

Scalable BGP FRR Protection Against Edge Node Failure

draft-bashandy-bgp-edge-node-frr-02

Authors :

Ahmed Bashandy, Cisco Systems
Keyur Patel, Cisco Systems
Burjiz Pithawala, Cisco Systems

Presenter :

Ahmed Bashandy

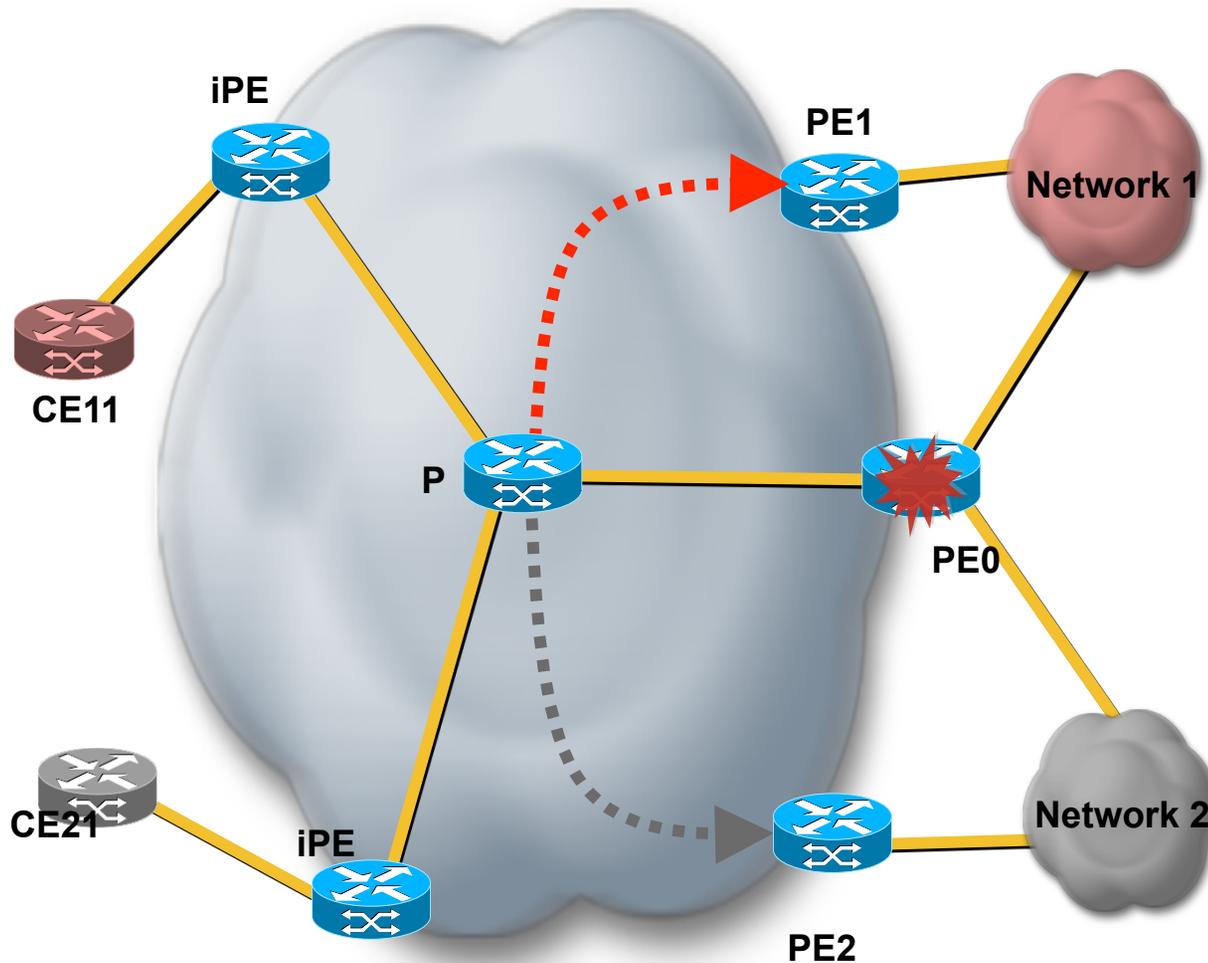
IETF83, Mar/2012

Paris, France

Agenda

- ◆ Problem and requirement
- ◆ Proposed Solution

Problem



- PE0 is primary for both **Red** and **Gray**.
- PE0 *fails* !!
- P router redirects traffic to the **correct** repair PE
 - PE1 for **Red**
 - PE2 for **Gray**
- Correct **BGP label** must exist for correct forwarding on repair PE

What we are trying to Achieve

- ◆ Packet must be forwarded to correct repair PE on primary PE failure
- ◆ Correct BGP label must be pushed when repairing
- ◆ Core remains BGP free
- ◆ Minimal provisioning
- ◆ Loop-free during repair

Control Plane Steps

1. Choose the repair PE
2. Assign and Advertise the next-hop for protected prefixes
3. Inform repairing core routers about primary to repair path mapping
4. Programming the forwarding plane on the repairing routers

Control Plane (1): Choosing repair PE

- ◆ Each PE that has an external path to a prefix P/m also chooses a repair PE
 - Any other PE that advertises an external path to P/m
 - The other PE MAY also advertises a **repair label (rL)**
 - **Optional non-transitive** as in `draft-bashandy-idr-bgp-repair-label`
 - **Same semantics as** `draft-bashandy-idr-bgp-repair-label`
 - Either send the packet out or drop it
 - **MUST Be Per-CE or Per-VRF**
 - To keep the core **BGP-free**
 - Needed for **good** attribute packing

Control Plane (2): Next-hop for P/m

- ◆ A PE now has a repair PE and possibly repair label for a protected prefix
- ◆ **Group** prefixes as follows:
 - Prefixes without repair labels
 - Two prefixes belong to the same group G_i if they share the same repair PE
 - Prefixes with repair label
 - Two prefixes belong to the same group G_i if they share the same repair PE and repair label
- ◆ Assign each a group **G_i** a separate protected next-hop **pNH_i**
 - Can be assign from a range
 - pNH_i must be unique within a core: must not collide with other next-hops
- ◆ Advertising the pNH_i to iBGP peers
 - May be the next-hop attribute of BGP: Good for backward compatibility
 - Can be optional non-transitive attribute: Less churn but requires ingress PEs to understand it

Control Plane (3): Informing Core Routers

- ◆ Egress PE needs to inform core routers about repair info: pNH, rNH, and rL
 - pNH s advertised into IGP
 - rNH is an IP address for the repair PE → it is advertised into IGP as usual
 - What is left is mapping of **pNH** to **rNH** and **rL**
- ◆ If there is no repair label **rL**
 - Advertise the pair (pNH,rNH) to repairing routers (e.g. through optional LDP or ISIS TLV)
 - The semantics of (pNH,rNH) is

*If the next-hop pNH becomes unreachable, then traffic tunneled to the next-hop pNH SHOULD be immediately **re-tunneled** to rNH, **without waiting for IGP or BGP to re-converge**, because rNH can reach protected prefixes reachable via pNH.*
- ◆ If there is a repair label **rL**
 - Advertise the quadruple (pNH,rNH, rL, Push) to repairing routers (e.g. through optional LDP or ISIS TLV)
 - The semantics the quadruple (pNH,rNH, rL,Push) is
 1. If the next-hop **pNH** becomes unreachable, then traffic tunneled to the next-hop **pNH** SHOULD be immediately **re-tunneled** to **rNH** , **without waiting for IGP or BGP to re-converge**, because **rNH** can reach protected prefixes reachable via **pNH**.
 2. If the **Push** flag is cleared, the label underneath the tunnel encapsulation PE MUST be swapped with the label **rL** before re-tunneling to the repair PE, irrespective of the value of the label below the tunnel encapsulation.
 3. If the **Push** flag is set, then the label **rL** MUST be pushed on the packet before re-tunneling to the repair PE* .

Control Plane (4): FIB in Core routers

- ◆ Assume pNH matches the IGP router pR
- ◆ Thus the FIB entry for pR is programmed as follows
 - Primary path: Next hop router on the path towards pNH
 - Repair path when the candidate repair router receives the pair (pNH, rNH)
 - Next-router on the path towards rNH
 - Repair path when the candidate repair router receives the quadruple $(pNH, rNH, rL, Push)$
 - Primary path: Next router on the path towards pNH
 - Repair path:
 - Next router towards rNH but with additional semantics
 - If the "**Push**" flag is cleared
 - » Pop label in the packet right under the tunnel header (irrespective of the value of that label)
 - EndIf
 - Push the underlying repair label **rL**

Forwarding Plane on Repairing Router on Failure

- ◆ The repairing P router detects that pNH is no longer reachable
- 1. Decapsulate the tunnel header to expose the tunneled packet
- 2. If the underlying repair label *rL* is programmed in the forwarding plane
 - 1. If the "**Push**" flag is set
Push the underlying repair label *rL*
 - 2. Else
Swap the label under the tunnel encapsulation (irrespective of the value of that label) with the underlying repair label *rL*
- 3. Tunnel the packet towards rNH

Q & A

