Micro-loop avoidance using SPRING

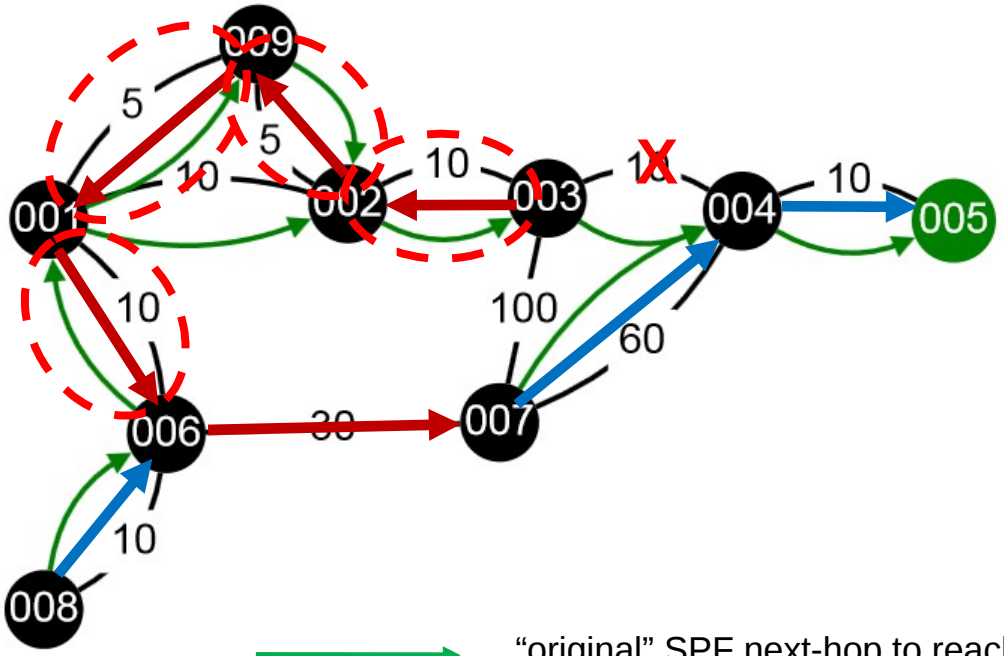
draft-hegde-rtgwg-microloop-avoidance-using-spring-00

Shraddha Hegde (shraddha@juniper.net)
Pushpasis Sarkar (psarkar@juniper.net)

AGENDA

- Problem Statement
- Solution using SPRING Tunnels
- Multiple events handling
- Partial deployment
- OSPF/ISIS Extensions

Problem Statement



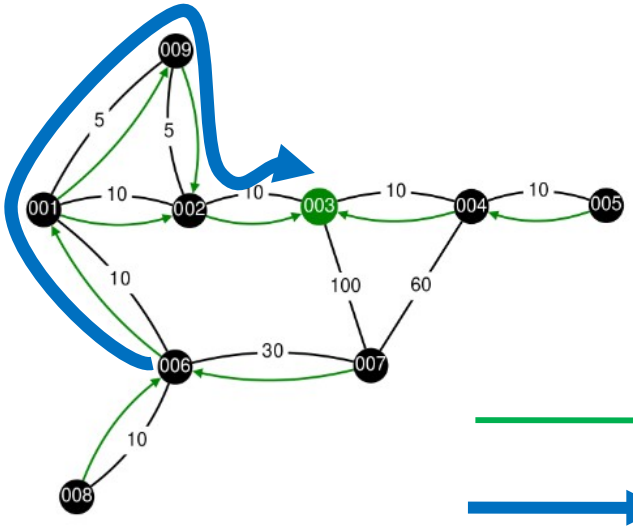
- Transient loops due to unsynchronized FIB state across nodes
- Certain topologies are more prone to micro-loops Ex: Rings
- Micro-loop between nodes 6,1,2, 9 and 3 for destination 5 when link between 3 and 4 goes down

"original" SPF next-hop to reach 005

Unchanged SPF next-hop to reach 005 after event

Changed SPF next-hop to reach 005 after event

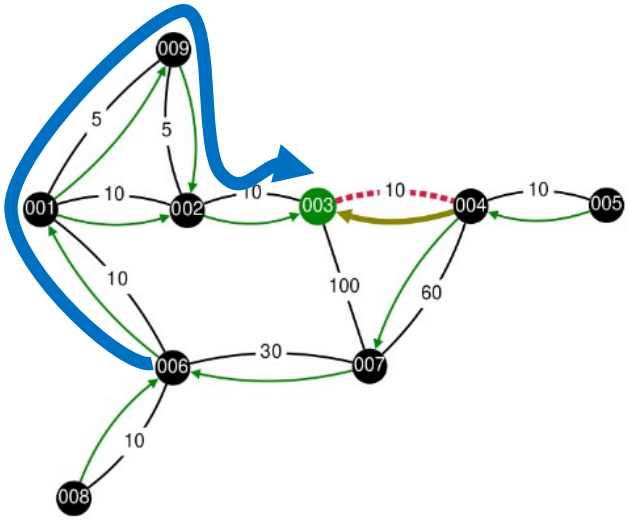
Solution (near-end tunnelling)



Before failure

SPF next-hops to reach 003

Shortest-path tunnel to PLR

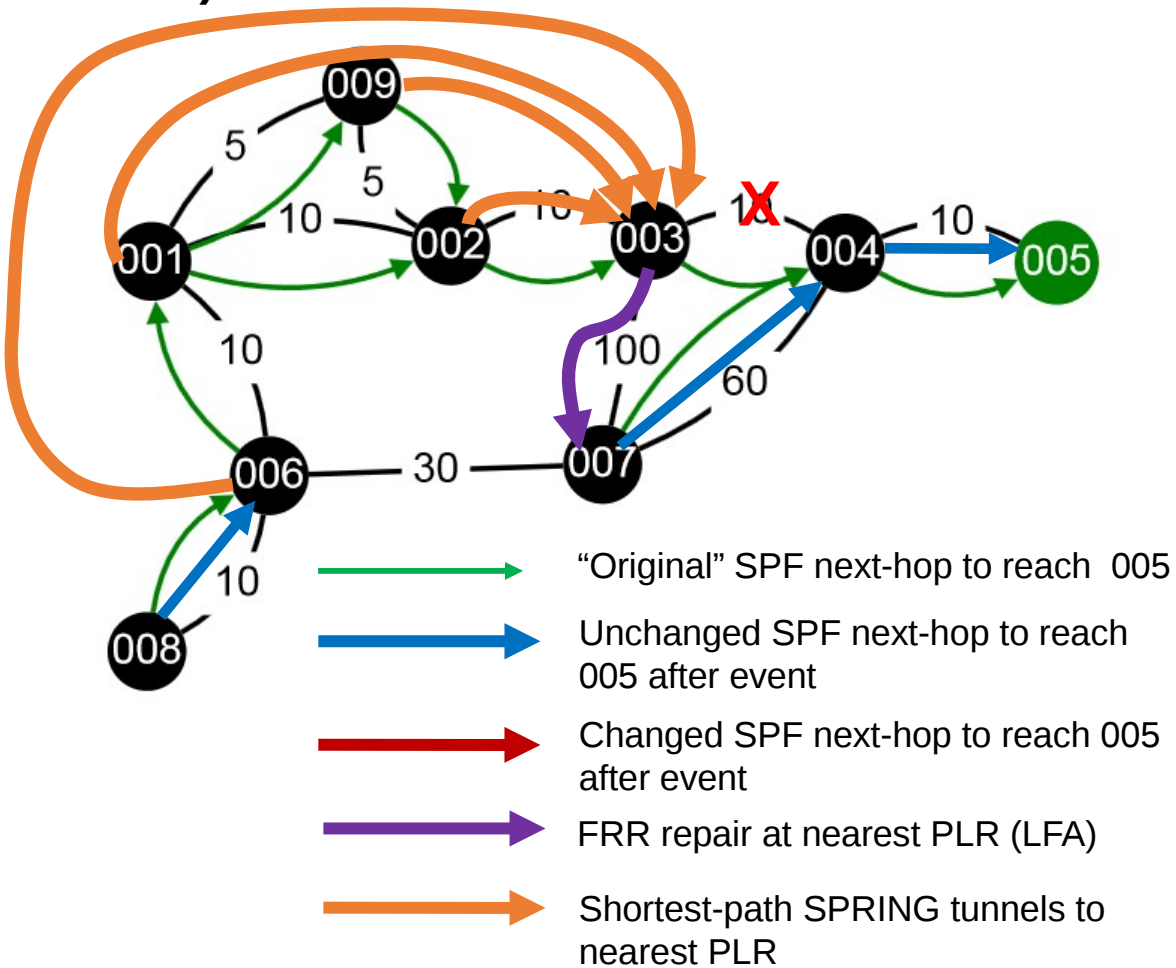


After failure

- **Shortest path to PLR is unaffected by failure for nodes that need to tunnel for that failure and hence microloop-free.**

Solution (near-end tunnelling)

D = MAX_CONVERGE_DELAY (network-wide)

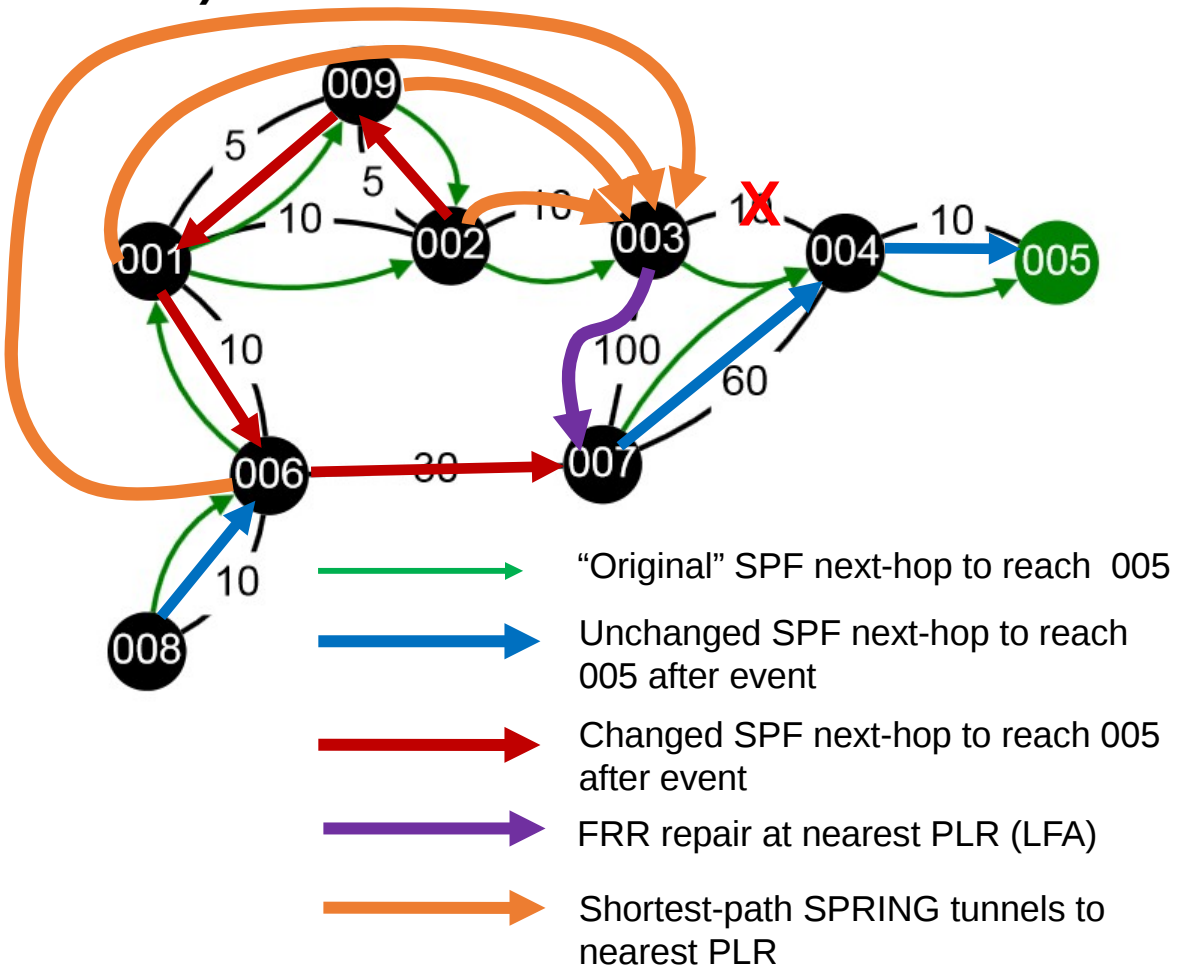


On node/link event(T0)

- On attached PLRs
  - FRR and delayed convergence
- On other routers, where nexthop to destination changed
  - Delay convergence to new SPF nexthops
  - Instead use 2 segment segment-list to tunnel traffic till all routers converge.
    - To nearest PLR (first segment)
    - From nearest PLR to final destination (next segment)

Solution using SPRING

D = MAX_CONVERGE_DELAY (network-wide)

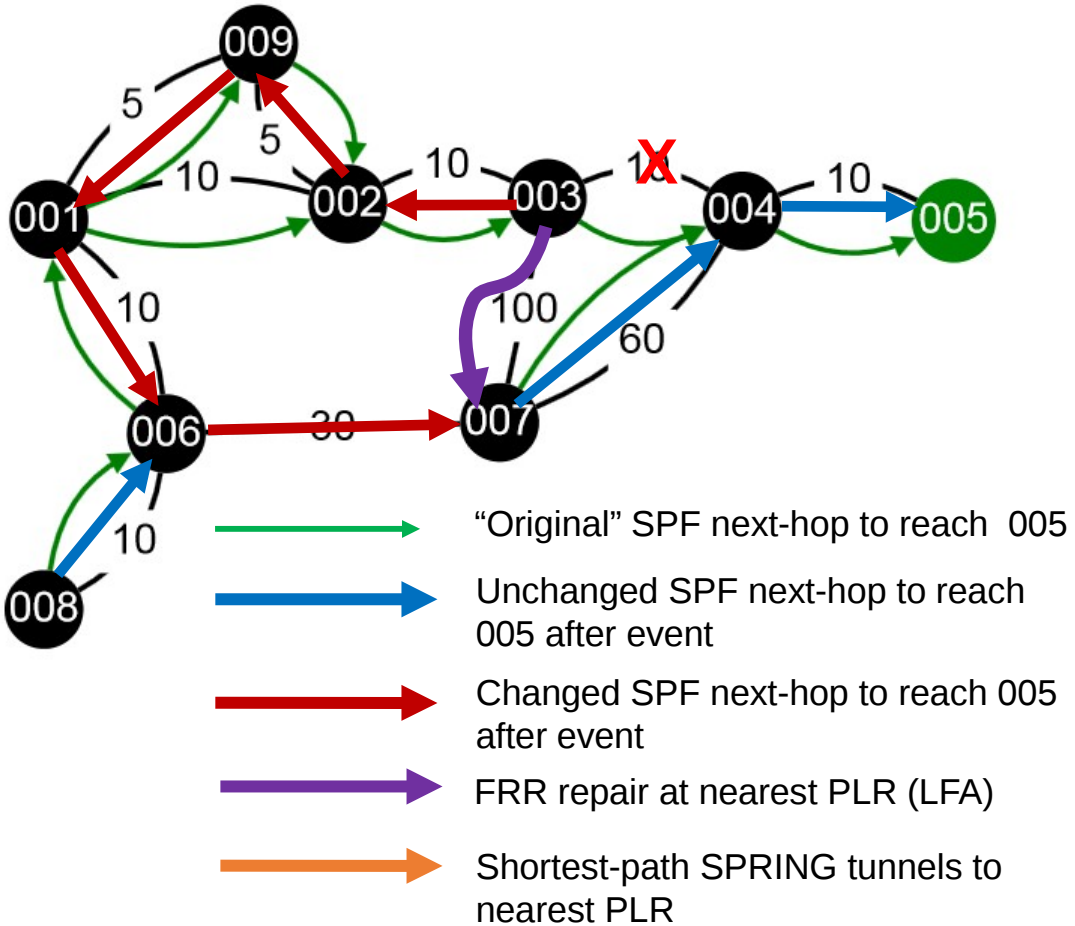


After time D (T1)

- On other routers, where nexthop to destination changed
  - Convergence to new SPF nexthops.

Solution using SPRING

D = MAX_CONVERGE_DELAY (network-wide)

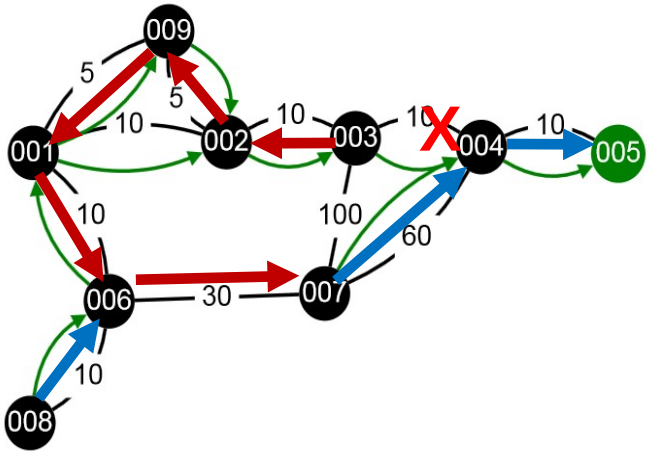


After time 2xD (T2)

- On PLR
  - Convergence to new SPF nexthops.

FIB table at various time intervals

Node	Before T0	T0-T1	T1-T2	After T2
001	Push 1005, Fwd to 002 Push 1005, Fwd to 009	Push **1005,**1003(top),Fwd to 003 Push **1005,**1003(top),Fwd to 009	Push **1005**, Fwd to 006	Push **1005**, Fwd to 006
008	Push 1005, Fwd to 006	Push 1005, Fwd to 006	Push **1005**, Fwd to 006	Push **1005**, Fwd to 006
003	Push 1005,Fwd to 004 *push **1005**, fwd to 007	*push **1005**, fwd to 007	*push **1005**, fwd to 007	push **1005**, fwd to 002



- Each node has SRGB range 1000-2000
- Each node is configured with index identified by node number

Procedures for various network events

- Link-down
- Link-up
- Metric increase
- Metric decrease
- Node-UP
- Node- Down
- SRLG failures

Handling Multiple events

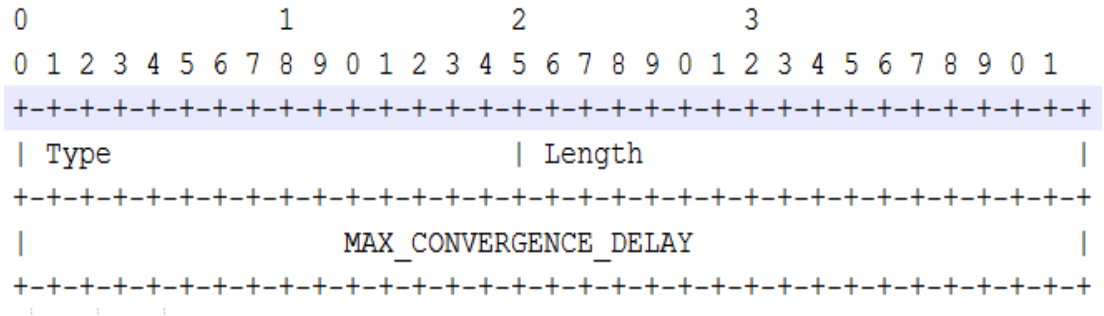
- Multiple network events which are not part of same SRLG are not handled and micro-loop prevention procedures are aborted
- Mechanisms to identify link-down/link-up events reported by both end points
- Mechanisms to identify node-down/node-up events reported by various neighbors of the node.

Partial Deployments

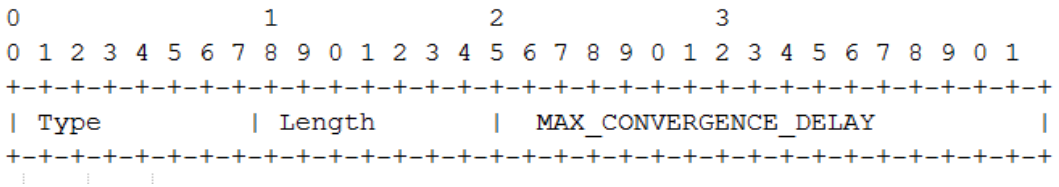
- All the nodes in the IGP flooding domain need to implement the micro-loop prevention procedures to work effectively
- Protocol extensions to advertise support of this feature.
- In some cases of partial deployment, traffic loss might increase if these procedures are followed by a few nodes and not followed by the PLR.

OSPF/ISIS Extensions

Micro-loop prevention TLV carried in RI-LSA in OSPF



Micro-loop prevention sub-TLV carried in RI- Capability TLV in ISIS



Next Steps

- Comments
- Suggestions

THANK YOU

Backup Slides

Micro-loop prevention procedures

- When microloop-prevention is enabled on a node,
  - The node is configured with **MAX_CONEVRGENCE_DELAY(D)**.
    - If not configured explicitly a good enough default value should be assumed.
  - The node should then advertise the capabilty in its IGP link-state advertisements along with the value of **MAX_CONVERGENCE_DELAY (D).**
  - The actual **value to be used on** is always derived from the **maximum value of MAX_CONVERGENCE_DELAYs** learnt across the entire IGP domain (learnt from all the nodes).

Micro-loop prevention procedures

LINK-DOWN Scenario

- At time T0,
  - PLRs,
    - Starts a timer T2 = 2 * MAX-CONEVRGENCE-DELAY.
    - Delays convergence and continues to use backup path.
  - Others, on receipt of the event, and finding change of next-hops for one or more destinations,
    - Starts a timer T1 = MAX-CONVERGENCE-DELAY.
    - Modifies the nexthop(s) for the affected destinations to tunnel the traffic to nearest PLR, using 2 segment stack.

Micro-loop prevention procedures

LINK-DOWN Scenario

- On Expiry of T1,
  - All nodes other than PLR(s)
    - Downloads the new SPF path(s).
      - Replaces the two-segment nexthop(s)
- On Expiry of T2,
  - PLR(s) stops using backup path(s)
    - Downloads new SPF path(s)