

Routing w/ BGP Communities

Akira Kato



WIDE Project
kato@wide.ad.jp

Extensive (but minimum) punching hole

- ☆ **We'd like to put traffic on the fattest pipe**
 - BW is one of the major concerns
 - JP may want to send traffic via US to EU
- ☆ **Some applications prefer pipe with the least delay**
 - Interactive applications, for example
 - Path with the least delay may not have biggest BW
- ☆ **PBR may not be good**
 - It's a "static" routing
 - Big performance impact may happen

Extensive (but minimum) punching hole

☆ So proposed scheme is

- Put regular traffic to fattest path
 - e.g. For JP-EU, traffic is sent via US
- For interactive applications, do "punching a hole"
 - Corresponding longer prefixes advertised
 - What is the longest prefix cover the range?
- Advertise it along a path with smallest delay
 - e.g. the prefix advertised through TEIN-2
 - How to control the advertisement is a key
- Care must be taken not to leak to commercial ISPs

BGP Communities

☆ Defined in RFC1997

- Carries a set of 32bit values

☆ Extended Communities defined in RFC4360

- Carries a set of 48bit values w/ Type

☆ Need to explicitly configuration to advertise them

- In some routers, such as cisco

```
neighbor x.x.x.x send-community
```

Usage of BGP Communities

☆ Communities can be set by the origin or a transient AS

- May be removed once interpreted
- May not necessary be removed, however

☆ A few values are predefined

```
NO_EXPORT :      0xfffff01
NO_ADVERTISE :  0xfffff02
NO_EXPORT_SUBCONFED : 0xfffff03
```

☆ No globally coordinated usage for lower 16bits

- We may need a consistent set for the R&E community

Usage of BGP Communities

☆ 32bit is usually divided into two 16bit values

- e.g. 1234:5678

☆ Typically higher 16bit indicates the target ASN

- For AS1234, please handle the prefix as "5678"
- The lower 16bits are defined per ASN
 - No common value set is defined globally
 - In some AS following values set to local preference
 - 90 : backup, lower priority
 - 100 : peer, middle priority
 - 110 : customer, higher priority

BGP Community Operation

☆ Many routers support "Exact Match" on communities

```
ip community-list MY-CUST permit 2500:110
```

```
route-map WIDE-IN permit 10  
  match community MY-CUST  
  set local-preference 110
```

☆ Some routers support Regexp Match

```
ip community-list expanded MY-COMMUNITY permit 2500:[0-9]+
```

☆ Capability is limited

- No rewriting rules are configurable
 - Unlike sendmail.cf
- No calculation machinery supported
 - Unlike perl/awk/...

Proposed Usage of BGP Communities

☆ One proposal was by Brent Sweeny from IU

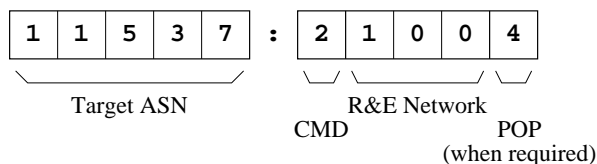
- BGP path 'hinting' proposal
- sweeny-hinting-mm.pdf available at RENOG web page
- Defines AS-path-hint-value
- Possible hint values (example)

```
AARNET          64101  
APAN/TransPAC  64102  
ASNET           64103  
CANET          64104  
...            ...
```

Kato's Proposal on Lower 16bits

☆ Divide the lower 16bits into three parts

- Not bit-boundary but decimal boundary
- Regexp is supported on decimal representation
- Easy to read by the human operators



☆ R&E Network Number need a registration service

- Unique number in range of 100 .. 990

Proposal on lower 16bits

☆ Command digit

- 1 : don't accept
- 2 : accept as lower priority
- 3 : accept as higher priority
- 4 : don't advertise
- of course as long as local policy permits

☆ POP digit

- 0 : entire network
- 1-9 : identify each POP
- If more than 9 POPs exist,
 - Another R&E number is assigned

Kato's Proposal on Lower 16bits

☆ Example values

- R&E Number
 - 100 Abilene (AS11537)
 - 101 TransPAC (AS22388)
 - 102 ASNET (AS9264)
 - ...
- POP for 100
 - 1 : Abilene at New York (MANLAN)
 - 2 : Abilene at Chicago (StarLight)
 - 3 : Abilene at Seattle (PWave North)
 - 4 : Abilene at LA (PWave South)

Kato's Proposal on Lower 16bits

☆ Let prefixes route through LA

```
route-map AS11537 permit 10
  match ip address prefix-list ABILENE-LA-priory
  set community 11537:31014
```

☆ In Abilene LA router (in cisco syntax, sorry)

```
ip community-list expanded MY-HIPRI permit 11537:31000
ip community-list expanded MY-LA-HIPRI permit 11537:31004
```

```
route-map AS11537-IN-LA permit 30
  match community MY PRIORITY MY-LA-HIPRI
  set local-preference 120
```

Configuration

- ☆ **The community list and rout-map configuration**
 - Could be large and complex to manual description
 - Language differs in each router vendor
 - Generator script is required to prevent "typos"
 - Need to care about conflicts with existing policy
- ☆ **Registration of R&E AS and the POPs if required**
 - Any volunteer? APAN/JP? Global NOC?

Registration

- ☆ **How the registration result is disclosed?**
 - Through "whois" service
 - Dumps the entire table
 - Through DNS
 - AXFR in TCP or "getnext" like query in UDP

```
$ORIGIN communities.renog.org
0 IN TXT "20070827 0325"
1 IN TXT "11537 1000 ABILENE"
2 IN TXT "11537 1001 ABILENE CHICAGO"
33 IN TXT "EOF"
```

Leakage prevention

☆ We need a common "R&E" community value

- All R&E routes should be with it
- Easy to accommodate a filter to ISPs
 - Not to advertise to them

☆ What value shall we use?

- Any volunteer for offering some community space?
 - 7660:7660, for example
 - May better to use RENOG ASN (waste an ASN)
- What to do with non-R&E but non-commercial prefixes?
 - Root DNS servers, ...

BGP Community

☆ It can be driven by users

- Dynamically (order of 10min) control the route

☆ The finer granularity of routing control/selection

- An user want to try "different" path
 - For debugging
 - For workaround of non-serious packet losses
- Some R&E applications will have benefit
 - "any path with less packet loss rate"

Considerations

- ☆ **The BGP community proposed here**
 - Applicable to IPv6 as well
 - Any consideration on multicasting on IPv4/IPv6?
- ☆ **It uses BGP (non-extended) communities**
 - How we support 4-byte ASNs?
 - Not our special problem, however
- ☆ **Anyway, it require support of route observation**
 - RouteView and/or ComPATH
 - As it doesn't prevent routing problem...