Network Working Group
Request for Comments: 2519
Category: Informational

E. Chen
Cisco
J. Stewart
Juniper

February 1999

A Framework for Inter-Domain Route Aggregation

Status of this Memo

Copyright Notice

Abstract

This document presents a framework for inter-domain route aggregation and shows an example router configuration which 'implements' this framework. This framework is flexible and scales well as it emphasizes the philosophy of aggregation by the source, both within routing domains as well as towards upstream providers, and it also strongly encourages the use of the 'no-export' BGP community to balance the provider-subscriber need for more granular routing information with the Internet's need for scalable inter-domain routing.

1. Introduction

The need for route aggregation has long been recognized. Route aggregation is good as it reduces the size, and slows the growth, of the Internet routing table. Thus, the amount of resources (e.g., CPU and memory) required to process routing information is reduced and route calculation is sped up. Another benefit of route aggregation is that route flaps are limited in number, frequency and scope, which saves resources and makes the global Internet routing system more stable.

Since CIDR (Classless Inter-Domain Routing) [2] was introduced, significant progress has been made on route aggregation, particularly in the following two areas:

- Formulation and implementation of IP address allocation policies by the top registries that conform to the CIDR principles [1].

This policy work is the cornerstone which makes efficient route aggregation technically possible.

- Route aggregation by large (especially "Tier 1") providers. To date, the largest reductions in the size of the routing table have resulted from efficient aggregation by large providers.

However, the ability of various levels of the global routing system to implement efficient aggregation schemes varies widely. As a result, the size and growth rate of the Internet routing table, as well as the associated route computation required, remain major issues today. To support Internet growth, it is important to maximize the efficiency of aggregation at all levels in the routing system.

Because of the current size of the routing system and its dynamic nature, the first step towards this goal is to establish a clearly defined framework in which scaleable inter-domain route aggregation can be realized. The framework described in this document is based on the predominant and current experience in the Internet. It emphasizes the philosophy of aggregation by the source, both within routing domains as well as towards upstream providers. The framework also strongly encourages the use of the "no-export" BGP community to balance the providersubscriber need for more granular routing information with the Internet's need for scalable inter-domain routing. The advantages of this framework include the following:

- Route aggregation is done in a distributed fashion, with emphasis on aggregation by the party or parties injecting the aggregatable routing information into the global mesh.

- The flexibility of a routing domain to be able to inject more granular routing information to an adjacent domain to control the resulting traffic patterns, without having an impact on the global routing system.

In addition to describing the philosophy, we illustrate it by presenting sample configurations. IPv4 prefixes, BGP4 and ASs are used in examples, though the principles are applicable to inter-domain route aggregation in general.

Address allocation policies and technologies to renumber entire networks, while very relevant to the realization of successful and sustained inter-domain routing, are not the focus of this document. The references section contains pointers to relevant documents [8, 9, 11, 12].

2. Route Aggregation Framework

The framework of inter-domain route aggregation we are proposing can be summarized as follows:

- Aggregation from the originating AS

  That is, in its outbound route announcements, each AS aggregates the BGP routes originated by itself, by dedicated AS and by private-ASs [10]. ("Routes originated by an AS" refers to routes which have that AS first in the AS path attribute. For example, routes statically configured and injected into BGP fall into this category.)

  This framework does not depend on "proxy aggregation" which refers to route aggregation done by an AS other than the originating AS. This preserves the capability of a multi-homed site to control the granularity of routing information injected into the global routing system. Since proxy aggregation involves coordination among multiple organizations, the complexity of doing proxy aggregation increases with the number of parties involved in the coordination. The complexity, in turn, impacts the practicality of proxy aggregation.

  An AS shall always originate via a stable mechanism (e.g., static route configuration) the BGP routes for the large aggregates from which it allocates addresses to customers. This ensures that it is safe for its customers to use BGP "no-export".

- Using BGP community "no-export" toward upstream providers

  That is, in its route announcements toward its upstream provider, an AS tags the BGP community "no-export" to routes it originates that do not need to be propagated beyond its upstream provider (e.g., prefixes allocated by the upstream provider).

This framework is illustrated in Figure 1. A "Tier 1" provider does not use "no-export" in its announcement as it does not have an upstream provider. However, it shall aggregate the routes it originates in its outbound announcements towards both peer providers and customers. An AS with an upstream provider shall aggregate the routes it originates and use "no-export" toward its upstream provider for routes that do not need to be propagated beyond its provider’s AS. This recursion shall apply to all levels of the routing hierarchy.

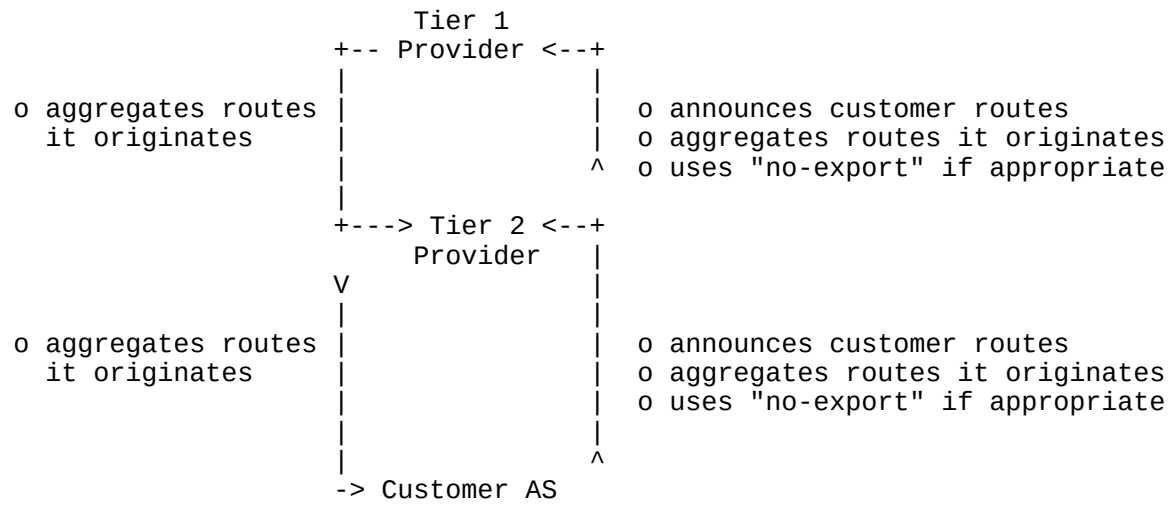

Figure 1

This framework scales well as aggregation is done at all levels of the routing system. It is flexible because the originating AS controls whether routes of finer granularity are injected to, and/or propagated by, its upstream provider. It facilitates multi-homing without compromising route aggregation.

This framework is detailed in the following sections.

3. Aggregation from the Originating AS

It has been well recognized that address allocation and address renumbering are keys to containing the growth of the Internet routing table [1, 2, 8, 9, 11, 12].

Although the strategies discussed in this document do not assume a perfect address allocation, it is strongly urged that an AS receive allocation from its upstream service providers' address block.

3.1 Intra-Domain Aggregation

To reduce the number of routes that need to be injected into an AS, there are a couple of principles that shall be followed:

- Carry in its BGP table the large route block allocated from its upstream provider or an address registry (e.g., InterNIC, RIPE, APNIC). This can be done by either static configuration of the large block or by aggregating more specific BGP routes. The former is recommended as it does not depend on other routes.

- Allocate sub-blocks to the access routers where further allocation is done. That is, the address allocation shall be done such that only a few, less specific routes (instead of many more, specific ones) need to be known to the other routers within the AS.

  For example, a prefix of /17 can be further allocated to different access routers as /20s which can then be allocated to customers connected to different interfaces on that router (as shown in Figure 2). Then in general only the /20 needs to be injected into the whole AS. Exceptions need to be made for multi-homed static routes.

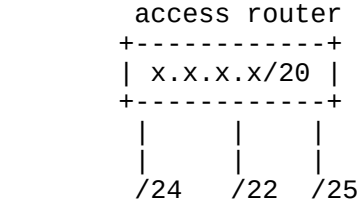

Figure 2

It is noted that rehoming of customers without renumbering even within the same AS may lead to injection of more specific routes. However, in general the more-specifics do not need to be advertised outside of that AS. Such routes can either be tagged with the BGP community "no-export" or filtered out by a prefix-based filter to prevent them from being advertised out.

3.2 Inter-Domain Aggregation

There are at least two types of routes that need to be advertised by an AS: routes originated by the AS and routes originated by its BGP customers. An AS may need to advertise full routes to certain BGP customers, in which case the routing announcements include routes originated by non-customer ASs. Clearly an AS can, and should, safely aggregate the routes originated by itself and by its BGP customers multi-homed only to it (using, e.g., the dedicated-AS and

by the private-AS mechanism [10]) in its outbound announcement. But it is far more dangerous to aggregate routes originated by customer ASs due to multi-homing.

However, there are several cases in which a route originated by a BGP customer (other than using the dedicated AS or private AS) does not need to be advertised out by its upstream providers. For example,

- The route is a more-specific of the upstream provider's block. However, the customer is either singly homed; or its connection to this particular upstream provider is used for backup only.

- The more-specifics of a larger block are announced by the customer in order to balance traffic over the multiple links to the upstream provider.

Our approach to suppress such routes is to give control to the ASs that originate the more-specifics (as seen by its upstream providers) and let them tag the BGP community "no-export" to the appropriate routes.

The BGP community "no-export" is a well known BGP community [6, 7]. A route with this attribute is not propagated beyond an AS boundary. So, if a route is tagged with this community in its announcement to an upstream provider and is accepted by the upstream provider, the route will not be announced beyond the upstream provider's AS. This achieves the goal of suppressing the more-specifics in the upstream provider's outbound announcement.

In this framework, the BGP community "no-export" shall be tagged to routes that are to be advertized to, but not propagated by, its upstream provider. They may include routes allocated out of its upstream provider's block or the more specific routes announced to its upstream provider for the purpose of load balancing. This aggregation strategy can be implemented via prefix-based filtering as shown in the example of Section 5.

For its own protection, a downstream AS shall announce only its own routes and its customer routes to its upstream providers. Thus, the outbound routing announcement and aggregation policy can be expressed as follows:

For routes originated by itself/dedicated-AS/private-AS:
tag with "no-export" when appropriate, and advertise the large block and suppress the more-specifics

For routes originated by customer ASs:
advertise to upstream ASs

For any other routes:
do not advertise to upstream ASs

This approach is flexible and scales well as it gives control to the party with the special needs, distributes the workload and avoids the coordination overhead required by proxy aggregation.

4. Aggregation by a Provider

A provider shall aggregate all the routes it originates, as documented in Section 3. The only difference is that the provider may be providing full routes to certain BGP customers where no outbound filtering is presently in place. Experience has shown that inconsistent route announcement (e.g., aggregate at the interconnects but not toward certain customers) can cause serious routing problems for the Internet as a whole because of longest-match routing. In certain cases announcing the more-specifics to customers might provide for more accurate IGP metrics and could be useful for better load-balancing. However, the potential risk seems to outweigh the benefit, especially given the increasing complexity of connectivity that a customer may have. As a result, every effort shall be made to ensure consistent route aggregation for all BGP peers. This means deploying filters for the BGP peers which receive full routes.

In summary, the aggregation strategy for a provider shall be:

- In announcing customer routes:

  For routes originated by itself/dedicated-AS/private-AS:
  tag with "no-export" when appropriate, and advertise the large block and suppress the more-specifics

  For routes originated by other customer ASs:
  advertise

  For any other routes:
  do not advertise

- In announcing full routes:

  For routes originated by itself/dedicated-AS/private-AS:
  tag with "no-export" when appropriate, and advertise the large block and suppress the more-specifics

  For any other routes:
  advertise

5. An Example

Consider the example shown in Figure 3 where AS 1000 is a "Tier 1" provider with two large aggregates 208.128.0.0/12 and 166.55.0.0/16, and AS 2000 is a customer of AS 1000 with a "portable address" 160.75.0.0/16 and an address 208.128.0.0/19 allocated from AS 1000. Assume that 208.128.0.0/19 does not need to be propagated beyond AS 1000.

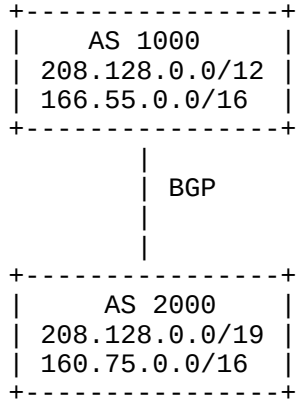

Figure 3

Then, based on the framework presented, AS 1000 would

- originate and advertise the BGP routes 208.128.0.0/12 and 166.55.0.0/16, and suppress more-specifics originated by itself/private-ASs/dedicated-ASs

- advertise the routes received from the customer AS 2000

and AS 2000 would

- originate BGP route 208.128.0.0/19 and 160.75.0.0/16

- advertise both 160.75.0.0/16 and 208.128.0.0/19 to its provider AS 1000 and suppress the more specifics originated by itself/private-AS/dedicated-AS, tagging the route 208.128.0.0/19 with "no-export"

- advertise both 160.75.0.0/16 and 208.128.0.0/19 to its BGP customers (if any) and suppress the more-specifics originated by itself/private-AS/dedicated-AS, plus any other routes the customers may desire to receive

The sample configuration which implement these policies (in Cisco syntax) is given in Appendix A.

6. Acknowledgments

The authors would like to thank Roy Alcala of MCI for a number of interesting hallway discussions related to this work. The IETF's IDR Working Group also provided many helpful comments and suggestions.

7. References

[1] Rekhter, Y. and T. Li, "An Architecture for IP Address Allocation with CIDR", RFC 1518, September 1993.

[2] Fuller, V., Li, T., Yu, J. and K. Varadhan, "Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy", RFC 1519, September 1993.

[3] Rekhter, Y., and T. Li, "A Border Gateway Protocol 4 (BGP-4)", RFC 1771, March 1995.

[4] Rekhter, Y. and P., Gross, "Application of the Border Gateway Protocol in the Internet", RFC 1772, March 1995.

[5] Rekhter, Y., "Routing in a Multi-provider Internet", RFC 1787, April 1995.

[6] Chandra, R., Traina, P. and T. Li, "BGP Communities Attribute", RFC 1997, August 1996.

[7] Chen, E. and T. Bates, "An Application of the BGP Community Attribute in Multi-home Routing", RFC 1998, August 1996.

[8] Ferguson, P. and H. Berkowitz, "Network Renumbering Overview: Why would I want it and what is it anyway?", RFC 2071, January 1997.

[9] Berkowitz, H., "Router Renumbering Guide", RFC 2072, January 1997.

[10] Stewart, J., Bates, T., Chandra, R., and Chen, E., "Using a Dedicated AS for Sites Homed to a Single Provider", RFC 2270, January 1998.

[11] Carpenter, B., Crowcroft, J. and Y. Rekhter, "IPv4 Address Behaviour Today", RFC 2101, February 1997.

[12] Carpenter, B. and Y. Rekhter, "Renumbering Needs Work", RFC 1900, February 1996.

[13] Cisco systems, Cisco IOS Software Version 10.3 Router Products Configuration Guide (Addendum), May 1995.

8. Authors' Addresses

Enke Chen
Cisco Systems
170 West Tasman Drive
San Jose, CA 95134-1706

Phone: +1 408 527 4652
EMail: enkechen@cisco.com

John W. Stewart, III
Juniper Networks, Inc.
385 Ravendale Drive
Mountain View, CA 94043

Phone: +1 650 526 8000
EMail: jstewart@juniper.net

A. Appendix A: Example Cisco Configuration

This appendix lists the Cisco configurations for AS 2000 of the examples presented in Section 5. The configuration here uses the AS-path for outbound filtering although it can also be based on BGP community. Several route-maps are defined that can be used for peering with the upstream provider, and for peering with customers (announcing full routes or customer routes).

```
!!# inject aggregates
ip route 160.75.0.0 255.255.0.0 Null0 254
ip route 208.128.0.0 255.255.224.0 Null0 254
!
router bgp 2000
network 160.75.0.0 mask 255.255.0.0
network 208.128.0.0 mask 255.255.224.0
neighbor x.x.x.x remote-as 1000
neighbor x.x.x.x route-map export-routes-to-provider out
neighbor x.x.x.x send-community
!
!!# match all
ip as-path access-list 1 permit .*
!
!!# List of internal AS and private ASs that are safe to aggregate
ip as-path access-list 10 permit ^$
ip as-path access-list 10 permit ^64999_
ip as-path access-list 10 deny .*
!
!!# list of other customer ASs
ip as-path access-list 20 permit ^3000_

!!# List of prefixes to be tagged with "no-export"
access-list 101 permit ip 208.128.0.0 0.0.0.0 255.255.224.0 0.0.0.0
!!# Filter out the more specifics of large aggregates, and permit the rest
access-list 102 permit ip 160.75.0.0 0.0.0.0 255.255.0.0 0.0.0.0
access-list 102 deny ip 160.75.0.0 0.0.255.255 255.255.128.0 0.0.127.255
access-list 102 permit ip 208.128.0.0 0.0.0.0 255.255.224.0 0.0.0.0
access-list 102 deny ip 208.128.0.0 0.0.31.255 255.255.240.0 0.0.16.255
access-list 102 permit ip any any
!

!!# route-map with the upstream provider
route-map export-routes-to-provider permit 10
match ip address 101
set community no-export
route-map export-routes-to-provider permit 20
match as-path 10
match ip address 102
```

```
route-map export-routes-to-provider permit 30
match as-path 20
!
!!# route-map with BGP customers that desire only customer routes
route-map export-customer-routes permit 10
match as-path 10
match ip address 102
route-map export-customer-routes permit 20
match as-path 20
!
!!# route-map with BGP customers that desire full routes
route-map export-full-routes permit 10
match as-path 10
match ip address 102
route-map export-full-routes permit 20
match as-path 1
!
```

Full Copyright Statement