Network Working Group J. Haas
Internet Draft NextHop
Expiration Date: August 2001 22 February 2001

Autonomous System Number Substitution on Egress
draft-jhaas-ase-00.txt

1. Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of Section 10 of RFC2026.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

2. Overview and Rational

Network Reachability information in the Internet is curretly exchanged through the use of the BGP-4 protocol[1]. BGP Speakers require an Autonomous System (AS) number in order to peer. The AS number is used to identify sets of routes sharing a common administrative policy.

Studies of the current allocation patterns of Autonomous System numbers have projected that all available Autonomous System numbers will be exhausted by 2005. As noted by recent presentations at IETF and NANOG, the allocation pattern has been roughly exponential.[2]

The CIDR report[3] has shown that a large number of ASs are

advertising only a single prefix into the global routing system. Additional data supplied by CAIDA[4] suggest that a large number of leaf AS's advertise a relatively small number of prefixes.

This data suggests that the large increase in the usage of AS numbers is due to small networks multihoming themselves. The reasons for the current increase in the rate of multihoming is outside the scope of this document.

Most of these ASs that announce a small number of prefixes could be adequately served in their multihoming by having each of their upstream providers originate the route. There are several issues that make this problematic:

+o RFC 1930[5], Section 7 states that a prefix should not be originated from more than one AS. However the CAIDA data also notes several ASs that already violate this. Operationally this may not be an issue where the networks originating the prefix are doing this for a stub AS.

+o Methods by which the upstream AS could originate the prefix:

a. Static configuration. This can result in blackholes if the customer link goes down.

b. Running an IGP on the customer link. This addresses the blackhole issue, however due to security concerns many providers do not wish to run an IGP on a customer link. Additionally, a flapping customer link would affect internal routing convergence.

c. Running BGP on the link using a private AS number and utilizing the ability of the ISP router to strip the private AS. This solves problems a and b. This unfortunately removes the ability of the customer to bias their incoming traffic by adjusting AS_PATH length.

In order to address AS number depletion, a recent Internet Draft (draft-chen-as4bytes-00 [6]) suggests extending the AS Path component size from 2 octets to 4 octets. This would address the issue of AS number depletion, but requires wide deployment throughout the Internet to be most useful. The method suggested by this draft works at the edges of a partcipating network and doesn't require additional functionality to be added to non-participating routers.

RFC 1965[7] (recently updated), AS Confederations for BGP,

introduced the concept of modifying a route's AS_PATH to remove confederation member AS numbers when that route is advertised outside of the AS confederation. The member AS numbers are replaced with the AS number for the AS confederation, thus representing the group of ASs as a single AS.

This draft recommends a simple modification similar to AS Confederations that helps conserve AS numbers.

3. Discussion

This draft attempts to address the issue of AS number exhaustion issue for a large and growing class of BGP speakers. This class includes entities that are multihoming to more than one network and do not provide transit service between the networks that they multihome to -- in other words, stub ASs.

RFC 1930 reserves AS numbers 64512 through 65535 for private use. These AS numbers should never be found in the global Internet routing tables.

The following diagram represents the typical customer multi-homing scenario:

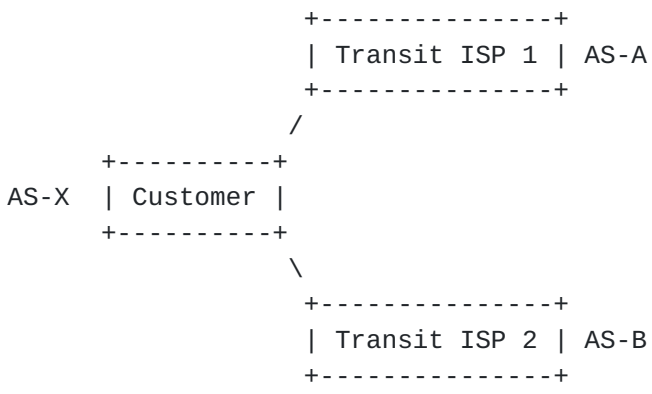

Figure 1

In the normal multi-homing scenario, the customer would need to request an AS number from its Regional Internet Registry (ARIN, RIPE, APNIC, etc.). This draft proposes that the customer is assigned a private AS number that is mutually agreeable to its transit providers. The customer may then originate its routes normally with BGP-4 using this private AS number.

The transit routers, when re-advertising routes originated from

the Customer AS will substitute its own AS number for each occurence of the customer's private AS. This is called AS number Substitution on Egress (ASE).

It can be noted that this methodology is analogous to two BGP confederations with an overlapping member AS.

ASE is intended to be applied only to non-transit AS's. As such it strictly prohibits advertisement of routes containing any AS number that is not the mutually agreed Customer AS number. If a router performing ASE for a peer has received a route that contains non-peer AS numbers in the AS_PATH, the router must terminate the peering session with a notification message of "Malformed AS Path Attribute".

In order to provide loop detection for the customer AS by proxy, a new non-transitive attribute will be added to the route when it is re-advertised.

4. ASE-ORIGINATOR attribute

This document creates the ASE-ORIGINATOR path attribute. This attribute is an optional transitive attribute with a fixed length of 10 octets. The attribute consists of three components:

a. The originating AS number (customer AS) which is two octets and is part of the private AS space.

b. The AS that is performing the ASE. Should this AS be a member of a BGP confederation, the AS Confederation Identifier should be used. This is inserted to guarantee uniqueness of the ASE-ORIGINATOR across the Internet.

c. The ASE client identifier, which is four octets. Unless configured otherwise, the ASE client identifier should default to the BGP Identifier of the peering session.

The Customer AS number and the ASE client identifier must be mutually agreed upon by the transit ISPs.

A BGP-4 speaker who is configured to perform ASE must not re-advertise a route to an ASE client when that route contains the ASE-ORIGINATOR attribute containing the peer's AS number and ASE client id.

The ASE-ORIGINATOR attribute has Type Code 255. (To be assigned by IANA.)

5. Implementation Issues

When a route is received by a BGP-4 speaker, administrative policy is often used to determine whether or not a route will be placed into the router's AdjRibsIn. Some of this policy is implemented based on filtering on the contents of the AS_PATH of the route. It is important for the router to retain the route in its original form so that filtering can happen normally. Performing the AS number substitution prior to egress can make it difficult to apply proper filtering.

Additionally, if policy were to change it is useful to be able to run policy on the originating customer AS number rather than on one's own AS number with additional criteria.

6. Operational Considerations

a. Receiving an ASE peer AS number from internal BGP peers:

AS Numbers in the private AS number space are often used for many things within a given network. For example, they are used as the AS number for a BGP confederation member. Additionally, private ASs are often used for stub BGP peers.

Thus, even though a router performing ASE for a peer will never propagate the AS number of the peer undergoing ASE (and thus knowledge of the ASE peering AS is localized to that router), this router may still receive BGP updates containing the AS number that is used for the ASE peering session from other BGP speakers.

Care must be taken by operators of routers running ASE when constructing policy for routes received from other BGP speakers.

Example of this issue:

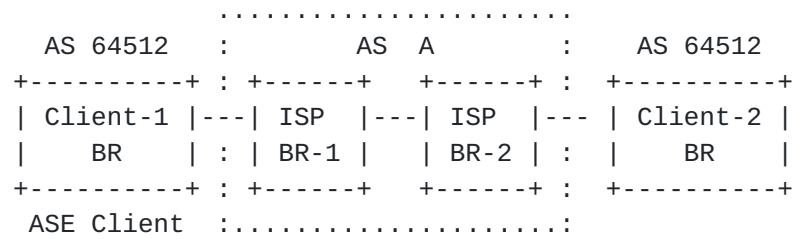

Figure 2

This diagram shows an AS with two external peers, both with AS 64512, which is a Private AS number. ISP BR-1 is performing ASE, ISP BR-2 is not. In this configuration, ISP BR-1 will receive routes from Client-1's border router with an AS_PATH containing one or more instances of 64512. When ISP BR-1 re-advertises the route to ISP BR-2, these instances of 64512 will be substituted with the AS number A.

ISP BR-1 will receive from ISP BR-2 routes containing AS 64512. Since the peering session between ISP BR-1 and ISP BR-2 is IBGP, the AS_PATH will not be modified. It is important for ISP BR-1 to be aware of this and deal with policy appropriately. Additionally, Client-1 will discard any routes propagated by ISP BR-1 that came from Client-2 since its own AS number occurs in the AS_PATH.

In general, having the same private AS used by peers of your AS that may also be used internally for ASE may cause problems. Care should be taken so this doesn't happen. This could be done, for example, by allocating a range of Private ASs that will only be used for ASE within one's AS.

b. Route Looping and Prevention:

Since the AS path information is substituted by the transit routers on egress, an ASE Client may receive an announcement of its own NLRI from the upstream routers. Since the AS_PATH has been modified to remove the private AS of the customer, standard AS_PATH loop detection will not work.

The ASE-ORIGINATOR attribute is meant to provide loop prevention by a router performing ASE from propagating known loops. Misconfiguration of an ASE speaker, for example by configuring the BGP peering session as a normal external peering session without ASE, may lead to this.

This may result in the customer router containing AdjRibsIn entries for its own NLRI. These routes will not usually become active due to the default route selection criteria of BGP-4. However, in the event of misconfiguration, route loops may take place if the externally received route is installed in preference to the internal routes.

Example:

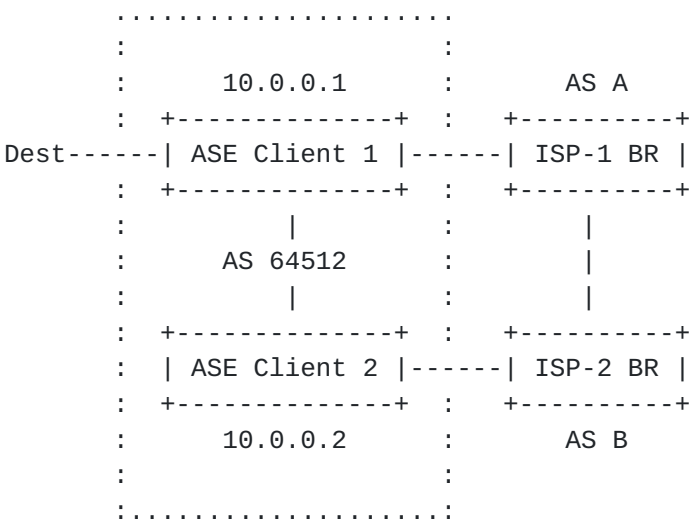

Figure 3

In this example, AS X is an ASE client of AS A and B. ASE Client 1 has a BGP Identifier of 10.0.0.1 and ASE Client 2 has a BGP Identifier of 10.0.0.2. To prevent loops, AS A and AS B should both agree on the same BGP Identifier for the ASE Client Identifier.

i. AS X, an ASE Client of AS A and AS B, advertises Dest to AS A. AS A thus has a route of Dest with a path of <X>.

ii. AS A re-advertises the route to AS B and performs the ASE on it. It appends the ASE-ORIGINATOR attribute of 64512:A:10.0.0.1 (where A is the AS number of AS A) to the route. AS B thus has the route Dest with a path of <A>.

iii. AS B is either misconfigured to not use ASE on this peering session, or has an incorrectly configured ASE Client Identifier and thus re-advertises the route to ASE Client 2. ASE Client 2 thus has the route Dest with a path of <B, A>.

iv. Normally this would be a loop and the route would be dropped. However, the path information to prevent loops has been lost. AS X must filter on the prefixes it advertises (normally a good thing) to prevent this route from being installed in its AdjRibsIn. More importantly, AS X must ensure that policy does not

select this route as being active and thus leading to a routing loop.

One case where this behaviour may be useful is in the case of a customer AS partition. This allows the customer AS to reach itself via its transit ASs.

c. ASE clients must NEVER re-advertise BGP routes they learn.

In the example above, if AS X receives the route Dest<B, A> and then re-advertises the route to AS A, it will lose its peering session with AS A. This is due to AS A receiving an AS_PATH from the ASE Client that contains an AS that is not the configured AS number of the ASE Client.

To reiterate, ASE clients must be configured to avoid propagating externally learned routes to peers. This behaviour, although operationally troublesome, is to prevent a stub AS with a Private AS number from becoming a transit AS.

7. Summary

ASE provides a simple mechanism to help slow the exhaustion of AS Numbers. ASE is very simple mechanism to implement and needs only be deployed at the edges of the network. Routers that are not participating in ASE do not need to understand ASE.

In short, ASE is meant to provide an analog to the benefits of Network Address Translation (NAT) at the AS level.

As noted throughout this draft, misconfiguration of routers performing ASE can lead to an ASE client receiving its own NLRI without enough information to perform loop detection and drop the route. However, the ASE mechanism prevents such loops from affecting the wider Internet by preventing re-advertisement of routes that are not locally originated.

8. Security Considerations

All security considerations of the BGP-4 protocol apply. In addition, ASE "hides" the originating entity and may cause parties who are troubleshooting routing issues to contact the transit ISP when contacting the customer directly may have sufficed.

9. Acknowledgements

The author wishes to thank Matt Richardson of Nexthop for valuable

comments.

10. References

[1] RFC 1771 - A Border Gateway Protocol 4 (BGP-4). Y. Rekhter, T. Li.

[2] NANOG-21 presentation on Global Routing System Scaling Issues. February 2001. http://www.nanog.org/mtg-0102/ppt/cathy/index.html

[3] The CIDR report. Author Tony Bates. http://www.employees.org:80/~tbates/cidr-report.html

[4] CAIDA study of Stub AS's injecting a few (< 10) routes. http://www.caida.org/~broido/bgp/stubas.html

[5] RFC 1930 - Guidelines for creation, selection, and registration of an Autonomous System (AS). J. Hawkinson, T. Bates.

[6] draft-ietf-idr-as4bytes-00 - BGP support for four-octet AS number space. Quaizar Vohra, Enke Chen.

[7] RFC 1965 - Autonomous System Confederations for BGP. P. Traina.

[8] draft-ramachandra-bgp-ext-communities-08 - BGP Extended Communities Attribute. Srihari Ramachandra, Yakhov Rekhter.

11. Author's Address

Jeffrey Haas
NextHop Technologies
517 W. William St.
Ann Arbor, MI 48103-4943

Phone: (734) 973-2200

EMail: jhaas@nexthop.com