Label Sharing for Fast PE Protection

draft-zhang-l3vpn-label-sharing-02

Mingui Zhang, Peng Zhou, Russ White

Advertising the same VPN route label

- PEs connected to the same set of CEs form a Redundant Group (RG).

- It's required that all egress PEs in an RG distribute the same VPN route label for the routes per VRF.

- Therefore, if a packet needs to be redirected to a backup egress PE at the PLR, its VPN route label remains intact.
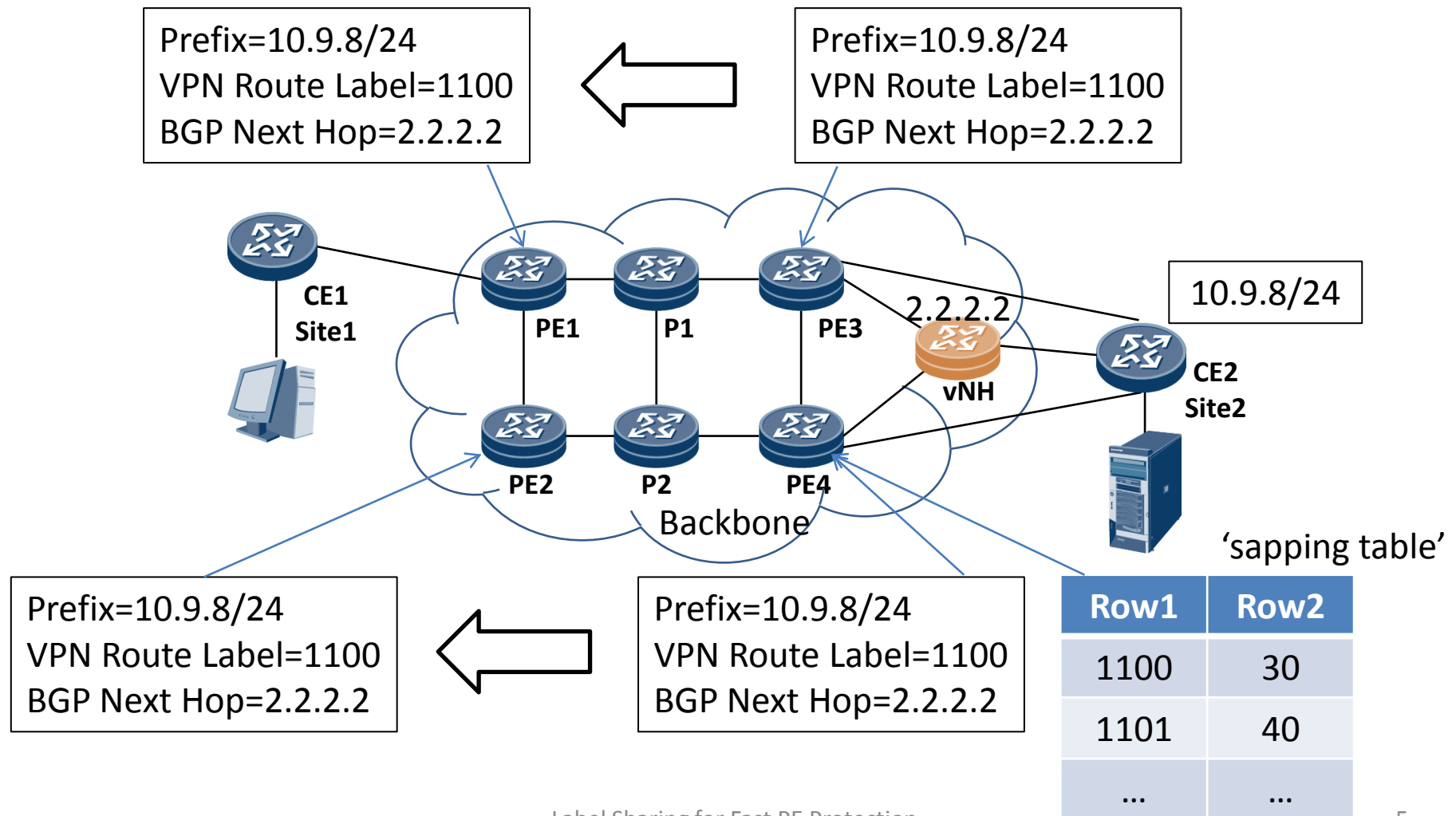
Primary & backup LSP tunnels

- A virtual router (vNH) is created in IGP to represent the RG.

- vNH will be used as the common last hop of the primary & backup LSP tunnels.

- In a normal VPN route propagation, egress PEs use their own IP addresses as the 'BGP next hop'. Here, all egress PEs use the IP address of the vNH as the 'BGP next hop'. In this way the upper-layer VPN route is bound to the under-layer LSP tunnels.
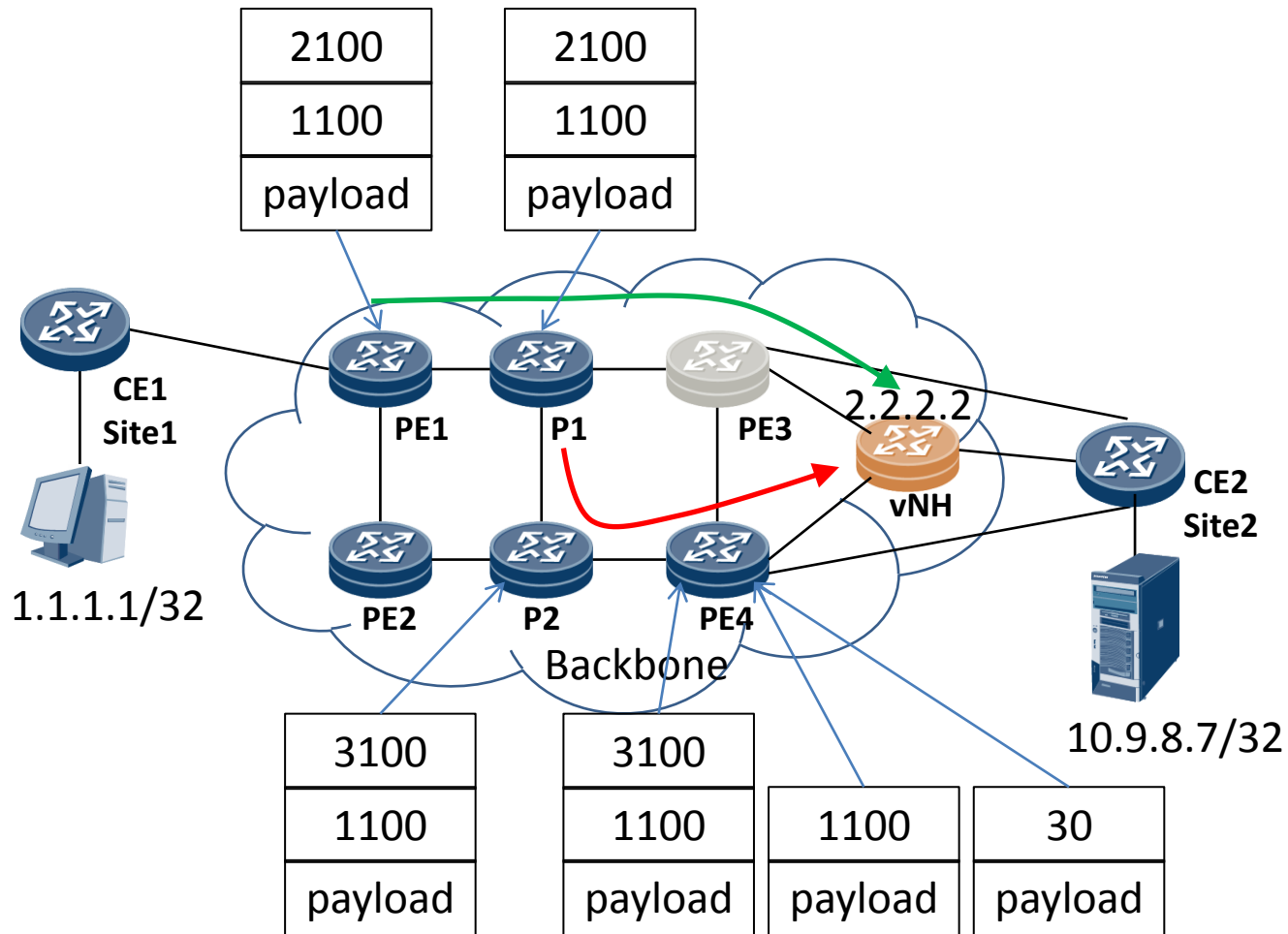
Overview
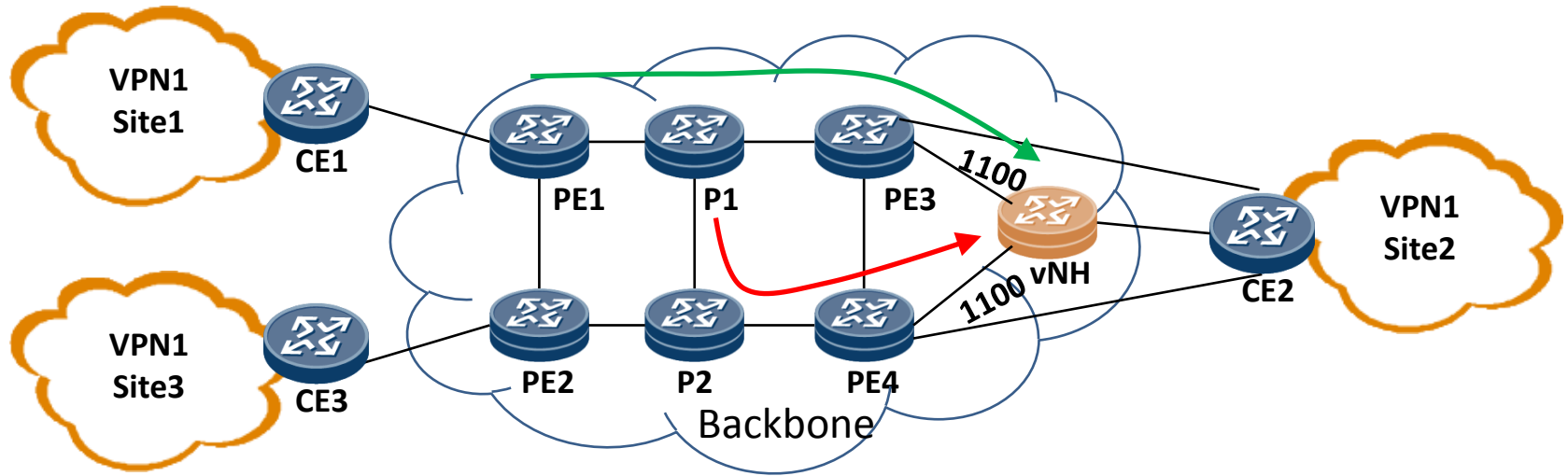
- Ingress PEs: PE1, PE2
- Egress PEs: PE3, PE4, they share the label 1100 for VPN1.
- Primary tunnel: PE1->P1->PE3-vNH
- When PE3 fails, backup tunnel P1->P2->PE4->vNH can be used.
- vNH: the virtual BGP Next Hop

Example on label distribution (Westbound)

Prefix=10.9.8/24
VPN Route Label=1100
BGP Next Hop=2.2.2.2

Prefix=10.9.8/24
VPN Route Label=1100
BGP Next Hop=2.2.2.2

CE1
Site1

PE1   P1   PE3

2.2.2.2

vNH

10.9.8/24

CE2
Site2

PE2   P2   PE4

Backbone

Prefix=10.9.8/24
VPN Route Label=1100
BGP Next Hop=2.2.2.2

Prefix=10.9.8/24
VPN Route Label=1100
BGP Next Hop=2.2.2.2

'sapping table'

Row1	Row2
1100	30
1101	40
...	...

Example on failure protection (Eastbound)

2100
1100
payload

2100
1100
payload

CE1
Site1

1.1.1.1/32

PE1 P1 PE3

2.2.2.2

vNH

CE2
Site2

10.9.8.7/32

PE2 P2 PE4

Backbone

3100
1100
payload

3100
1100
payload

| 1100 |
|---|
| payload |

30
payload

Incorporated comments

- How to guarantee two PEs allocate/advertise a common VPN label?
  - Option A: reserved label ranges per RG
  - Option B: the label swapping table
- Protection exceptions
  - The primary route has a longer prefix
  - The primary route has a preferred Local-Pref

Option A: reserved label ranges per RG

- The essence is to have an entity to allocate the VPN label for the whole RG. In this document, this entity is the primary PE.

- The operator needs to plan a space for RGs. Let each RG reserve the label range(s).

- This option should only be used in one administrative domain. Not over the public internet.

10, 000

300

RG2

200

RG1

100

0
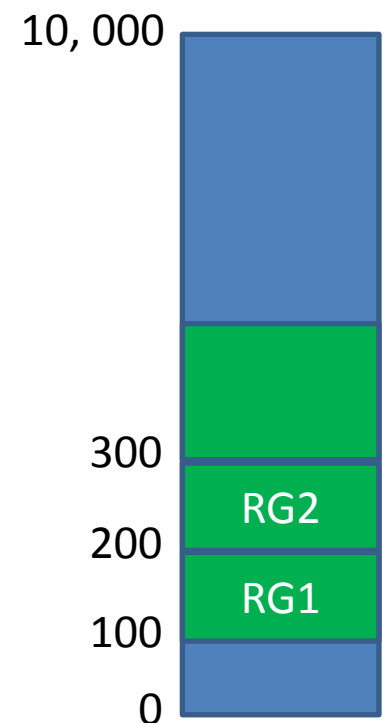
Option B: label swapping table

- The backup PE continues to allocate its own VPN route labels.

- It monitors the label allocated by the primary PE for VPN routes in a VRF and uses a swapping table to record the mapping between the heard label and the locally assigned label.

Row1	Row2
1100	30
1101	40
…	…

Row1: heard & distributed
Row2: locally allocated

Protection exceptions

- An operators is possible of violating the following requirement.

  - The primary and backup PEs are the penultimate hop of the LSP tunnels .

- Two exceptions had been proposed on the list.

  - Longer prefix
  - Preferred Local-Pref

- The backup LSP tunnel actually goes via the primary PE (e.g., PE4->PE3->vNH). Generate an alarm to the operator.

Next steps

- Get directions from the WG.

- We are ready to incorporate further comments to reach a consensus.

- Let's work together to enrich it and ease its implement and deployment.

Thanks!