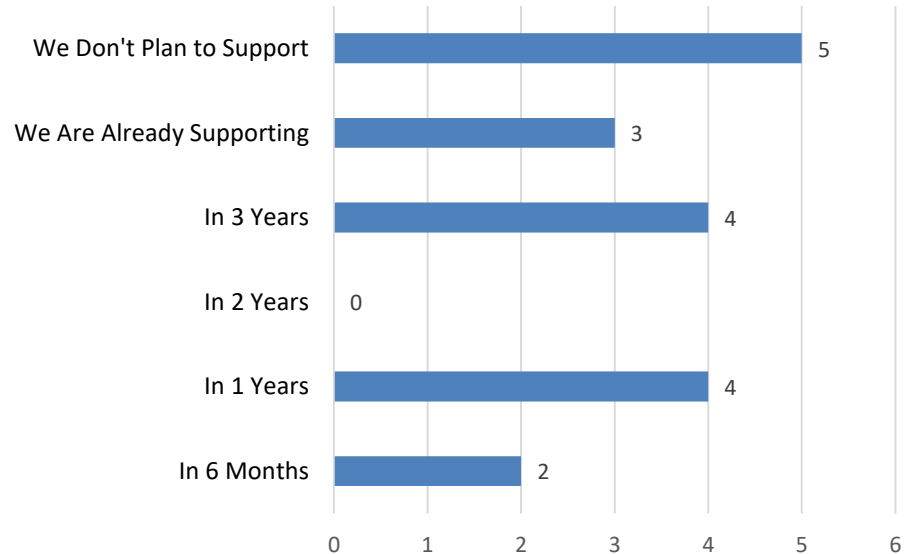
# Encoding and Usage



- A unique namespace for all 16-bit and 32-bit ASNs
  - No namespace collisions between ASNs
- Large communities are encoded as a 96-bit quantity and displayed as “32-bit ASN:32-bit value:32-bit value”
- Canonical representation is \$Me:\$Action:\$You

# Demand for BGP Large Communities at CIX

- CIX 2016 members survey
- 18/31 members responded
- “When do you plan to support BGP large communities?”
- Results show existing use and near-term demand



# Planning for Large Communities

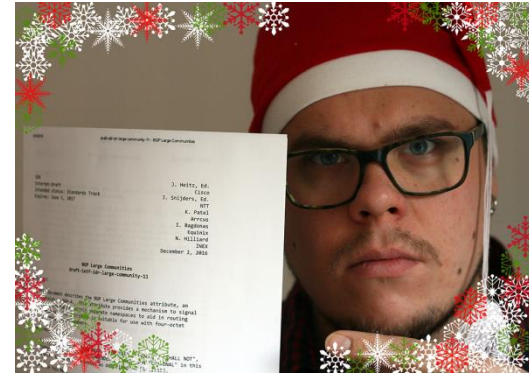
- The entire network ecosystem needs to support large communities in order to provision, deploy and troubleshoot them
- Ask your vendors and implementers for software support
- Update your tools and provisioning software
- Extend your routing policies, and openly publish this information
- Train your technical staff

Image sources: <https://www.sunet.se/blog/all-i-want-for-christmas-is-large-bgp-communities/>

“All i want for christmas is ... Large BGP Communities” by Fredrik "Hugge" Korsbäck

2017-04-11

30th Euro-IX Forum, Barcelona



# Develop a Comprehensive Communities Policy

- Classic RFC 1997 communities will continue to be used together with large communities
  - There's no flag day to convert, large communities simply provide an additional way to signal information
- Your existing routing policy with classic communities is still valid
- Well-known communities such as “no-advertise”, “no-export”, “blackhole”, etc. are still used
- Extend your policy with large communities that allow members to signal the same information as they can with classic communities



# Communities Policy Development

- [draft-ietf-grow-large-communities-usage](#) is a new [RFC 1998](#) style I-D in the IETF GROW Working Group
- Provides examples and inspiration for network operators to use large communities
- Also provides many examples on how to develop a communities policy
  - Informational communities
  - Action communities

# Informational Communities

- An informational label to mark a route with
  - Its origin: ISO 3166-1 numeric country ID and UIC M.49 geographic region
  - Relation or propagation: internal, member, peer, transit
- Provides information for debugging or capacity planning
- The Global Administrator field is set to the ASN that labels the routes
- Most useful for downstream networks and the Global Administrator itself

# Information Communities Example

ISO 3166-1 Country ID		+	UN M.49 Region		+	Relation	
Large Community	Description		Large Community	Description		Large Community	Description
64497:1:528	Netherlands		64497:2:2	Africa		64497:3:1	Internal
64497:1:392	Japan		64497:2:9	Oceania		64497:3:2	Member
64497:1:840	USA		64497:2:145	Western Asia		64497:3:3	Peering
			64497:2:150	Europe		64497:3:4	Transit

- For example, a communities value of “64497:1:528 64497:2:150 64497:3:2” would indicate that it was learned in the Netherlands, in Europe, from a member

# CDN / Eyeball Example – You do a lot with 32 bits!

British Postal Codes (~31 Bits)		or	GPS Coordinates	
Large Community	Postal Code		Large Community	Location
64497:9:849701135	E1W 1LB (London)		64497:10:1281024	Amsterdam
64497:9:1345374681	M90 1QX (Manchester)			(52.37783, 4.87995)

- Location encoding can be used to provide very accurate location information attached to more-specific routes announced to CDN caches
- British postal codes can be encoded by stripping the whitespace and doing a simple base36 to base10 conversion
- GPS coordinates can be encoded with Geohash
  - For example 52.37783, 4.87995 (Amsterdam) encoded with 600 meter precision
  - Python: `import Geohash; Geohash.encode(52.37783, 4.87995, precision=6)`
  - Geohash result: “u173zp”
  - Convert “u173zp” from base32 to base10 = 1281024

# Action Communities

- An action label to request that a route be treated in a particular way within an AS
  - Propagation characteristics: export, selective export, no export
  - Local preference: influence ingress traffic within the AS
  - AS Path: influence traffic from outside the AS
- The Global Administrator field is set to the ASN which has defined the functionality of the community
  - Also is the AS that is expected to perform the action
- Most useful for transit providers taking action on behalf of a member or the Global Administrator

# Action Communities Example

- Selective no export
  - ASN based selective no export
  - Location based selective no export
- Selective AS path prepending
  - ASN based selective AS path prepending
  - Location based selective AS path
- Blackhole
  - Local blackhole
  - Remote blackhole

ASN Based No Export	
Large Community	Description
64497:4:64498	AS 64498
64497:4:64499	AS 64499
64497:4:65551	AS 65551

Location Based No Export	
Large Community	Description
64497:5:528	Netherlands
64497:5:392	Japan
64497:5:840	USA

# Getting Started With Large Communities

- 2018 is the year of large BGP communities
  - Preparation, testing, training and deployment can take some time
  - Start the work now, so you are ready when members want to use large communities
- Lots of resources are available to help IXPs learn about large communities
  - BGP speaker implementations
  - Analysis and ecosystem tools
  - Presentations (<http://largebgpcommunities.net/talks/>)
  - Documentation for each implementation
  - Configuration examples (<http://largebgpcommunities.net/examples/>)

# Large Communities Beacon Prefixes

- The following prefixes are announced with AS path 2914\_15562\$
  - 192.147.168.0/24 ([looking glass](#))
  - 2001:67c:208c::/48 ([looking glass](#))
  - BGP Large Community: 15562:1:1

## Cisco IOS Output (Without 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 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)
```



# BGP Speaker Implementation Status

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

Route Servers are Done!

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

# Tools and Ecosystem Implementation Status

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

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

# Testing Large Communities

- The BGP Large Communities Playground provides an easy way run several implementations together in a lab environment
- Supports BIRD, ExaBGP, GoBGP, Quagga and pmacct
- Docker images are available
- Use the playground to
  - Become familiar with large communities
  - Test interoperability with your vendor's BGP implementations
  - Design, configure and verify your new community policies

# Questions?

Presentation created by:



Greg Hankins

Nokia

[greg.hankins@nokia.com](mailto:greg.hankins@nokia.com)

[@greg\\_hankins](#)



Job Snijders

NTT Communications

[job@ntt.net](mailto:job@ntt.net)

[@JobSnijders](#)

Visit <http://LargeBGPCommunities.net/> for the Latest Info

**Reuse of this slide deck is permitted and encouraged!**

# Configuration and Output Examples

# BIRD Configuration

```
# match
```

```
if ((8283, 1, 2) ~ bgp_large_community) then return true;
```

```
# scrub / delete
```

```
bgp_large_community.delete([(8283, *, *)]);
```

```
bgp_large_community.delete([(8283, 0, 1)]);
```

```
# set
```

```
bgp_large_community.add((8283, 0, 100));
```

```
bgp_large_community.add([(8283, 0, 100), (8283, 2, 333)]);
```

# IOS XR Configuration (EFT – Beta “Just Like Community”)

```
# match
route-policy set-something
  if large-community matches-any (8283:4:3) then
    set local-preference 120
  endif
end-policy
```

```
# scrub / delete
route-policy set-something
  delete large-community in (8283:*:*)
  delete large-community in (8283:4:3)
end-policy
```

```
# set
route-policy set-something
  set large-community (8283:45:29) additive
end-policy
```

# Nokia SR OS Configuration

```
policy-options
  community "set" members "8283:45:29"
  community "match" members "8283:4:3"
  community "delete" members "8283:4:3"
```

```
policy-statement "set-something"
  entry 10
    description "match"
    from
      community "match"
    exit
    action accept
      local-preference 120
    exit
  exit
  entry 20
    description "scrub / delete"
    action accept
      community remove "delete"
    exit
  exit
  entry 30
    description "set"
    action accept
      community add "set"
    exit
  exit
exit
```



# OpenBGPD Configuration

## # match

```
allow from any large-community 8283:1:2
match from any large-community 8283:1:2 set localpref 300
deny to any peer-as neighbor-as \
    large-community 8283:6:neighbor-as
```

## # scrub / delete

```
match from any set { large-community delete 8283:*:* }
match from any set { large-community delete 8283:1:2 }
```

## # set

```
match from any set { large-community 8283:1:2 }
match from any set { large-community 8283:1:2 \
    large-community 8283:4034:24824 }
```

# tcpdump 4.9.0 Packet Capture

```
# ./tcpdump -i eth3 -n -v -c 1 src port 179
tcpdump: listening on eth3, link-type EN10MB (Ethernet), capture size 262144 bytes
16:22:08.992920 IP (tos 0xc0, ttl 64, id 41807, offset 0, flags [DF], proto TCP (6), length 181)
 94.142.247.3.179 > 94.142.247.6.33785: Flags [P.], cksum 0xabce (incorrect -> 0x1e40), seq
58743671:58743800, ack 2012368616, win 2270, options [nop,nop,TS val 857977378 ecr 149127175],
length 129: BGP
    Update Message (2), length: 129
      Origin (1), length: 1, Flags [T]: IGP
      AS Path (2), length: 34, Flags [T]: 38930 1299 3910 721 27065 1554 1555 1501
      Next Hop (3), length: 4, Flags [T]: 94.142.247.3
      Multi Exit Discriminator (4), length: 4, Flags [O]: 0
      Local Preference (5), length: 4, Flags [T]: 100
      Atomic Aggregate (6), length: 0, Flags [T]:
      Aggregator (7), length: 8, Flags [OT]: AS #1501, origin 144.105.202.0
      Community (8), length: 8, Flags [OT]: 1299:20000, 8283:14
      Large Community (32), length: 12, Flags [OTP]:
      8283:6:14
    Updated routes:
      136.210.249.0/24
```

# Wireshark 2.3.0 (Prerelease) Packet Capture

▼ Path Attribute - LARGE\_COMMUNITY: 65535:1:1 4294967295:4294967295:4294967295

▷ Flags: 0xc0, Optional, Transitive: Optional, Transitive, Complete

Type Code: LARGE\_COMMUNITY (32)

Length: 24

▼ Large communities: 65535:1:1

Global Administrator: 65535

Local Data Part 1: 1

Local Data Part 2: 1

▼ Large communities: 4294967295:4294967295:4294967295

Global Administrator: 4294967295

Local Data Part 1: 4294967295

Local Data Part 2: 4294967295

▷ Network Layer Reachability Information (NLRI)

0000	02 42 c0 00 02 03 02 42	c0 00 02 02 08 00 45 00	.B.....B	.....E.
0010	01 ab 39 26 40 00 40 06	7c 21 c0 00 02 02 c0 00	..9&@.e.	1.....
0020	02 03 b6 0d 00 b3 4d 13	33 db f4 01 bc ba 80 18	.....M.	3.....
0030	01 c9 23 2f 00 00 01 01	08 0a 00 01 0b 0f 00 01	..#/.....	.....
0040	0b 0f ff ff ff ff ff ff	ff ff ff ff ff ff ff ff	.....	.....
0050	ff ff 00 4b 02 00 00 00	2f 40 01 01 00 40 02 06	...K....	/e...e..
0060	02 01 00 01 00 00 40 03	04 c0 00 02 02 c0 20 18	.....e.	.....
0070	00 00 ff ff 00 00 00 01	00 00 00 01 ff ff ff ff	.....	.....