Network working group
Internet Draft
Category: Informational

X. Xu
Huawei

S. Hares

Y. Fan
China Telecom

C. Jacquenet
Orange

T. Boyes
Bloomberg LP

B Fee
Extreme Networks

Expires: August 2014 February 14, 2014

FIB Reduction in Virtual Subnet

draft-xu-l3vpn-virtual-subnet-fib-reduction-00

Abstract

Virtual Subnet is a L3VPN-based subnet extension solution which can be used to build Layer3 network virtualization overlays within and/or across data centers. This document describes a mechanism for reducing the FIB size of PE routers in the Virtual Subnet context.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

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."

This Internet-Draft will expire on August 14, 2014.

Copyright Notice

Conventions used in this document

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].

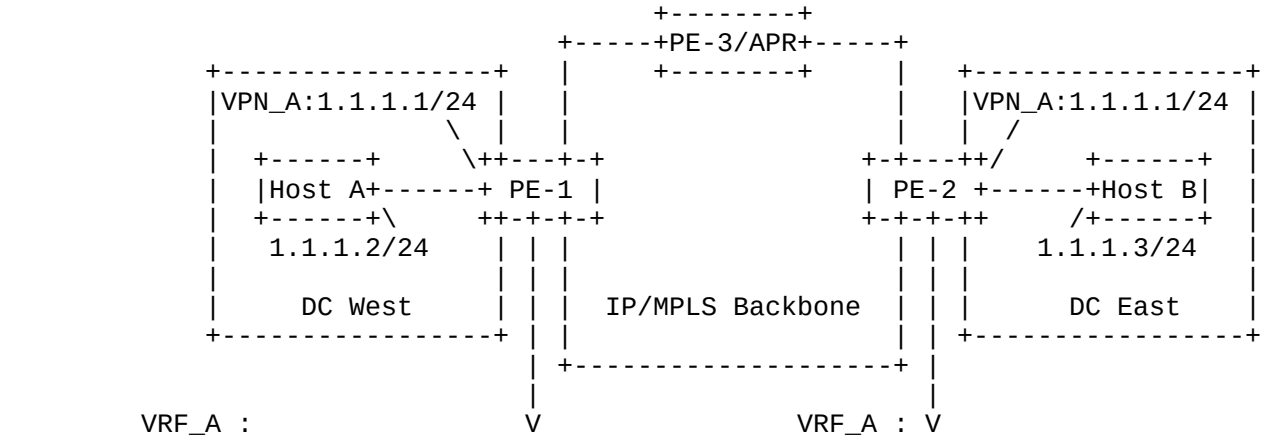
Table of Contents

1. Introduction

Virtual Subnet [VS] is a L3VPN-based subnet extension solution which can be used to build Layer3 network virtualization overlays within and/or across data centers. Since host routes of a given VPN instance are usually exchanged among PE routers participating in that VPN instance in the context of Virtual Subnet, the forwarding table (a.k.a. FIB) size of PE routers may become a scaling concern once they need to install a huge amount of host routes into their forwarding tables, especially in the particular cloud data center interconnect scenario where millions of host routes are there.

To address the above FIB scaling concern, 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 installed into the FIB on demand, while the remaining routes including local host routes are installed into the FIB by default as before.

2. Solution Overview



VRF_A :

Prefix	Nexthop	Protocol	In_FIB
1.1.1.1/32	127.0.0.1	Direct	Yes
1.1.1.2/32	1.1.1.2	Direct	Yes
1.1.1.3/32	PE-2	IBGP	No
1.1.1.0/25	APR	IBGP	Yes
1.1.1.128/25	APR	IBGP	Yes
1.1.1.0/24	1.1.1.1	Direct	Yes

VRF_A :

Prefix	Nexthop	Protocol	In_FIB
1.1.1.1/32	127.0.0.1	Direct	Yes
1.1.1.2/32	PE-1	IBGP	No
1.1.1.3/32	1.1.1.3	Direct	Yes
1.1.1.0/25	APR	IBGP	Yes
1.1.1.128/25	APR	IBGP	Yes
1.1.1.0/24	1.1.1.1	Direct	Yes

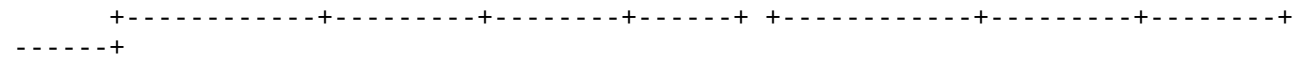


Figure 1: FIB Reduction Example

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

1) Multiple more specific prefixes (e.g., 1.1.1.0/25 and 1.1.1.128/25) corresponding to an extended subnet (i.e., 1.1.1.0/24) are specified as Virtual Prefixes (VPs). Meanwhile, one or more PE routers 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 which are APRs for any VP covering that host route would install it into the FIB.

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.

3. Security Considerations

This document doesn't introduce additional security risk to BGP/MPLS IP VPN, nor does it provide any additional security feature for BGP/MPLS IP VPN.

4. IANA Considerations

There is no requirement for any IANA action.

5. Acknowledgements

Thanks Robert Raszuk for his review and suggestions on this document.

6. References

6.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[VS] Xu, X et al., "Virtual Subnet: A L3VPN-based Subnet Extension Solution", draft-xu-l3vpn-virtual-subnet-03 (Work in Progress), January 2014.

6.2. Informative References

[RFC4364] Rosen. E and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, February 2006.

[VA] Francis, P., Xu, X., Ballani, H., Jen, D., Raszuk, R., and L. Zhang, "Virtual Aggregation", draft-ietf-grow-va-05.txt, (Work in Progress), December 2011.

Authors' Addresses

Xiaohu Xu
Huawei Technologies,
Beijing, China.
Phone: +86 10 60610041
Email: xuxiaohu@huawei.com

Susan Hares
Email: shares@ndzh.com

Yongbing Fan
Guangzhou Institute, China Telecom
Guangzhou, China.
Phone: +86 20 38639121
Email: fanyb@gsta.com

Christian Jacquenet
Orange
Rennes France
Email: christian.jacquenet@orange.com

Truman Boyes
Bloomberg LP
Phone: +1 2126174826
Email: tboyes@bloomberg.net

Brendan Fee
Extreme Networks
9 Northeastern Blvd.
Salem, NH, 03079
Email: bfee@enterasys.com