Geoff Huston
November 2016

BGP Large Communities

IPv4 addresses are not the only Internet number resource that has effectively run out in recent times. Another pool of Internet numbers under similar consumption pressures has been the numbers that are intended to uniquely identify each network in the Internet's inter-domain routing space. These are Autonomous System numbers (ASNs).

Like IPv4 address runout, the exhaustion of the ASN number pool was a well forecasted event, and in the same way that IPv6 was developed in response to address runout, changes have been made to the BGP routing protocol specification almost a decade ago to address this other runout problem. The specific change for ASNs was to allow BGP implementations to support an ASN field that had doubled in size, from 16-bits (or two-octets) to 32-bits (four-octets).

Unlike IPv4 and IPv6, some effort was made to support a limited form of backward compatibility in BGP, allowing older BGP implementations to interoperate with new BGP implementations in ways that were largely compatible. The approach was to perform a simple form of ASN translation. A two-octet ASN BGP speaker would see all four-octet ASNs as instances of the same two-octet ASN, namely AS23456. In addition, the original four-octet AS path would be tunnelled across two-octet ASN domains by adding an opaque transitive path attribute to BGP. This way the four-octet supporting versions of BGP could be deployed incrementally, without any overall orchestration.

This transition in the ASN number space was managed very effectively. The relatively small pool of BGP speakers (certainly small as compared to the pool of IPv4 users) and their vendor community were able to field new versions of BGP, and we were able to start incrementally using four-octet ASNs by 2007, well before we completely ran out of the original two-octet ASN number pool. So disaster was averted in plenty of time, the Internet was saved once more and we all headed to the bar!

Right?

Well, not really.

The issue is that the transition appears to have stalled, as it appears that a significant number of networks want to continue to use two-octet ASNs.

Figure 1 shows the number of allocated ASNs on each day since January 1997. Four-octet ASNs were introduced to the registry system in 2007, yet the rate of allocation of two-octet ASNs continued unabated until late 2013. A year later some of the RIRs, notably ARIN, RIPE NCC and APNIC, implemented ASN transfer policies. The outcome of these policies is evident in this figure, where the number of unadvertised ASNs has declined from a peak of 7,300 in mid-2013 to a current count of 3,300. Some 4,000 two-octet ASNs are now back in the BGP system as announced ASNs.

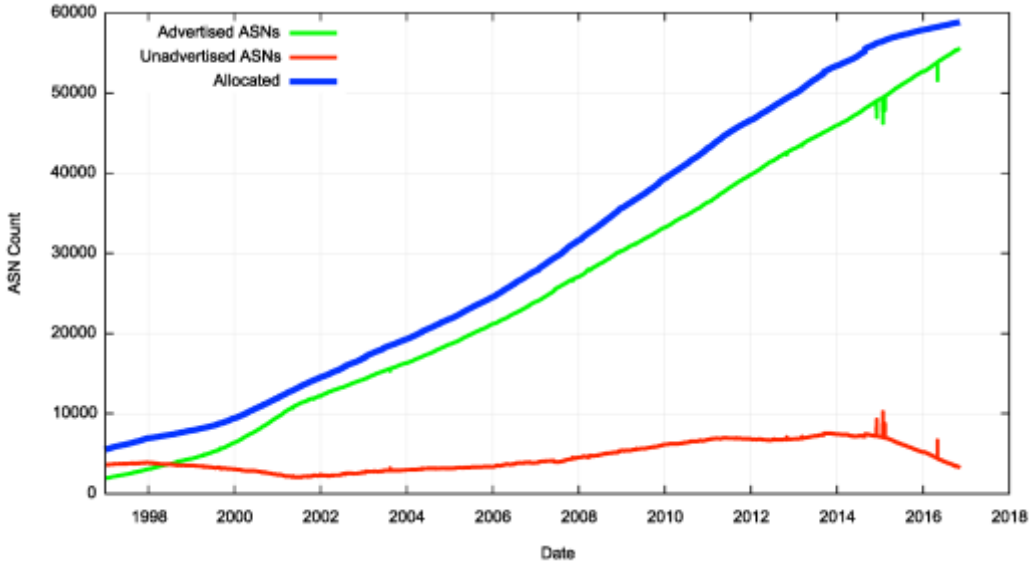

Figure 1 – Time Series of Two-octet ASN Number Use

If four-octet ASNs were truly equivalent in every way to two-octet ASNs then they would have no particular value over and above four-octet ASNs, and we probably would not see the pool of unadvertised two-octet ASNs declining in recent months. However, this appears to not be the case, and two-octet ASNs are clearly preferred by some networks. Why could this be?

The answer could well be contained in the state of BGP Communities.

BGP Communities are an instance of a BGP Update Message Path Attribute. A set of values in a BGP Communities attribute of a BGP Update Message is associated with all of the network prefixes, or "routes", listed in the Update. It is commonly used by network operators to express a desired handling for routes that could not otherwise be expressed in BGP. In particular, it is intended to be used as an additional signal that may control how these routes are to be accepted by other BGP speakers, to influence the preference of a route, to control the onward distribution of a route, or to describe other attributes of a route that are not directly expressed in specific Path Attributes.

The original specification for BGP Communities, RFC 1997, defined the Communities attribute as a variable length attribute consisting of a set of four-octet values. There were few constraints imposed on what could be used as a value for a community. The specification did define three common community attribute values:

- **NO_EXPORT**, an attribute that in intended to prevent the routes from being distributed outside of an AS, or, in the case of a networtk that uses the BGP Confederation construct, out of the BGP Confederation.

- **NO_ADVERTISE**, an attribute intended to prevent the route from being advertised to any BGP peer.

- **NO_EXPORT_SUBCONFED**, an attribute intended to prevent the route from being advertised to any eBGP peer, including eBGP peers in the same BGP Confederation

This set of defined community values has been expanded since the publication of RFC1997, and there is a maintained IANA registry of "well-known" four-octet community values that is maintained at http://www.iana.org/assignments/bgp-well-known-communities/bgp-well-known-communities.xhtml.

BGP Communities were subsequently extended in RFC4360 in the form of "Extended Communities". This refinement to the Communities specification added additional control flags in the form of a Type field of one or two octets, and a value field that used the remaining space up to a fixed total size of 8 octets for each community value instance. There were a number of variants in the interpretation of the value field based on the type flags. One variant allowed the specification of a two-octet ASN in the "Global Administrator" sub-field, and the remaining four octets contained a value that was specific to the ASN. This variant appears to be widely used by network operators to implement a rich policy language to BGP.

For example, if you want to advertise a route towards AS2914 (NTT America) and alter this network's local preference setting, then you achieve this by selecting a community value defined by https://www.us.ntt.net/support/policy/routing.cfm. An aggregate collection of published BGP Extended Community values that influence the handling of a route object by a number of networks can be found at https://onestep.net/communities/. It appears that common use of the BGP Extended Communities attribute uses the first two octets to denote the target network where the policy setting is to be invoked, and the remaining four octets express an intended policy.

This construct can allow very precise policy controls on the propagation of a route in BGP. For example, the four-octet second field of the extended community may itself contain a two-octet peer ASN and a two-octet action, so that, for example, you may be able to direct the network AS 64496 to prepend its ASN three times to the AS path of all routes propagated towards the peer AS 64500 by using the BGP Extended Community value 64496:64500:3. Another use case is in defense against various DDOS attacks, using a technique called "Remotely Triggered Black Holes", where the network under attack may opt to push a traffic drop filter into its peers and further out in the network to allow the attack traffic to be dropped closer to its origination points. The target route is filtered early, allowing otherwise legitimate traffic directed to other routes in the victim's network to have a highly likelihood of successful transmission (RFC5635).

The reason why BGP Extended Communities have proved to be so useful is that BGP operates in reverse. In other words, a BGP route propagates across the inter-domain routing space in one direction, while the traffic directed towards the destinations encompassed by this BGP route flows in the opposite direction. Without BGP communities it is challenging to orchestrate incoming traffic across multiple potential paths such that the traffic load is even spread across all paths. Network operators use selective advertising of more specific routes and AS prepending in an effort to bias the remote network's path selection process, but such an effort can be challenging and the outcomes are often fragile. In response many of the larger ISP transit providers have added BGP Extended communities to their service offerings, allowing their ISP customers to trigger particular responses. For example, a multi-homed customer could use two upstream transit ISPs, and use one to prefer to receive traffic originating in Europe and the other to prefer traffic originating in Asia.

While BGP Extended Communities have wide acceptance in the network operations area, they are effectively limited to express policy referring to two-octet ASNs. If one were to use a four-octet ASN in the Global Administrator sub-field of a BGP Extended Community value, this would only leave two octets for the policy action, which appears to be inadequate for many of today's use cases. The result was that in those parts of the Internet where two-octet ASN use in BGP Extended communities was well established, notably North America, there was little demand for four-octet ASNs. As the pool of remaining two-octet ASNs dwindled there was demand to open the pool of two-octet ASNs to trading and transfers, but this was inevitably only a stopgap measure.

What was needed is a standard and widely implemented method of specifying BGP Community values with the same expressive capability as Extended Communities, but able to use four-octet ASNs.

There have been a number of efforts to address this need. Four-Octet AS Specific Extended Communities (RFC5668) proposed a minor change to the Extended Communities specification by lengthening the Global Administrator sub-field from two to four octets. However, this implied that the Local Administrator-defined value field was reduced from four to two octets. Flexible BGP communities (https://tools.ietf.org/html/draft-lange-flexible-bgp-communities) tried a different approach, using four-octet Originating ASN field, and a variable-length value, allowing up to 255 octets of an associated value. This was not widely supported at the time (around 2004) and did not progress any further. Wide Communities (https://tools.ietf.org/html/draft-ietf-idr-wide-bgp-communities) took a similar stance, with a Type-Length-Value approach to each community value. The two-octet length field permitted a maximum value field of 65,535 octets. Wide Communities certainly appear to have taken a kitchen sink approach to BGP communities, attempting to offer what is essentially a universal mechanism for expressing BGP communities.

But sometimes when the hammer you are using is too small for the job, it's not a variably-regulated customizable percussive tool that you are after. It's just a bigger hammer!

Such a bigger hammer is the subject of an internet draft that proposes BGP Large Communities (https://tools.ietf.org/html/draft-ietf-idr-large-community). This proposal takes the 6-octet payload of the BGP Extended Communities specification (RFC4360) and simply doubles the size of the individual field values, so that the Large Community values are each 12 octets in length. The proposal also applies a simplification to remove the Type field, a feature that appears to be largely unused by the operator community. The variable length approach of Wide Communities has been avoided. What is left could be seen as a Community structure that is close to the original approach of RFC1997, but with sufficient length to be useful for four-octet ASNs in the same style of Extended Communities. This means that the format of the Large Community value has a four-octet Global Administrator field capable of carrying a four octet ASN, and an eight-octet Local Data field, whose interpretation is defined by the Global Administrator (Figure 2). This is an optional transitive path attribute.

```
The attribute consists of one or more twelve-octet values.  Each
twelve-octet Large BGP Communities value represents three four-octet
values, as follows:

 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      Global Administrator                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Local Data Part 1                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Local Data Part 2                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

Global Administrator:  A four-octet namespace identifier.  This
   SHOULD be an Autonomous System Number.

Local Data Part 1:  A four-octet operator-defined value.

Local Data Part 2:  A four-octet operator-defined value.
```

Figure 2 – BGP Large Communities – draft-ietf-idr-large-community-06.txt

The result is a means of associating community values to routes in the presence of four-octet ASNs with exactly the same expressive capability as that provided by Extended Communities for two-octet ASNs.

The draft specification progressed to the point of a provisional IANA code allocation of attribute code 30 for these large community attributes and then it emerged that the vendor Huawei has been squatting on this same BGP attribute code value in its deployed equipment. Such blatant disregard for the IETF Standards progress by any vendor is to be deplored, and Huawei is no exception here, but at the same

time expediency demands working around such clashes and draw a different number from the BGP attribute number space. The BGP Path Attribute code for Large Communities is now 32.

Hopefully, we are close to an acceptable simple solution to this issue of using four-octet ASNs in BGP Communities. And if we are so close to an outcome that addresses this issue of four-octet ASN BGP Communities, then perhaps we can now stop looking at four-octet ASNs as second class values and use them in precisely the same way was we become accustomed to using two-octet ASNs.

It's about time we completed this work. There are 3,172 available two-octet ASNs left in the various RIR-administered pools. ARIN has none, the RIPE NCC has 2,283, APNIC has 471, LACNIC has 339 and Afrinic has 79. At this time, particularly in North America, we are down to the bottom of the barrel, and we have already "mined" one half of the old unadvertised two-octet ASNs and recycled them back into BGP. This work on usable four-octet ASN communities is already very late, and we need to complete the work on vendor support, dev-ops tools, and operational deployment as soon as we possibly can.

For some further information on this proposal http://largebgpcommunities.net/ contains a useful collection of material and status of implementations.

Author

Geoff Huston B.Sc., M.Sc., is the Chief Scientist at APNIC, the Regional Internet Registry serving the Asia Pacific region. He has been closely involved with the development of the Internet for many years, particularly within Australia, where he was responsible for building the Internet within the Australian academic and research sector in the early 1990's. He is author of a number of Internet-related books, and was a member of the Internet Architecture Board from 1999 until 2005, and served on the Board of Trustees of the Internet Society from 1992 until 2001 and chaired a number of IETF Working Groups. He has worked as an Internet researcher, as an ISP systems architect and a network operator at various times.

www.potaroo.net

Disclaimer