

GROW Working Group
Internet-Draft
Intended status: Informational
Expires: February 17, 2013

R. Raszuk
NTT MCL
J. Heitz
Ericsson
A. Lo
Arista
L. Zhang
UCLA
X. Xu
Huawei
August 16, 2012

Simple Virtual Aggregation (S-VA)
draft-ietf-grow-simple-va-12.txt

Abstract

All BGP routers in the Default Free Zone (DFZ) are required to carry all the routes in the Default Free Routing Table (DFRT). A technique is described that allows some BGP routers not to install all of those routes into the Forwarding Information Base (FIB).

Some routers in an Autonomous System (AS) announce an aggregate (the VA prefix) in addition to the routes they already announce. This enables other routers not to install the routes covered by the VA prefix into the FIB as long as those routes have the same next-hop as the VA prefix.

The VA prefixes that are announced within an AS are not announced to any other AS. Described functionality is of very low operational complexity by proposing a confined BGP speaker solution without any dependency on network wide configuration or requirement for any form of intra-domain tunneling.

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

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

Internet-Draft

S-VA

August 2012

material or to cite them other than as "work in progress."

This Internet-Draft will expire on February 17, 2013.

Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Internet-Draft

S-VA

August 2012

Table of Contents

- [1. Introduction](#) [4](#)
- [1.1. Scope of this Document](#) [4](#)
- [1.2. Requirements notation](#) [4](#)
- [1.3. Terminology](#) [4](#)
- [2. Operation of S-VA](#) [5](#)
- [3. Deployment considerations](#) [7](#)
- [4. IANA Considerations](#) [7](#)
- [5. Security Considerations](#) [8](#)
- [6. Acknowledgements](#) [8](#)
- [7. Normative References](#) [8](#)
- [Authors' Addresses](#) [8](#)

1. Introduction

A technique called Simple Virtual Aggregation (S-VA) is described. It allows some routers not to have to store some routes in the Forwarding Information Base (FIB) while still advertising and receiving the full Default Free Routing Table (DFRT) in BGP.

A typical scenario is as follows. Core routers in the ISP maintain the full DFRT in the FIB and RIB. Edge routers maintain the full DFRT in the BGP Loc-RIB, but do not install certain routes in the RIB and FIB. Edge routers may install a default route to core routers, to Area Border Routers (ABR) which are installed on the Point of Presence (POP), to core boundary routers or to Autonomous System Border Routers (ASBR).

S-VA must be enabled on an edge router that needs to save its RIB and FIB space. The core routers must announce a new prefix called virtual aggregate (VA prefix).

1.1. Scope of this Document

The VA prefix is not intended to be announced from one AS into another, only between routers of the same AS.

S-VA can be used for IPv4 and IPv6 both unicast and multicast address families.

S-VA does not need to operate on every router in an AS.

1.2. Requirements notation

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [[RFC2119](#)].

1.3. Terminology

RIB/FIB-Installing Router (FIR): A router that does not suppress any routes and announces the VA prefix. Typically a core router, a POP to core boundary router or an ASBR would be configured as an FIR.

RIB/FIB-Suppressing Router (FSR): An S-VA router that installs the VA prefix, and does not install into its FIB routes that are covered by and have the same next-hop as the VA prefix. Typically an edge router would be configured as an FSR.

Suppress: Not to install a route that is covered by the VA prefix into the global RIB or FIB.

Legacy Router: A router that does not run S-VA, and has no knowledge of S-VA.

Global Routing Information Base (RIB): All the routing protocols in a router install their selected routes into the RIB. The routes in the RIB are used to resolve next-hops for other routes, to be redistributed to other routing protocols and to be installed into the FIB.

Local/Protocol Routing Information Base (Loc-RIB): The Loc-RIB contains the routes that have been selected by the local BGP speaker's Decision Process as in [[RFC4271](#)].

NLRI: Network Layer Reachability Information [[RFC4271](#)]

2. Operation of S-VA

There are three types of routers in S-VA, FIB-Installing routers (FIR), FIB-Suppressing routers (FSR) and optionally, legacy routers. While any router can be an FIR or an FSR, the simplest form of deployment is for AS border routers to be configured as FIRs and for customer facing edge routers to be configured as FSRs.

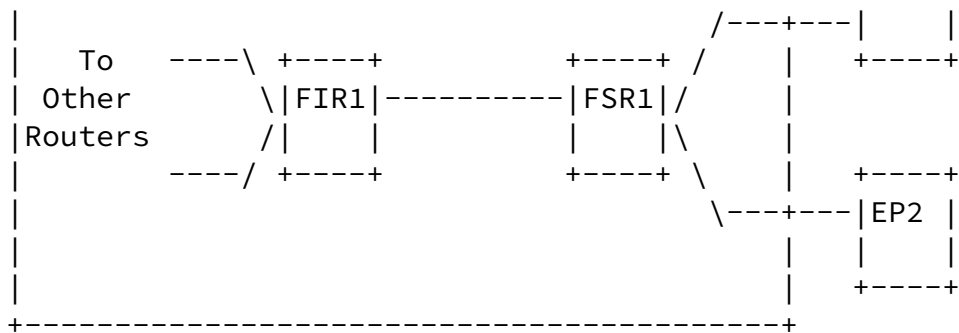


Figure 1

Suppose that FSR1 has been enabled to perform S-VA. Originally it receives all routes from FIR1 (doing next-hop-self) as well as from EP1 and EP2. FIR1 now will advertise a VA prefix 0/0 with next-hop set to itself. That will cause FSR1 to suppress all routes with the same next-hop as the VA prefix. However, FSR1 will not suppress any routes received from EP1 and EP2, because their next-hops are different from that of the VA prefix.

Several FIRs may announce different S-VA prefixes. For example, in a POP, each edge router can announce into the POP an S-VA prefix that covers the addresses of the customers it services.

Several FIRs may announce the same S-VA prefix. In this case an FSR must choose to install only one of them. For example, two redundant ASBRs, both of which announce the complete DFRT may each also announce the default route as an S-VA prefix into the AS.

S-VA may be used to split traffic among redundant exit routers. For example, referring to Figure 1, suppose EP1 and EP2 are two redundant ASBRs that announce the complete DFRT. Each may also announce two S-VA prefixes into the AS: 0/1 and 128/1. EP1 might announce 0/1 with higher preference and EP2 might announce 128/1 with higher

preference. FIR1 will now install into its FIB 0/1 pointing to EP1 and 128/1 pointing to EP2. If either one of EP1 or EP2 were to fail, then FSR1 would switch the traffic to the other exit router with a single FIB installation of one S-VA prefix.

3. Deployment considerations

BGP routes may be used to resolve next-hops for static routes or other BGP routes. Because the default route does not imply reachability of any destination, a router can be configured not to resolve next-hops using the default route. In this case, S-VA should not suppress from installation into the RIB a route that may be used to resolve a next-hop for another route. It may still suppress it from installation into the FIB.

Selected BGP routes in the RIB may be redistributed to other protocols. If they no longer exist in the RIB, they will not be redistributed. This is especially important when the conditional redistribution is taking place based on the length of the prefix, community value etc. In those cases where redistribution policy is in place S-VA implementation should refrain from suppressing installation into the RIB routes matching such policy. It may still suppress them from installation into the FIB.

A router may originate a network route or an aggregate route into BGP. Some addresses covered by such a route may not exist. If this router were to receive a packet for an unreachable address within an originated route, it must not send that packet to the VA prefix route. There are several ways to achieve this. One is to have the FIR aggregate the routes instead of the FSR. Another is to install a blackhole route for the nonexistent addresses on the originating router. This issue is not specific to S-VA, but applicable to the general use of default routes.

Like any aggregate, an S-VA prefix may include more address space than the sum of the prefixes it covers. As such, the S-VA prefix may provide a route for a packet for which no real destination exists. An FSR will forward such a packet to the FIR.

If an S-VA prefix changes its next-hop or is removed, then many routes may need to be downloaded into the FIB to achieve convergence.

[4.](#) IANA Considerations

There are no IANA considerations.

[5.](#) Security Considerations

The authors are not aware of any new security considerations due to S-VA. The local nature of the proposed optimization eliminates any external exposure of the functionality. The presence of more specifics which are used as VA prefixes is also a normal BGP behaviour in current networks.

6. Acknowledgements

The concept for Virtual Aggregation comes from Paul Francis. In this document authors only simplified some aspects of its behavior to allow simpler adoption by some operators.

Authors would like to thank Clarence Filsfils, Nick Hilliard, S. Moonesamy and Tom Petch for their review and valuable input.

7. Normative References

- [RFC1997] Chandrasekeran, R., Traina, P., and T. Li, "BGP Communities Attribute", [RFC 1997](#), August 1996.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC5082] Gill, V., Heasley, J., Meyer, D., Savola, P., and C. Pignataro, "The Generalized TTL Security Mechanism (GTSM)", [RFC 5082](#), October 2007.

Authors' Addresses

Robert Raszuk
NTT MCL
101 S Ellsworth Avenue Suite 350
San Mateo, CA 94401
US

Email: robert@raszuk.net

Jakob Heitz
Ericsson
300 Holger Way
San Jose, CA 95135
USA

Phone:
Email: jakob.heizt@ericsson.com

Alton Lo
Arista Networks
5470 Great America Parkway
Santa Clara, CA 95054
USA

Phone:
Email: altonlo@aristanetworks.com

Lixia Zhang
UCLA
3713 Boelter Hall
Los Angeles, CA 90095
US

Phone:
Email: lixia@cs.ucla.edu

Xiaohu Xu
Huawei Technologies
No.3 Xixi Rd., Shang-Di Information Industry Base, Hai-Dian District
Beijing, Beijing 100085
P.R.China

Phone: +86 10 82836073
Email: xuxh@huawei.com

