Network Working Group
Internet-Draft
Intended status: Informational
Expires: February 4, 2017

X. Xu
Huawei
C. Jacquenet
Orange
T. Boyes
Bloomberg LP
B. Fee
Extreme Networks
W. Henderickx
Alcatel-Lucent
August 3, 2016

FIB Reduction in Virtual Subnet
draft-ietf-bess-virtual-subnet-fib-reduction-03

Abstract

Virtual Subnet is a BGP/MPLS IP VPN-based subnet extension solution which is intended for building Layer3 network virtualization overlays within and/or between data centers. This document describes a mechanism for reducing the FIB size of PE routers in the Virtual Subnet context.

Table of Contents

1. Introduction

Virtual Subnet [RFC7814] is a BGP/MPLS IP VPN [RFC4364] -based subnet extension solution which is intended for building Layer3 network virtualization overlays within and/or across data centers. In the Virtual Subnet context, since CE host routes of a given VPN instance need to be exchanged among PE routers participating in that VPN instance, the resulting forwarding table (a.k.a. FIB) size of PE routers may become a big concern in large-scale data center environment where they may need to install a huge amount of host routes into their forwarding tables. In some cases where host routes need to be maintained on the control plane, it needs a method to reduce the FIB size of PE routers without any change to the RIB and the routing table. Therefore, this document proposes a very simple mechanism for reducing the FIB size of PE routers. The basic idea of this mechanism is: Those host routes learnt from remote PE routers are selectively installed into the FIB while the remaining routes including local CE host routes are installed into the FIB by default as before.

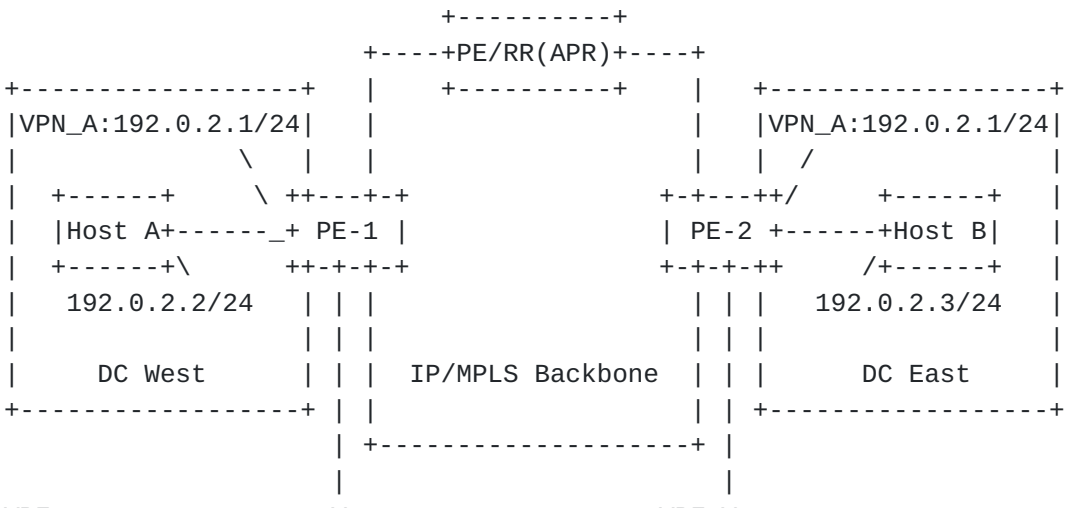
1.1. Requirements Language

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 RFC 2119 [RFC2119].

2. Terminology

This memo makes use of the terms defined in [RFC4364].

3. Solution Description



Prefix	Nexthop	Protocol	In_FIB
192.0.2.1/32	127.0.0.1	Direct	Yes
192.0.2.2/32	192.0.2.2	Direct	Yes
192.0.2.3/32	PE-2	IBGP	No
192.0.2.0/25	APR	IBGP	Yes
192.0.2.128/25	APR	IBGP	Yes
192.0.2.0/24	192.0.2.1	Direct	Yes

Prefix	Nexthop	Protocol	In_FIB
192.0.2.1/32	127.0.0.1	Direct	Yes
192.0.2.2/32	PE-1	IBGP	No
192.0.2.3/32	192.0.2.3	Direct	Yes
192.0.2.0/25	APR	IBGP	Yes
192.0.2.128/25	APR	IBGP	Yes
192.0.2.0/24	192.0.2.1	Direct	Yes

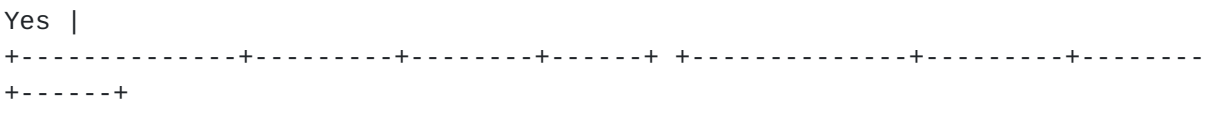


Figure 1: Selective IPv4 FIB Installation Example

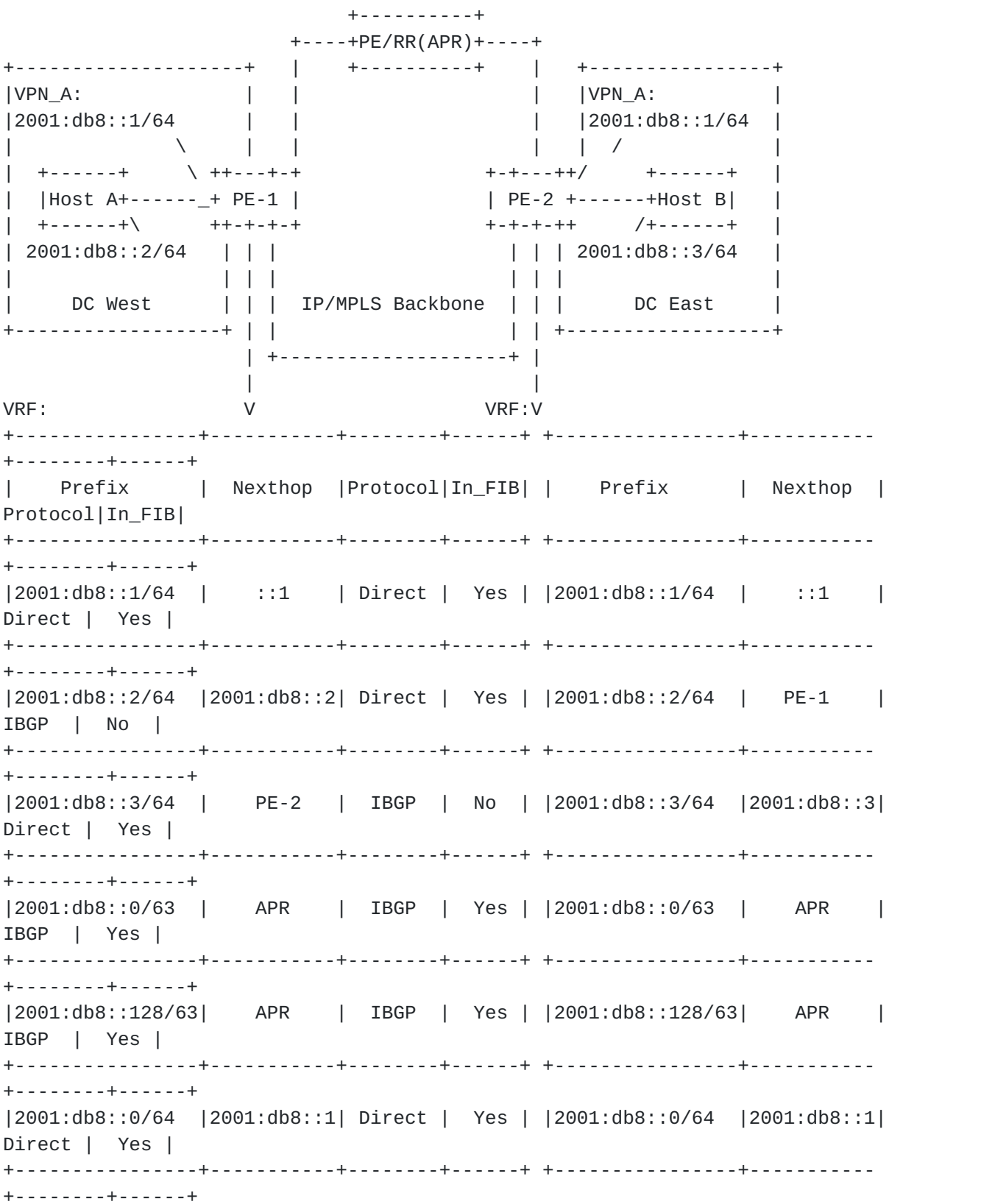


Figure 2: Selective IPv6 FIB Installation Example

To reduce the FIB size of PE routers, the selective FIB installation concept as described in [I-D.ietf-grow-va] can be leveraged in the Virtual Subnet context. Take the VPN instance demonstrated in Figure 1 or Figure 2 as an example, the FIB reduction procedures are described as follows:

1. Multiple more specific prefixes (e.g., 192.0.2.0/25 and 192.0.2.128/25 in IPv4 example, or 2001:db8::0/63 and 2001:db8::128/63 in IPv6 example) corresponding to an extended subnet (i.e., 192.0.2.0/24 in IPv4 example, or 2001:db8::0/64 in IPv6 example) are specified as Virtual Prefixes (VPs). Meanwhile, one or more PE routers (or route reflectors) are configured as Aggregation Point Routers (APR) for each VP. The

APRs for a given VP would install a null route to that VP while propagating a route to that VP via the L3VPN signaling.

2. For a given host route in the routing table which is learnt from any remote PE router, PE routers which are non-APRs for any VP covering this host route would not install it into the FIB by default. In contrast, PE routers (or route reflectors) which are APRs for any VP covering that host route would install it into the FIB. If one or more particular remote host routes need to be installed by non-APR PE routers by default as well for whatever reasons, the best way to realize such goal is to attach a special extended communities attribute to those particular host routes either by originating PE routers or by route reflectors. Upon receiving any host routes attached with the above extended communities attribute, non-APR PE routers SHOULD install them by default.

3. Upon receiving a packet destined for a given remote CE host, if no host route for that CE host is found in the FIB, the ingress PE router would forward the packet to a given APR according to the longest-matching VP route, which in turn forwards the packet to the final egress PE router. In this way, the FIB size of those non-APR PE routers can be greatly reduced at the potential cost of path stretch.

In order to forward packets destined for remote CE hosts directly to the final egress PE routers without the potential path stretch penalty, non-APR PE routers could perform on-demand FIB installation for remote host routes which are available in the routing table. For example, upon receiving an ARP request or Neighbor Solicitation (NS) message from a local CE host, the non-APR PE router would perform a lookup in the routing table. If a corresponding host route for the target host is found but not yet installed into the FIB, it would be installed into the FIB. Another possible way to trigger on-demand FIB installation is as follows: when receiving a packet whose longest-matching FIB entry is a particular VP route learnt from any APR, a copy of this packet would be sent to the control plane while this original packet is forwarded as normal. The above copy sent to the control plane would trigger a lookup in the routing table. If a corresponding host route is found but not yet installed into the FIB, it would be installed into the FIB. To provide robust protection against DoS attacks on the control plane, rate-limiting of the above packets sent to the control plane MUST be enabled. Those FIB entries for remote CE host routes which are on-demand installed on non-APR PE routers would expire if not used for a certain period of time.

4. Acknowledgements

The authors would like to thank Susan Hares, Yongbing Fan, Robert Raszuk, Bruno Decraene and Fred Baker for their valuable suggestions on this document.

5. IANA Considerations

The type value for the Extended Communities Attributes as described in this doc is required to be allocated by the IANA.

6. Security Considerations

Those security considerations as described in [RFC7814] are applicable to this document. This document does not introduce any new security risk.

7. References

7.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, DOI 10.17487/RFC4364, February 2006, <http://www.rfc-editor.org/info/rfc4364>.

[RFC7814] Xu, X., Jacquenet, C., Raszuk, R., Boyes, T., and B. Fee, "Virtual Subnet: A BGP/MPLS IP VPN-Based Subnet Extension Solution", RFC 7814, DOI 10.17487/RFC7814, March 2016, <http://www.rfc-editor.org/info/rfc7814>.

7.2. Informative References

[I-D.ietf-grow-va] Francis, P., Xu, X., Ballani, H., Jen, D., Raszuk, R., and L. Zhang, "FIB Suppression with Virtual Aggregation", draft-ietf-grow-va-06 (work in progress), December 2011.

Authors' Addresses

Xiaohu Xu
Huawei

Email: xuxiaohu@huawei.com

Christian Jacquenet
Orange

Email: christian.jacquenet@orange.com

Truman Boyes
Bloomberg LP

Email: tboyes@bloomberg.net

Brendan Fee
Extreme Networks

Email: bfee@enterasys.com

Wim Henderickx
Alcatel-Lucent

Email: wim.henderickx@alcatel-lucent.com