SRv6 Service SID Flag Extension for Multi-homed SRv6 BGP Services

draft-liu-bess-multihome-srv6-service-sid-flag-00

Yisong Liu (China Mobile)

Presenter: Changwang Lin (New H3C Technologies)

Mengxiao Chen (New H3C Technologies)

IETF-116, March 2023

Introduction

In the multi-homing scenarios

- Ingress PE can send traffic to the egress PE in load sharing mode or active/standby mode;
- Multiple egress PEs can also implement local protection when the CE side link fails.

So to meet different requirements, the egress PE may need to allocate and advertise multiple service SIDs for the same service

This draft describes the use cases for two types of service SID, **anycast service SID** and **bypass service SID**, and propose to define the flags when advertising these two type of service SIDs through BGP message
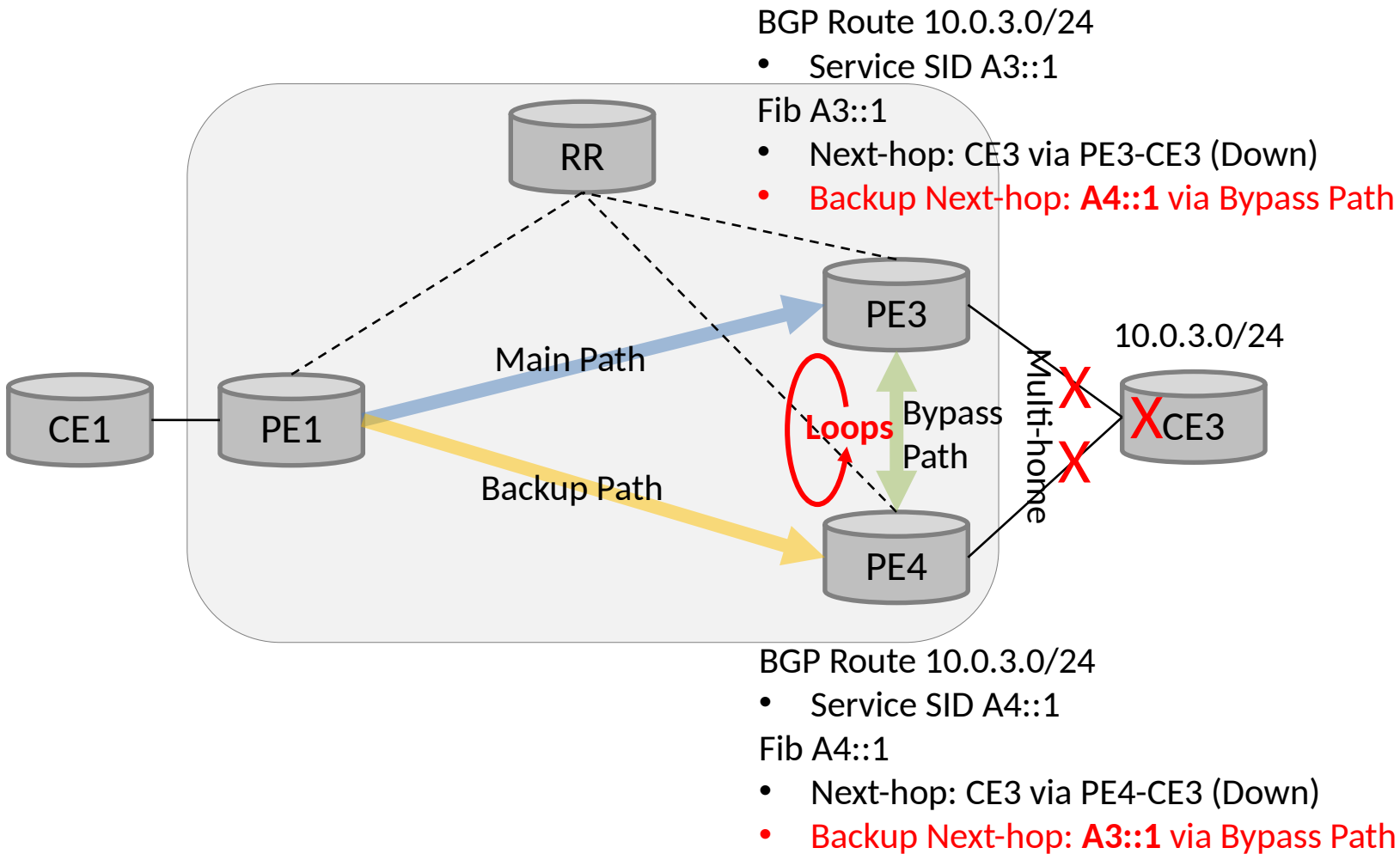
Use Case: Anycast SRv6 Service SID
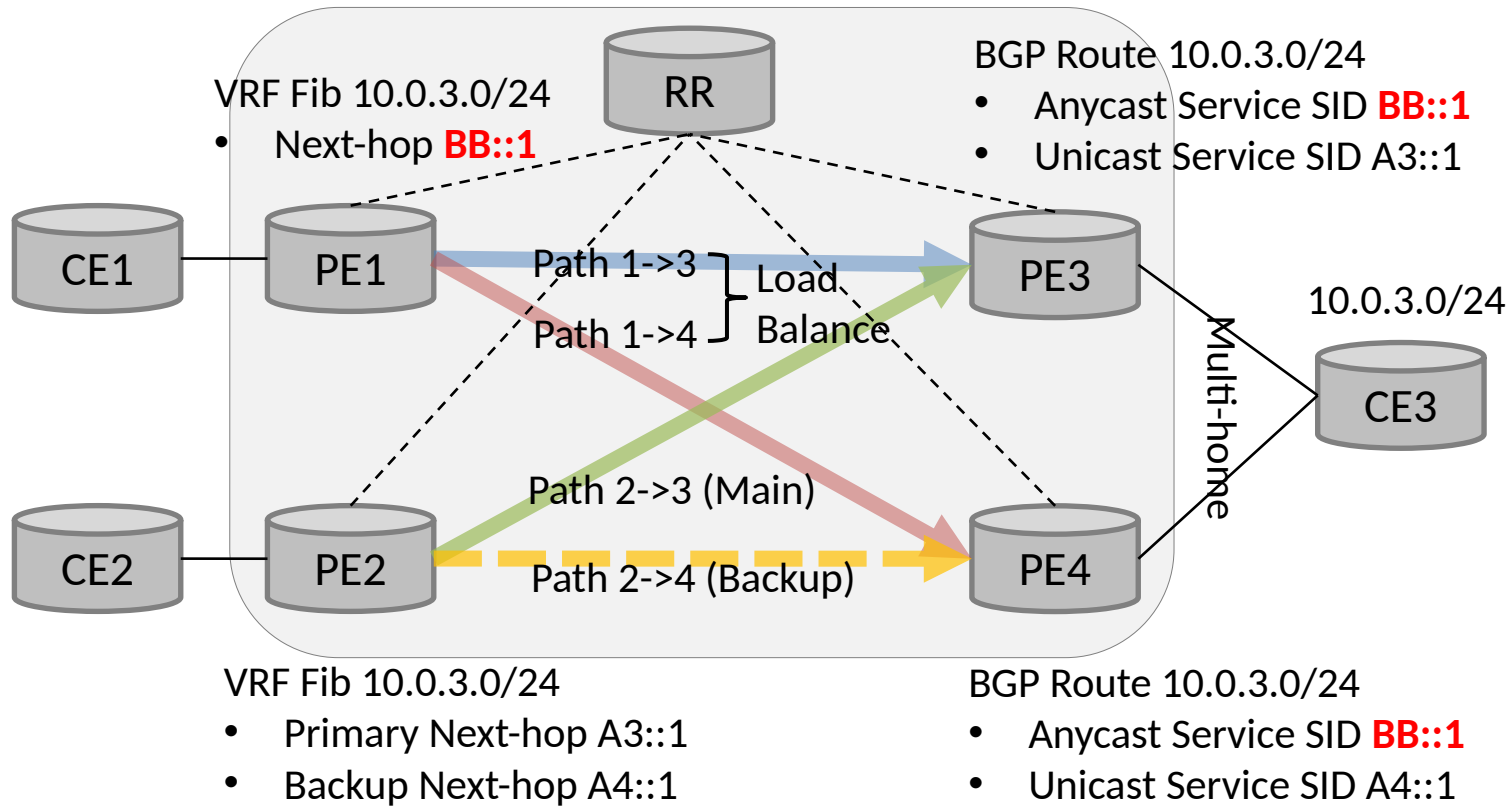
- CE3 is multi-homed to PE3 and PE4;
- PE3 and PE4 allocate unicast service SID for the VPN service.
- PE3 and PE4 can allocate a **anycast** service SID with same value.
- Anycast service SID needs to be allocated from a dedicated locator.
- Egress PE advertises both the anycast service SID and unicast service SID through BGP.
- Ingress PE1 and PE2 have different routing policies to select next-hop, result in using different service SID to forward traffic.
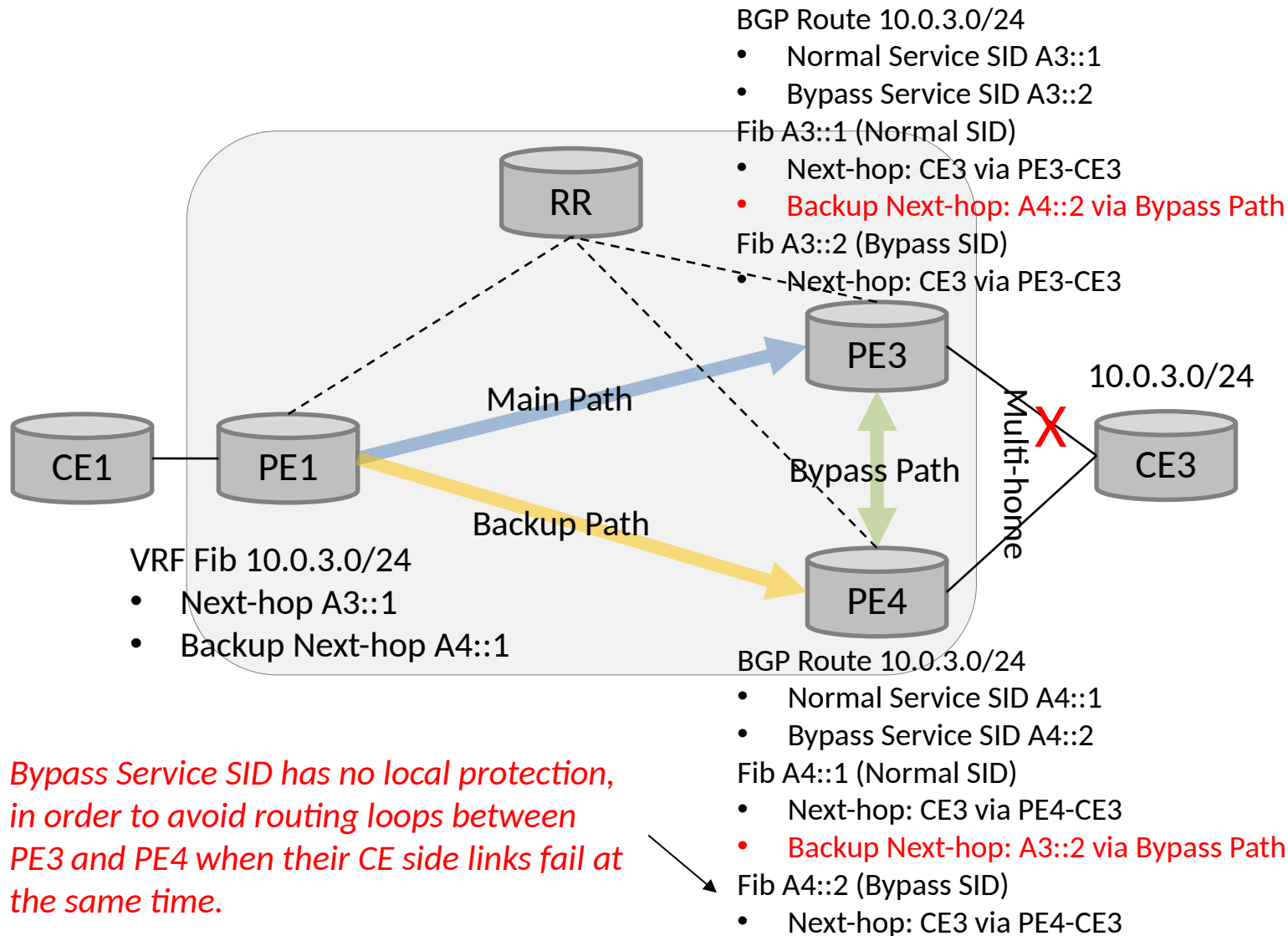
- PE1 uses the **anycast** service SID, and the traffic can be forwarded to PE3 and PE4 in a load-balanced manner
- PE2 can deploy VPN FRR, to uses the service SID of PE3 as the primary next hop, and PE4 as the backup next hop

Use Case: Bypass SRv6 Service SID

BGP Route 10.0.3.0/24
- Service SID A3::1

Fib A3::1
- Next-hop: CE3 via PE3-CE3 (Down)
- Backup Next-hop: **A4::1** via Bypass Path

BGP Route 10.0.3.0/24
- Service SID A4::1

Fib A4::1
- Next-hop: CE3 via PE4-CE3 (Down)
- Backup Next-hop: **A3::1** via Bypass Path

- PE3 and PE4 can use each other's Service SID to implement local protection for CE side link failure;

- When the CE-side link of PE3 fails, PE1 is not aware of the failure immediately and continues to forward packets to PE3

- PE3 can use the service SID of PE4 to send the traffic to PE4 and continue to forward it to CE3.

- However, if link PE3-CE3 and PE4-CE3 fail at the same time, loops occur between PE3 and PE4 until routing convergence;

- Additional bypass Service SID can be used to avoid loops:
  - **Egress PE does not compute backup next-hop for its own bypass Service SID;**
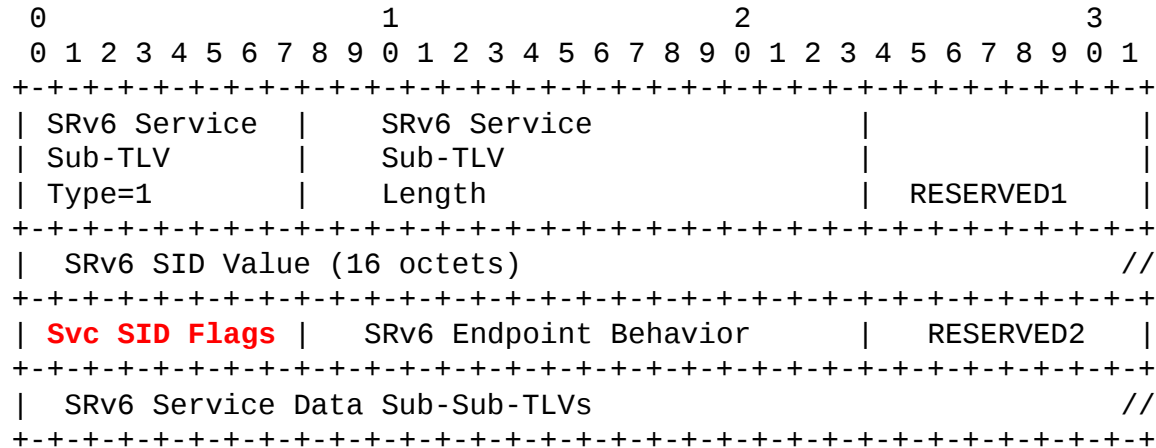 - **Egress PE use each other's bypass Service SID as backup;**

Use Case: Bypass SRv6 Service SID (Cont.)

BGP Route 10.0.3.0/24
- Normal Service SID A3::1
- Bypass Service SID A3::2

Fib A3::1 (Normal SID)
- Next-hop: CE3 via PE3-CE3
- Backup Next-hop: A4::2 via Bypass Path

Fib A3::2 (Bypass SID)
- Next-hop: CE3 via PE3-CE3

RR

PE3

Main Path

10.0.3.0/24

CE1

PE1

Multi-home

CE3

Bypass Path

Backup Path

PE4

VRF Fib 10.0.3.0/24
- Next-hop A3::1
- Backup Next-hop A4::1

BGP Route 10.0.3.0/24
- Normal Service SID A4::1
- Bypass Service SID A4::2

Fib A4::1 (Normal SID)
- Next-hop: CE3 via PE4-CE3
- Backup Next-hop: A3::2 via Bypass Path

Fib A4::2 (Bypass SID)
- Next-hop: CE3 via PE4-CE3

Bypass Service SID has no local protection, in order to avoid routing loops between PE3 and PE4 when their CE side links fail at the same time.
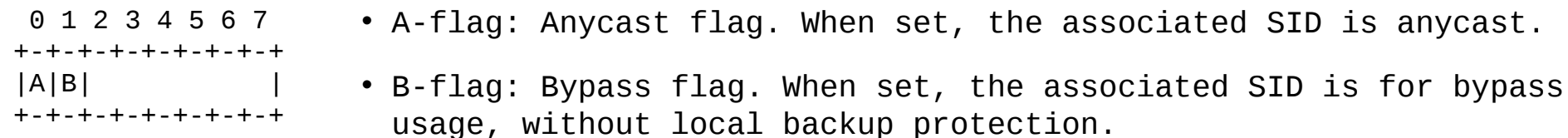
- Egress PE allocate bypass service SID and advertise it.
- Ingress PE always use unicast or anycast service SID to forward traffic.
- Egress PE use each other's bypass service SID as backup nexthop to CE.
- Bypass service SID has no backup nexthop
- When the CE-side link of PE3 fails, use the PE4's bypass service SID to send traffic to PE4.
- If the CE-side link of PE4 also fails at this time, PE4 will discard the traffic, thus avoiding a loop.
- The link failure on the CE side is not easy for the Ingress PE to detect. In a multi-homing scenario, Using bypass service SID, Egress PE could implement local protection y, avoiding traffic being discarded during ingress PE's perception.

New Flag Definition

[RFC9252] defines the SRv6 SID Information Sub-TLV to carry SRv6 Service SID in BGP messages.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| SRv6 Service  |    SRv6 Service                |             |
| Sub-TLV       |    Sub-TLV                     |             |
| Type=1        |    Length                      | RESERVED1   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  SRv6 SID Value (16 octets)                                 //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Svc SID Flags |    SRv6 Endpoint Behavior      |  RESERVED2  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  SRv6 Service Data Sub-Sub-TLVs                             //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

This draft defines two new flags in the SRv6 Service SID Flags field:

```
 0 1 2 3 4 5 6 7
+-+-+-+-+-+-+-+-+
|A|B|           |
+-+-+-+-+-+-+-+-+
```

- A-flag: Anycast flag. When set, the associated SID is anycast.
- B-flag: Bypass flag. When set, the associated SID is for bypass usage, without local backup protection.

These two flags cannot be set for the same service SID

Backward Compatibility Consideration

- About "Multiple SRv6 SIDs" in [RFC9252]:

    When multiple SRv6 SID Information Sub-TLVs are present, the ingress PE SHOULD use the SRv6 SID from the first instance of the Sub-TLV. An implementation MAY provide a local policy to override this selection.

    When the egress PE advertises multiple service SIDs, the unicast service SID needs to be carried in the first instance of sub-tlv

- About "SRv6 Service SID Flags" in [RFC9252]:

    Any unknown flags in the SRv6 Service SID Flags field MUST be ignored by the receiver.

    When there are PE routers not supporting the anycast and bypass flags, The egress PE may expect those routers to use the first SID and ignore the new-defined flags.

Next Steps

- Any questions or comments are Welcomed.

Thanks