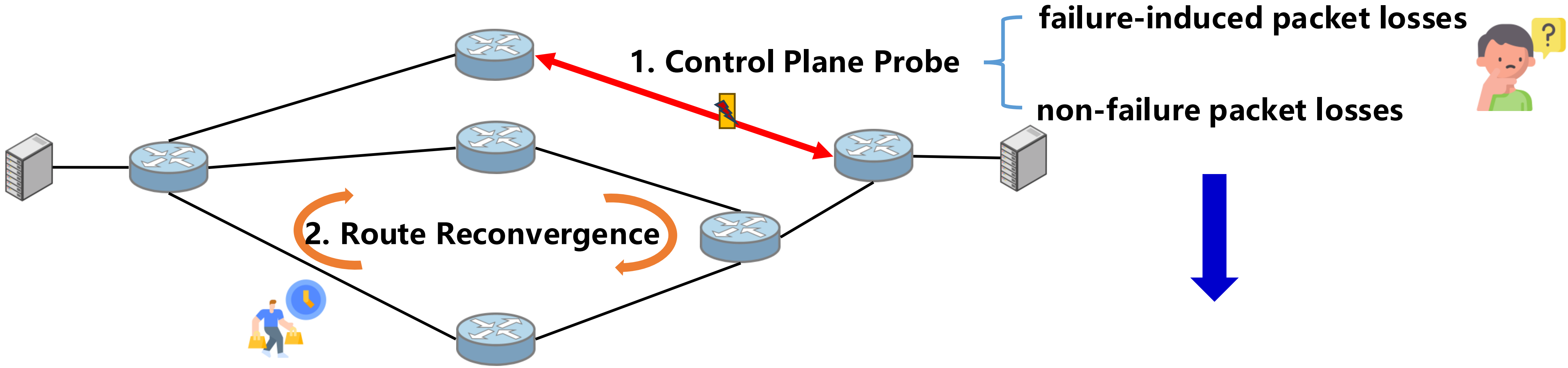
IETF 122 FANTEL

Fast Reroute based on Programmable Data Plane (PDP-FRR)

Dan Li, Kaihui Gao, **Shuai Wang**, Li Chen, Xuesong Geng

Fantel Side Meeting, IETF 122
2025-03-18

Motivation



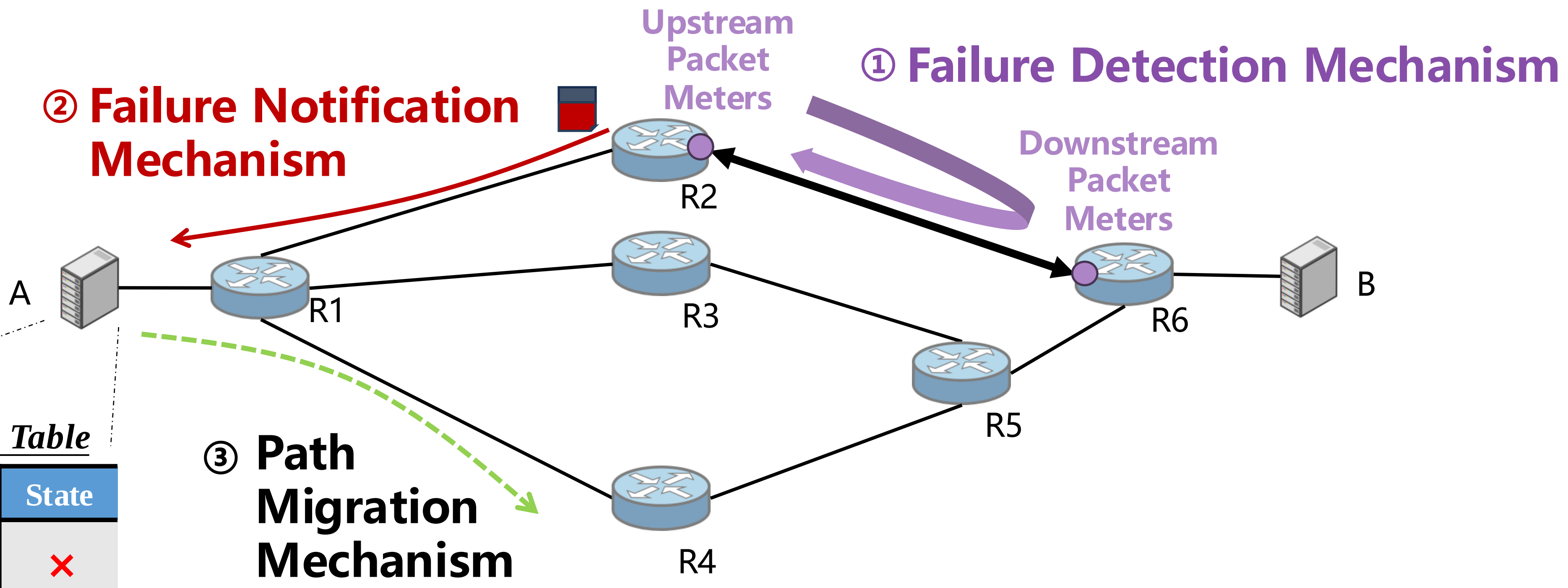
- Traditional network failure detection methods generate probe packets through the control plane (such as BFD), **treating the network data plane as a black box**.
  - If there is no response to a probe, it is assumed that a link failure has occurred, **without the ability to distinguish between fault-induced packet loss and non-fault packet loss (such as congestion loss, policy loss, etc.).**
- **Route reconvergence** in the control plane is time-consuming and results in slow reroute speed

PDP-FRR Architecture

Fast reroute based on programmable data plane (PDP-FRR) architecture **leverages the capabilities of the programmable data plane to significantly reduce the time required to detect link failures and reroute traffic**, thereby enhancing the overall robustness of datacenter networks.

- PDP-FRR architecture stands at the forefront of innovation by **integrating in-band network telemetry (INT) with source routing (SR) to facilitate rapid path migration directly within the data plane**
- PDP-FRR adopts a **white box** modeling of the data plane's packet processing logic
  - By deploying packet counters at both ends of a link and comparing them periodically, PDP-FRR can **identify failure-induced packet losses with unprecedented speed and accuracy**.
  - By pre-maintaining a path information table and utilizing SR (e.g., SRv6 and SR-MPLS), PDP-FRR **enables the sender to quickly switch traffic to alternative paths without the need for control plane intervention**.

PDP-FRR Architecture

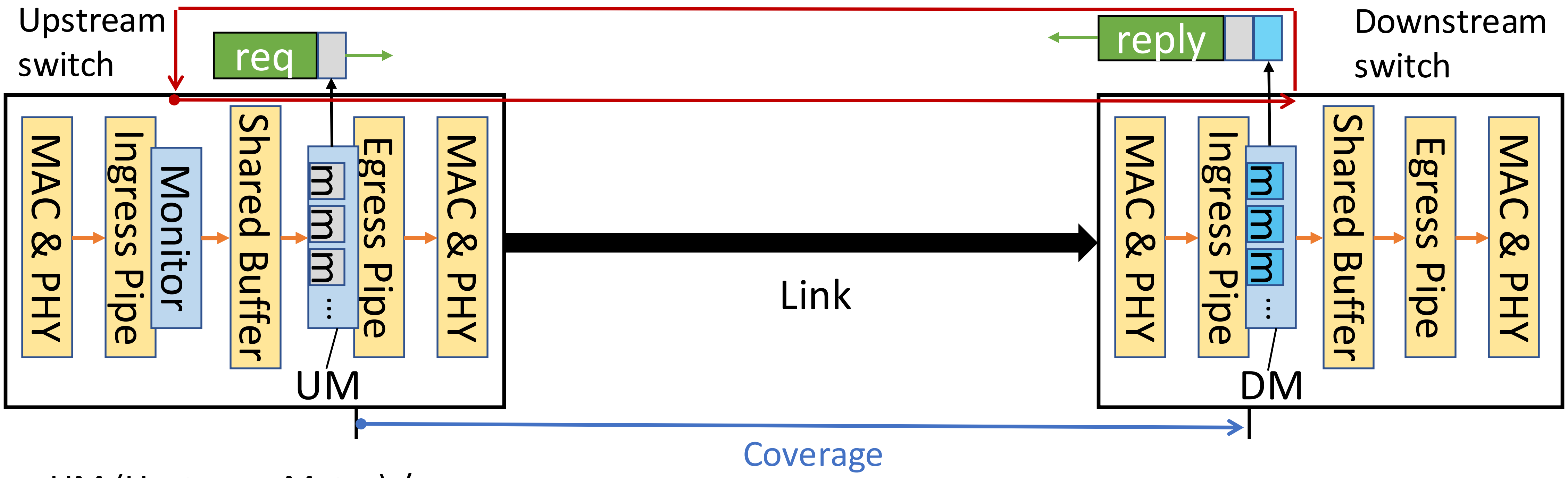


Path Information Table

Dest.	Path	State
B	R1→R2→R6	×
B	R1→R3→R5→R6	√
B	R1→R4→R5→R6	√

- **Failure Detection Mechanism:** distinguishing between packet losses caused by failures and normal packet losses
- **Failure Notification Mechanism:** leveraging switches convey failure information back to the end hosts via INT
- **Path Migration Mechanism:** The end hosts utilize SR to change the paths used by the traffic

1. Failure Detection Mechanism



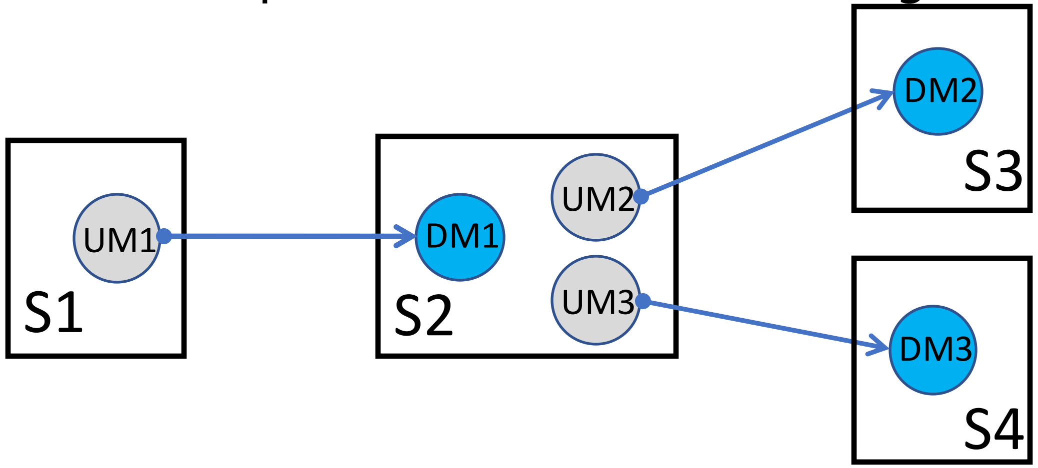
UM (Upstream Meter) /
DM (Downstream Meter):

- m — Total number of packets
- m — Number of packet losses due to **congestion**
- m — Number of packet losses due to **ACL**

1. UM/DM is used to **count the number of packets**, both flowing through, and actively dropped.
2. For a link, the upstream switch **sends request packets** to the outgoing port **to record these meters**.
3. The downstream switch **bounces it back**, and the upstream switch **calculates the packet loss rate** due to failure.

1. Failure Detection Mechanism

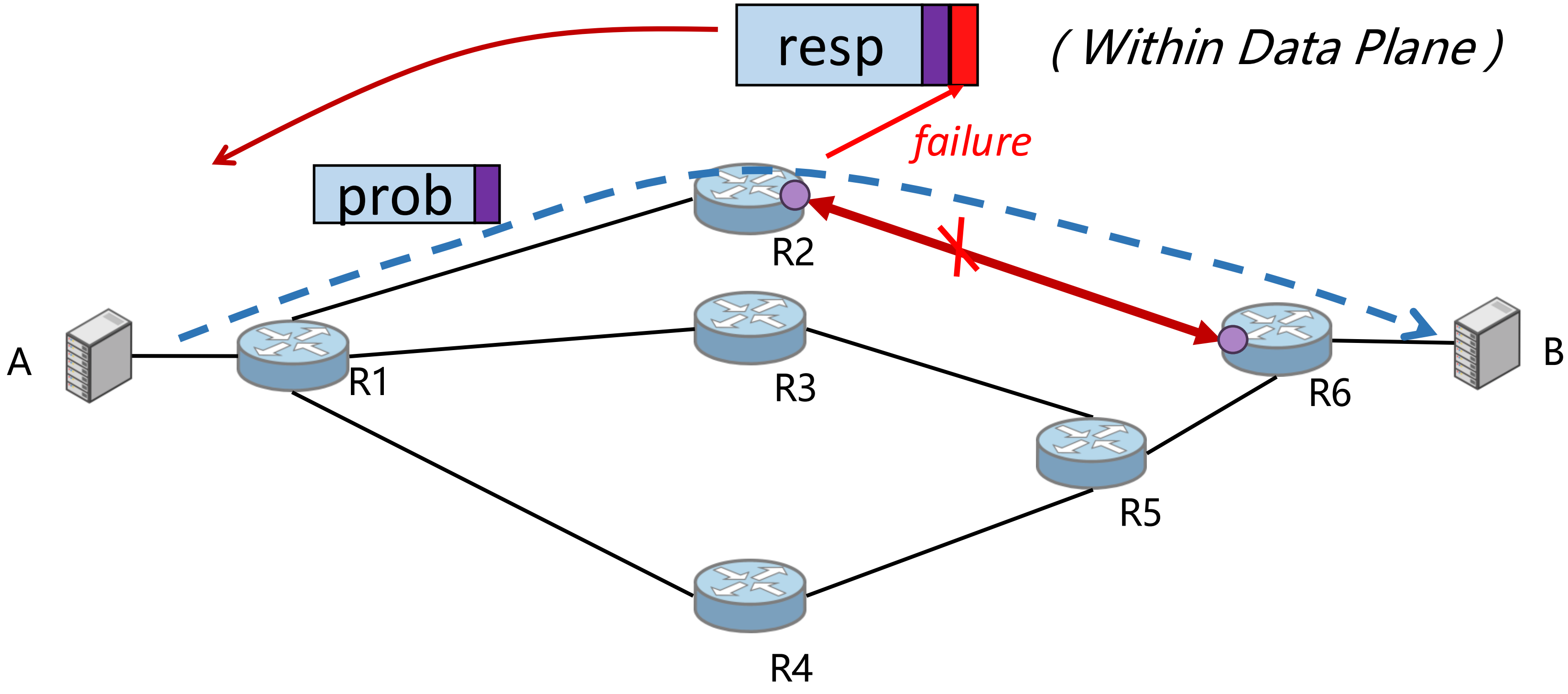
- **FDM (UM and DM) are deployed on all network links**
  - Adjacent switches can collaborate to detect failures of any type (including gray failures), and the mechanism is capable of accurately distinguishing non-failure packet losses, thus avoiding false positive.



- **An example**
  - Assume that 100 packets pass through the upstream switch UM, which records [100,0], with 0 representing no non-fault-related packet loss.
  - Suppose 8 packets are dropped on the physical link and 2 packets are dropped at the ingress pipeline of the downstream switch due to ACL rules.
  - Then, the DM records [90,2], where **90 represents the number of packets that passed through DM, and 2 represents the number of packets dropped due to non-fault reasons**.
  - Finally, by comparing the UM with DM, FDM calculates the packet loss rate of the link as 8% ((100-90-2)/100), rather than 10%.

2. Failure Notification Mechanism

Traditional control plane reroute schemes require several steps after detecting a failure, including **failure notification, route learning, and routing table updates**, which can take **several seconds** to modify traffic paths.

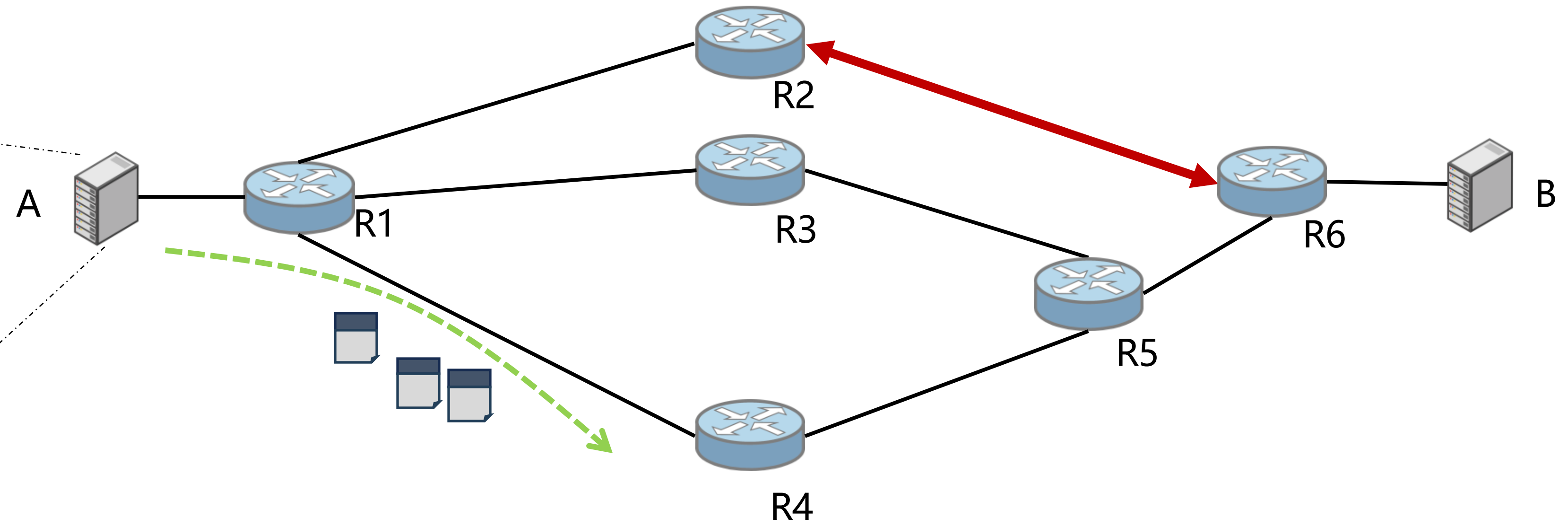


1. If there is a link failure, the upstream switch **marks the link as the failure state**
2. The sender sends **periodic probe packets** along the path of the data traffic
3. The switch bounces it back to the sender to **notify the failure**

3. Path Migration Mechanism

Path Information Table

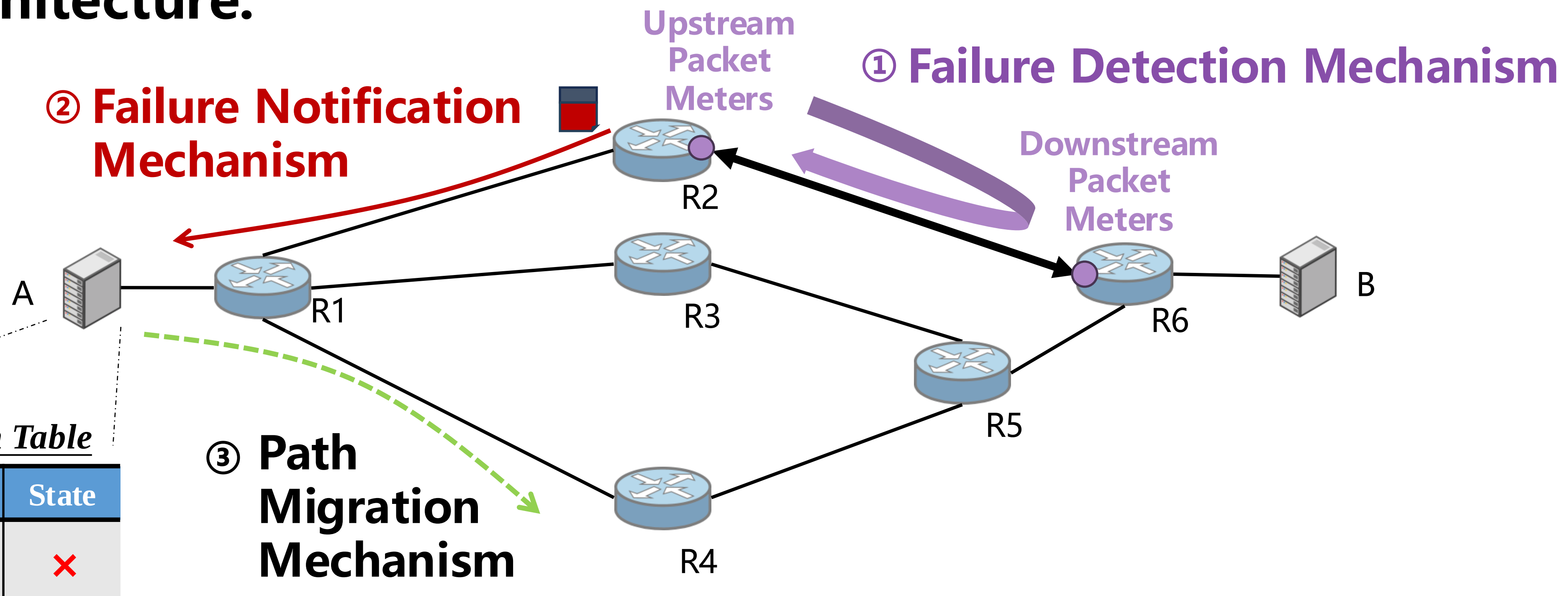
Dest.	Path	State
B	R1→R2→R6	✗
B	R1→R3→R5→R6	√
B	R1→R4→R5→R6	√



- To enable sender-driven fast reroute within data plane, the sender needs to **maintain a path information table in advance** so that it can quickly switch to another available path upon detecting network failure.
  - Within the transport layer protocol stack of the sender, this document designs a Path Migration Mechanism (PMM), which **periodically probes all available paths to other destinations**.
    - This information can also be obtained through other means, such as from an SDN controller.
  - Then, for a new flow, the sender will **randomly select an optimal available path** from the path information table and use source routing (e.g., SRv6 and SR-MPLS) to control the path of this flow.

Thanks! ☺

PDP-FRR Architecture:



Path Information Table

Dest.	Path	State
B	R1→R2→ R6	✕
B	R1→R3→ R5→R6	√
B	R1→R4→ R5→R6	√

- **Running entirely within the data plane**
- Relying on **programmable network devices**, such as P4 switches, Smart NICs
- The sender sends **periodic probe packets** along the path of the data traffic