Path-aware Remote Protection Framework

draft-liu-rtgwg-path-aware-remote-protection-01

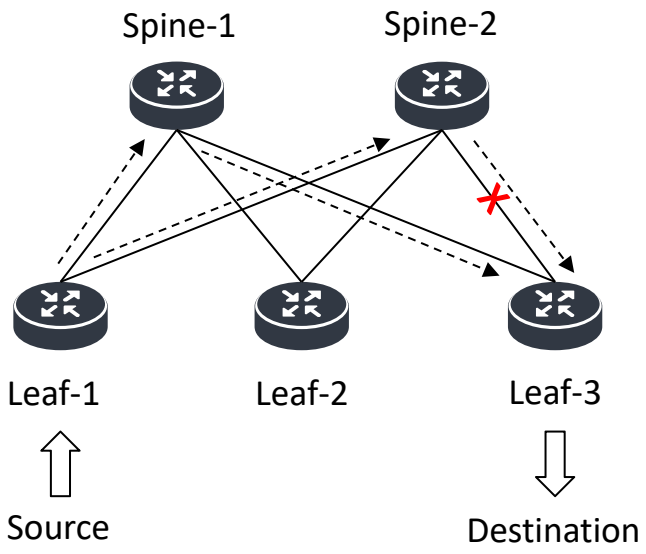
Presenter: Yisong Liu (China Mobile)

Co-authors: Yisong Liu (China Mobile)
Changwang Lin (New H3C Technologies)
Mengxiao Chen (New H3C Technologies)
Zheng Zhang (ZTE Corporation)

IETF-119, March 2024

Motivation

- Current IP network protection mechanisms can be mainly divided into local protection and end-to-end protection.
  - Local protection technologies, such as ECMP, LFA, and TI-LFA, can only perceive local failures and requires IGP.
  - End-to-end protection technologies are used for end-to-end TE paths. The head-end performs detection and switchover.
- There are some networks where current protection mechanisms cannot cover.



In a spine-leaf based DC network:

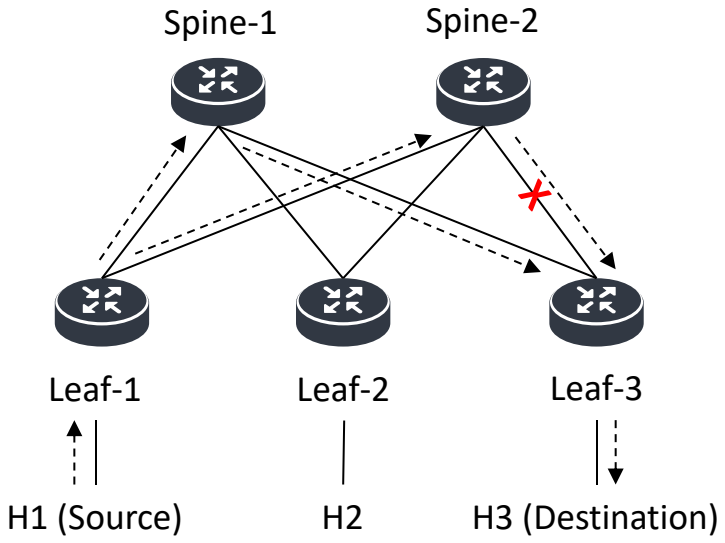
- Only BGP protocol is deployed, no IGP.
- IP-based BE forwarding, no TE.

When the failure occurs:

- Leaf-1 will continue to send traffic to both Spine-1 and Spine-2, until Leaf-1 receives BGP withdrawn routes from Spine-2.
- Waiting for control plane convergence would be quite long when there is a large number of BGP routes.

The idea is to allow Leaf-1 to detect the remote failure on the link between Spine-2 and Leaf-3, and then to invoke fast repairs!

Use Case 1 - Spine-Leaf Network



Forwarding table of Leaf-1:

Path-aware

H2:

Next-hop: Spine-1 -> Path: Spine-1, Leaf-2

Next-hop: Spine-2 -> Path: Spine-2, Leaf-2

H3:

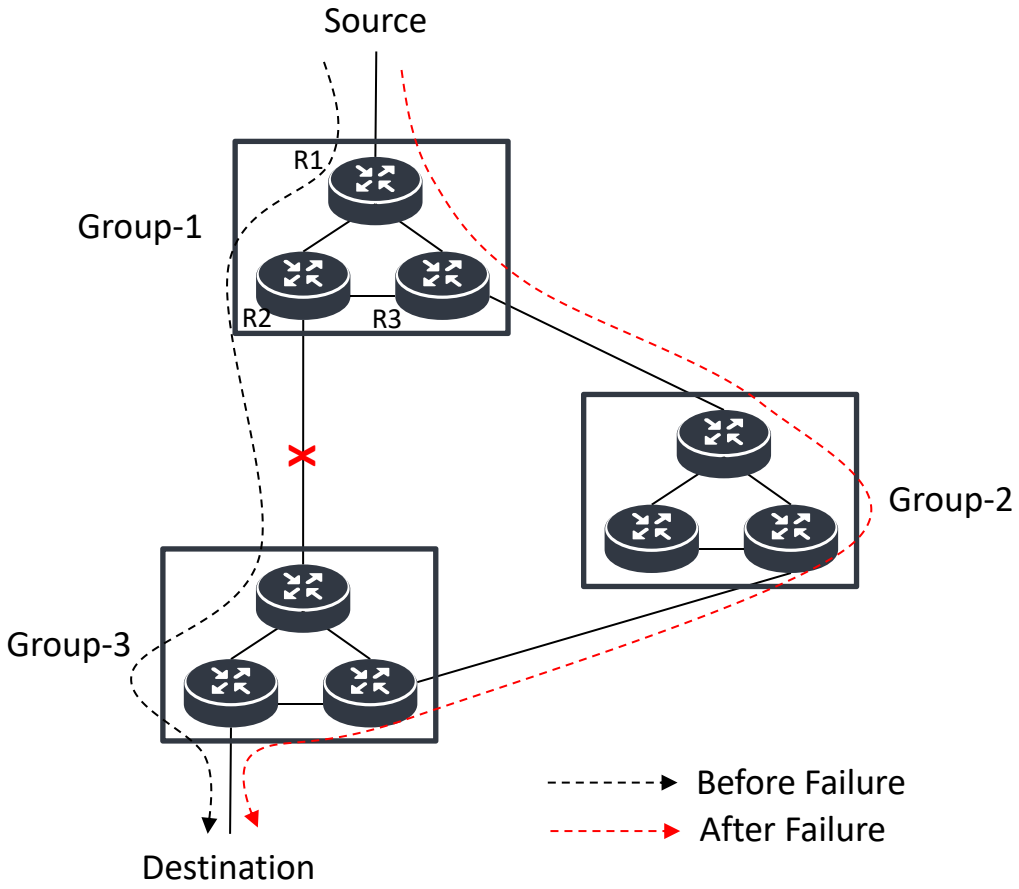
Next-hop: Spine-1 -> Path: Spine-1, Leaf-3

Next-hop: Spine-2 -> Path: Spine-2, Leaf-3

When the failure occurs on the link between Spine-2 and Leaf-3 :

1. Spine-2 detects the failure, and then notifies Leaf-1 of the failure on the link between Spine-2 and Leaf-3.
2. Leaf-1 finds the next-hop whose path has failure, and removes it from ECMP.

Use Case 2 - Dragonfly Network



Forwarding table of R1:

Path-aware

Destination:

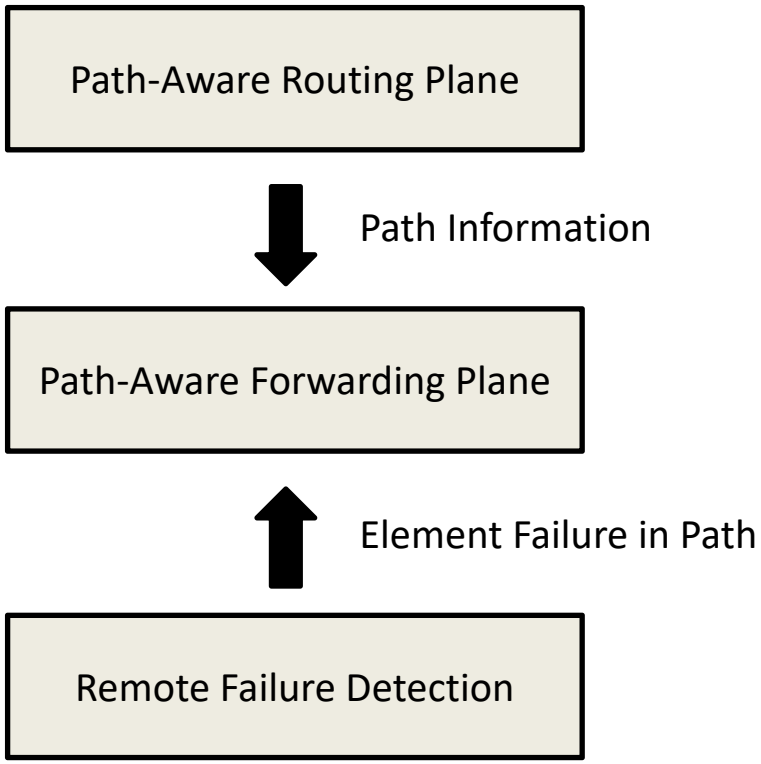
Primary Next-hop: R2 -> Path: R2, Group-3

Backup Next-hop: R3 -> Path: R3, Group-2, Group-3

When the failure occurs on the link between Group-1 and Group-3:

1. R2 detects the failure, and then notifies R1 of the failure on the link between Group-1 and Group-3.
2. R1 finds the next-hop R2 whose path has failure, and performs switchover from primary next-hop R2 to backup next-hop R3.

Framework



On the routing plane, the route calculation is not limited to the next hop, but requires path awareness.

The path information is downloaded to the forwarding plane.

When a failure occurs in any component along the path, it is required to quickly detect the failure and invoke repairs.

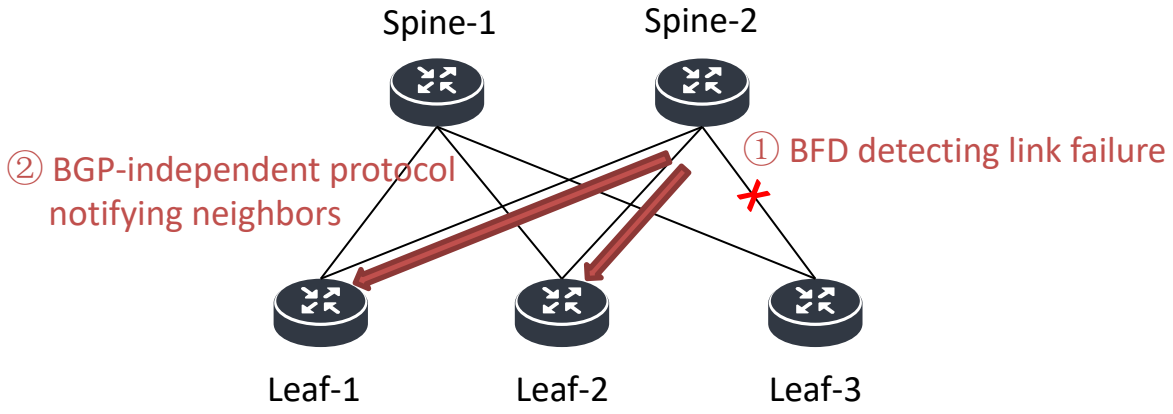
Remote Failure Detection

When a failure occurs, it is first detected by the router adjacent to it. The local failure detection may be based on existing techniques such as BFD. Then, that router notifies its neighbors of the failure, especially the upstream neighbors.

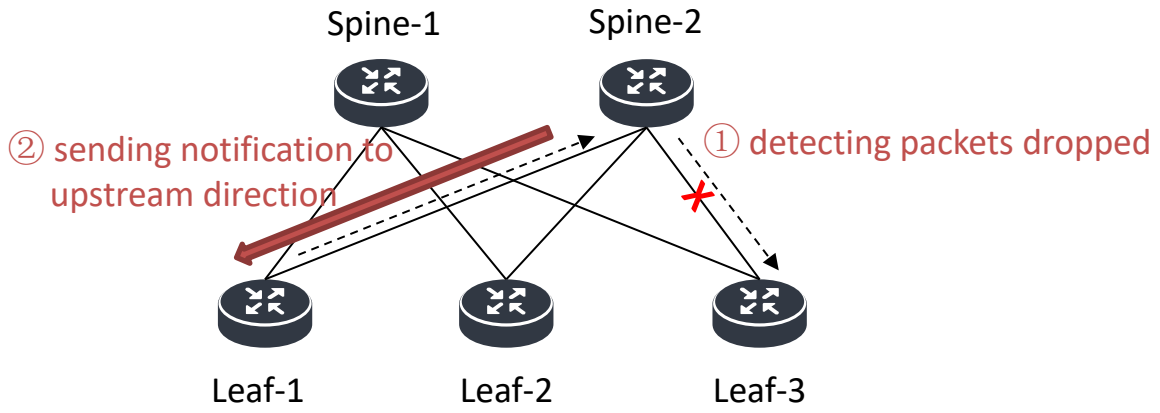
The failure notification between neighboring routers has the following requirements:

- Independent of routing protocols.
- Avoiding broadcast flooding.

Example-1:

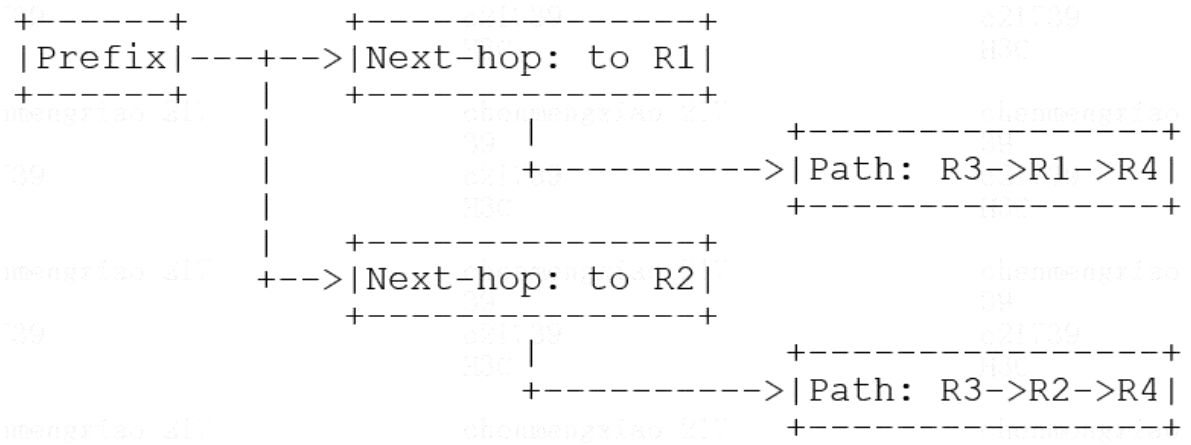


Example-2:



Path-Aware Forwarding Plane

Each next-hop is associated with a path.



When receiving failure notification from a neighbor, the next-hop entries corresponding to that neighbor will be checked to determine whether the associated path information contains the failed component.

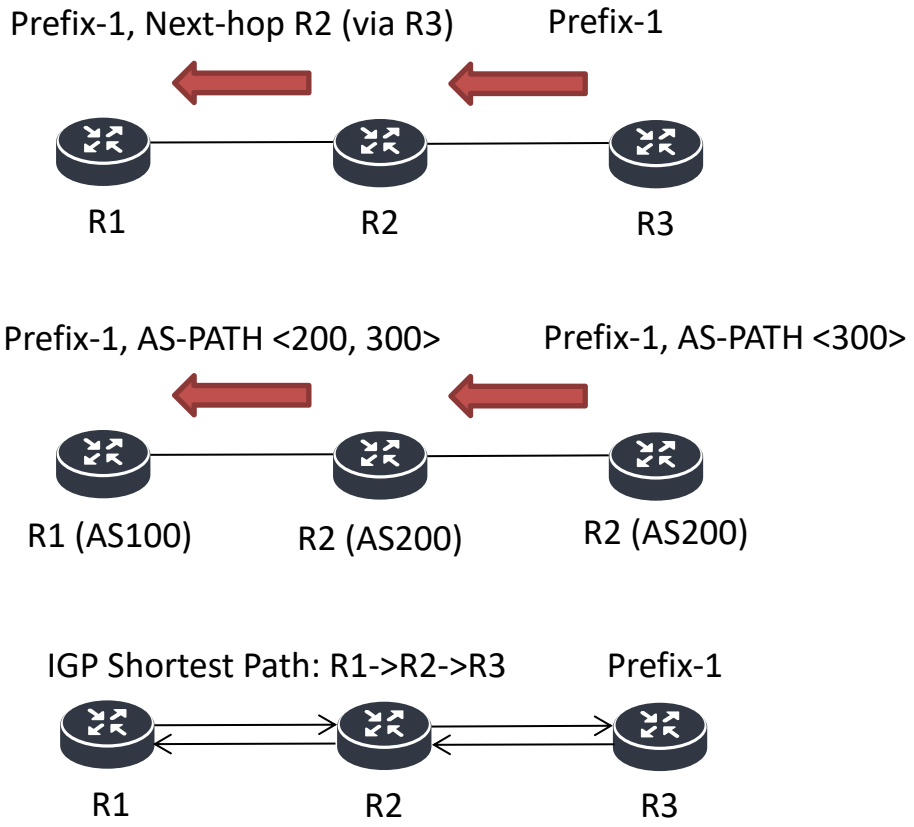
If detecting any failure in the path, the corresponding next-hop is regarded as failed.

Path-Aware Routing Plane

When calculating routes, the path needs to be perceived and the path information will be attached to the next hop.

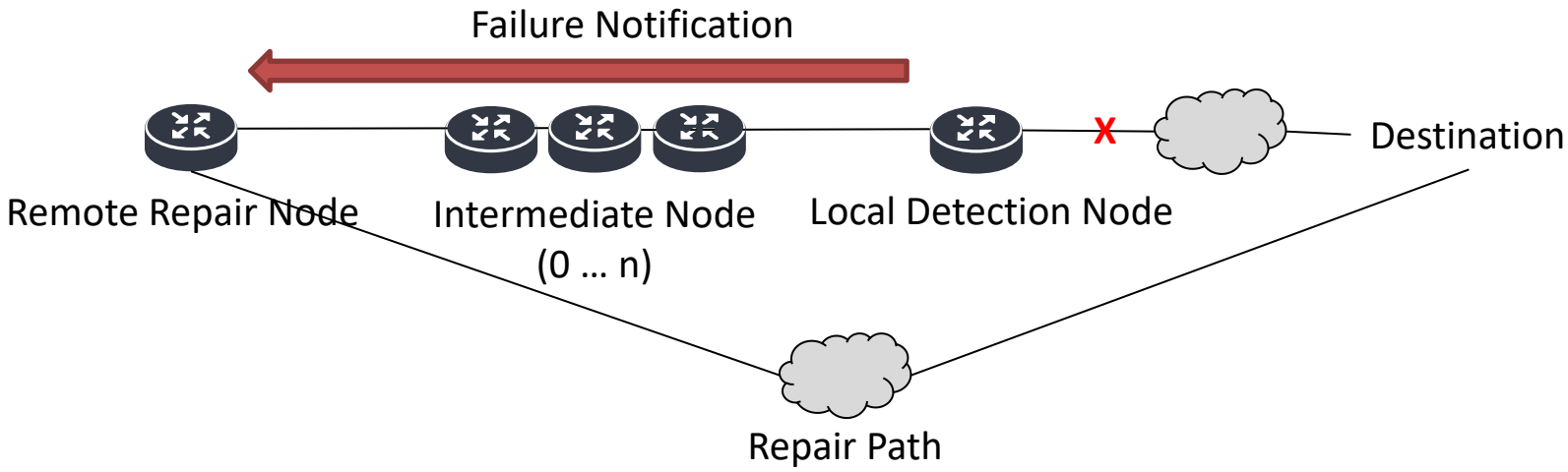
For examples:

- In a BGP-based network, a BGP route may carry the router-id of the peer from which that route is received, and the router-id will be added into the path information when calculating that route. The BGP protocol may needs some extensions to support such feature.
- For an EBGP-based DC network, a router may use the AS-PATH attribute (with SEQUENCE type) in the BGP route as the path information, without any protocol extensions.
- In an IGP-based network, a router may compute the path information based on the SPF tree and attach it into the next hop.



Role Types

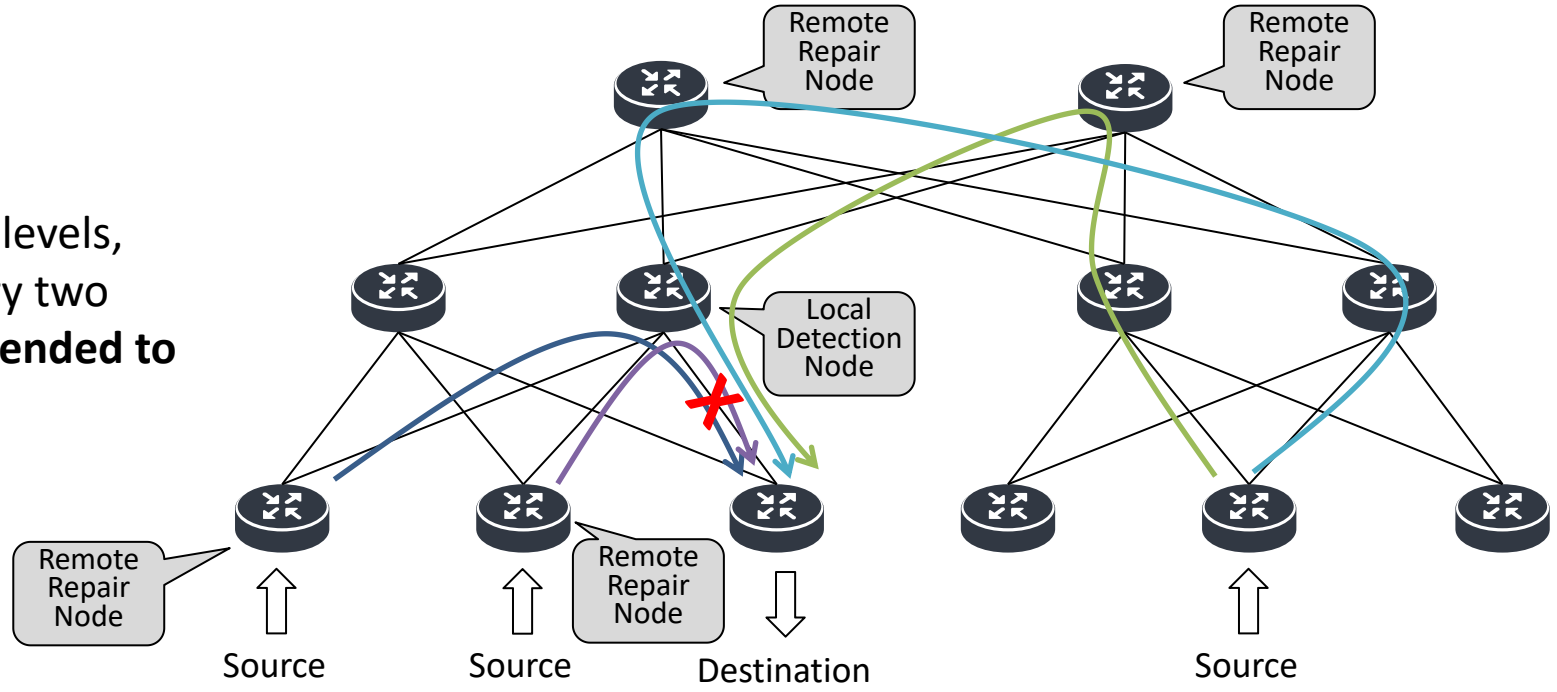
- **Remote repair node**: It has the repair path and provides the remote protection function.
- **Local detection node**: It is adjacent to the failure and detects the failure first. Then, it sends failure notification messages to the remote repair node.
- **Intermediate node**: It exists only if there are multiple hops between the remote repair node and the local detection node. It helps deliver the failure notification messages from the local detection node to the remote repair node.



Network Protection Coverage Scope

- The scope of path-aware remote protection covers from the remote repair node to the failure. The minimum scope covers two hops
- As the protection scope increases, the number of intermediate nodes increases, which may slow down the convergence time and expand the propagation of fault notification

For example:
In a spine-leaf network with multiple levels, usually there are ECMP paths on every two levels. **Remote protection is recommended to cover two hops**.



Next Steps

- Seeking for more scenarios
- Welcome any questions or comments

Thanks