FIB Reduction in Virtual Subnet

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

Xiaohu Xu (Huawei)

Susan Hares

Yongbing Fan (China Telecom)

Christian Jacquenet (France Telecom)

Truman Boyes (Bloomberg)

Brendan Fee (Extreme Networks)

IETF90, Toronto

Motivation

- Virtual Subnet [draft-ietf-l3vpn-virtual-subnet] is intended for building L3 network virtualization overlays within and/or across data centers.
  - Since a subnet is extended across multiple PE routers, CE host routes need to be exchanged among PE routers.
- The resulting FIB size of PE routers may become a major concern in large-scale data center environments.
- In some cases where host routes may 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.
- This draft describes a simple mechanism for reducing the FIB size of PE routers.
  - Remote CE host routes are selectively installed into the FIB.
  - The remaining routes including local CE host routes are installed into the FIB as before.

Steps to Reduce FIBs in Virtual Subnet

- Multiple more specific prefixes (e.g., 1.1.1.0/25 and 1.1.1.128/25) corresponding to the extended subnet (i.e., 1.1.1.0/24) are specified as **Virtual Prefixes (VPs).**

- More than one PE routers or RRs are configured as **Aggregation Point Routers (APR)** for each VP.

- **The APRs** for a given VP would originate a **null route** to that VP and then advertise it via L3VPN signaling.

- Remote CE host routes covered by a given VP would not be installed on non-APRs for that VP by default. Instead, they should be installed on APRs for that VP by default.
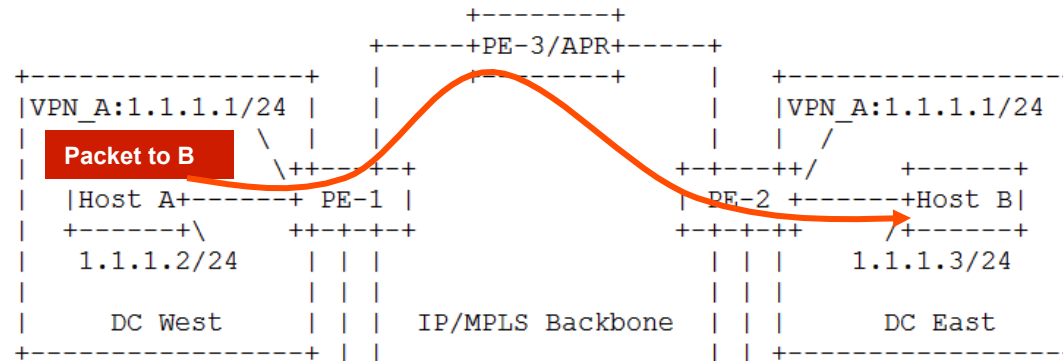
FIB Reduction in Virtual Subnet Context (1)

APR installs all received CE host routes into the FIB while originating two VPs (e.g., 1.1.1.0/25 and 1.1.1.128/25) corresponding to the extended subnet (e.g., 1.1.1.0/24).
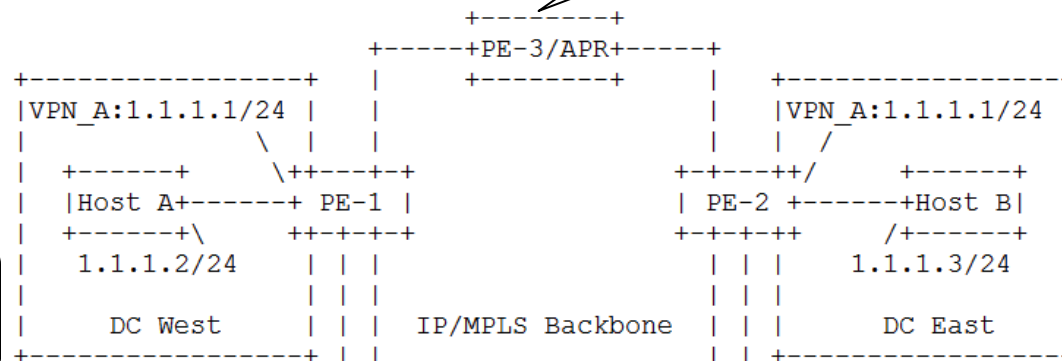
Non-APR would not install remote host routes into the FIB by default.

```
                              +--------+
                      +-----+PE-3/APR+-----+
  +----------------+  |      +--------+    |  +----------------+
  |VPN_A:1.1.1.1/24 | |                    |  |VPN_A:1.1.1.1/24 |
  |              \  | |                    |  |  |  /           |
  |  +------+    \++---+-+                 +-+---++/   +------+  |
  |  |Host A+------+ PE-1 |                | PE-2 +------+Host B|  |
  |  +------+\    ++-+-+-+                 +-+-+-++   /+------+  |
  |  1.1.1.2/24   | | |                    | | |  1.1.1.3/24    |
  |              | | |                    | | |                |
  |  DC West     | | |  IP/MPLS Backbone  | | |  DC East       |
  +----------------+ | |                  | | +----------------+
                   | +------------------+ |
                   |                      |
  VRF_A :          V         VRF_A : V
```

VRF_A :		VRF_A : V	
Prefix	Nexthop	Protocol	In_FIB	Prefix	Nexthop	Protocol	In_FIB
1.1.1.1/32	127.0.0.1	Direct	Yes	1.1.1.1/32	127.0.0.1	Direct	Yes
1.1.1.2/32	1.1.1.2	Direct	Yes	1.1.1.2/32	PE-1	IBGP	No
1.1.1.3/32	PE-2	IBGP	No	1.1.1.3/32	1.1.1.3	Direct	Yes
1.1.1.0/25	APR	IBGP	Yes	1.1.1.0/25	APR	IBGP	Yes
1.1.1.128/25	APR	IBGP	Yes	1.1.1.128/25	APR	IBGP	Yes
1.1.1.0/24	1.1.1.1	Direct	Yes	1.1.1.0/24	1.1.1.1	Direct	Yes

FIB Reduction in Virtual Subnet Context (2)

```
                                +--------+
                      +-----+PE-3/APR+-----+
    +----------------+    |   +------+    |   +----------------+
    |VPN_A:1.1.1.1/24 |   |   |        |   |   |VPN_A:1.1.1.1/24 |
    |   ┌──────────┐  \   |   |        |   |   |    /            |
    |   │Packet to B│   \++---+-+     +-+---++/  +------+       |
    |   |Host A+------+ PE-1 |         | PE-2 +------+Host B|   |
    |   +------+\      ++-+-+-+         +-+-+-++   /+------+   |
    |   1.1.1.2/24   | | |               | | |   1.1.1.3/24    |
    |               | | |               | | |                 |
    |   DC West     | | |  IP/MPLS Backbone | | |   DC East    |
    +----------------+ | |               | | +----------------+
                     | +-------------------+ |
                     |                       |
    VRF_A :          V                 VRF_A : V
```

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 : V

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

On-demand FIB Installation of Remote CE Host Route

- To avoid any potential path stretch penalty, non-APR PE routers could perform on-demand FIB installation of remote CE host routes.

  - Upon receiving an ARP request from a local CE host, the non-APR PE router would perform a lookup in the routing table. If the corresponding host route for the target host is found but not yet installed into the FIB, it would be installed into the FIB accordingly.

  - Alternatively, 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 to trigger the possible FIB installation.

    - 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.
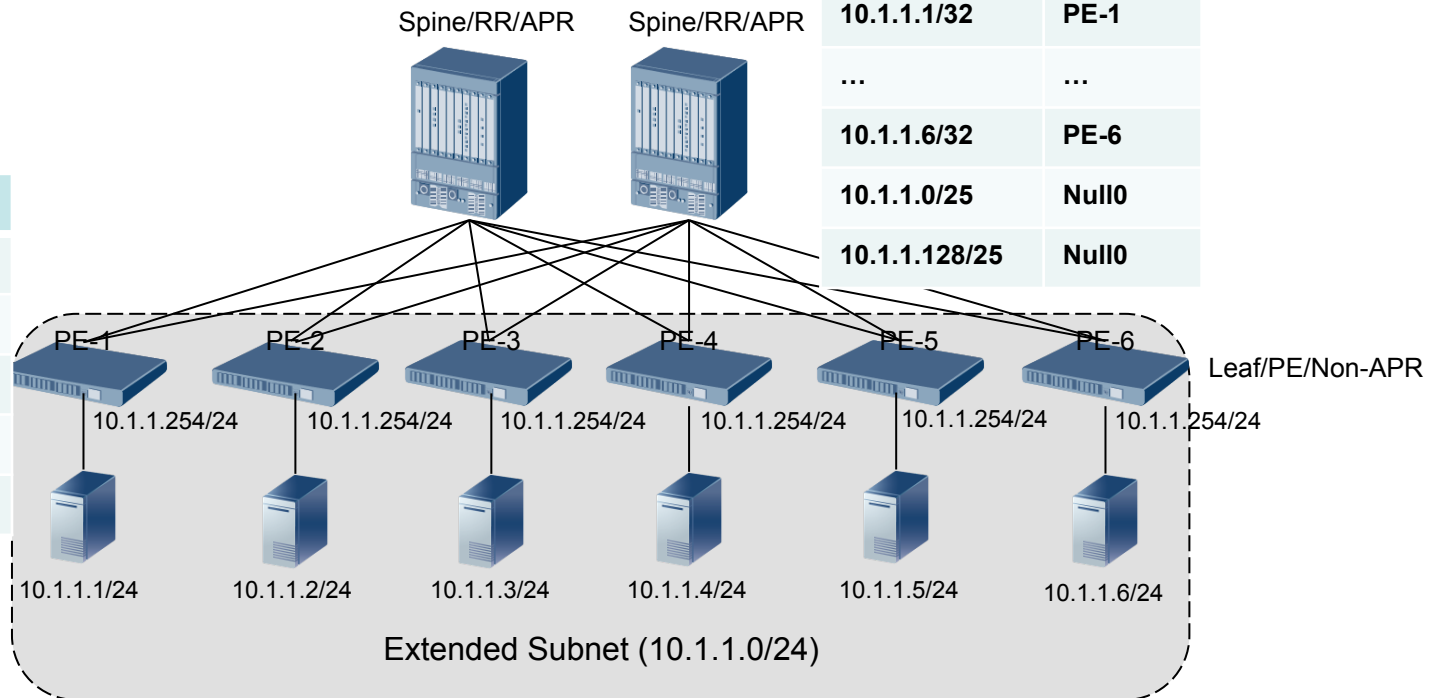
FIB Reduction in Spine-Leaf Topology

- In the spine-leaf topology, there is no need for the on-demand FIB installation of remote CE host routes since those packets destined for remote CE hosts would have to traverse one of the spine nodes anyway.
  - PE routers just need to install local routes and those VP routes learnt from APRs into the FIB.

FIB on RR:

Prefix	Nexthop
10.1.1.1/32	**PE-1**
…	…
10.1.1.6/32	**PE-6**
10.1.1.0/25	**Null0**
10.1.1.128/25	**Null0**

FIB on PE-1:

Prefix	Nexthop
10.1.1.1/32	**10.1.1.1**
10.1.1.254/32	**127.0.0.1**
10.1.1.0/24	**10.1.1.254**
10.1.1.0/25	**APR**
10.1.1.128/25	**APR**

Spine/RR/APR   Spine/RR/APR

PE-1   PE-2   PE-3   PE-4   PE-5   PE-6   Leaf/PE/Non-APR

10.1.1.254/24   10.1.1.254/24   10.1.1.254/24   10.1.1.254/24   10.1.1.254/24   10.1.1.254/24

10.1.1.1/24   10.1.1.2/24   10.1.1.3/24   10.1.1.4/24   10.1.1.5/24   10.1.1.6/24

Extended Subnet (10.1.1.0/24)

Next Steps

- WG adoption as an informational draft?