BGP Deterministic Path Forwarding (DPF)

draft-wang-idr-dpf

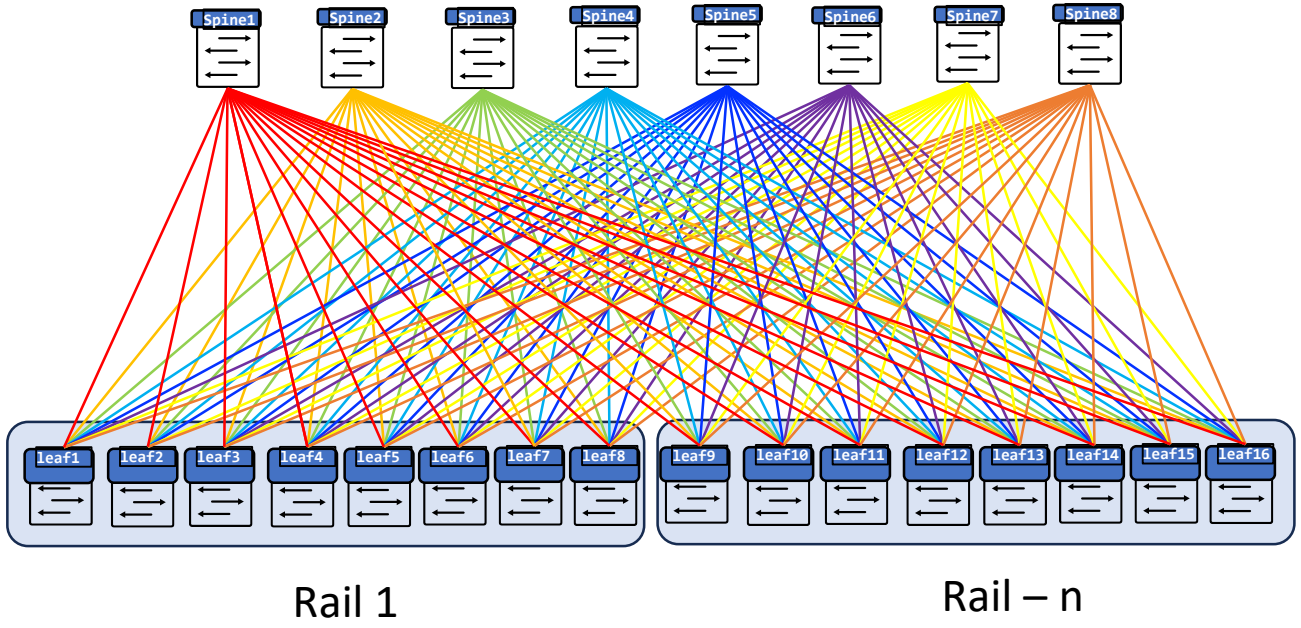
IETF 125, Mar 17th, 2026

Kevin Wang, Michal Styszynski, Wen Lin,

Mahesh Subramaniam, Thomas Kampa, Diptanshu Singh

BGP DPF Overview

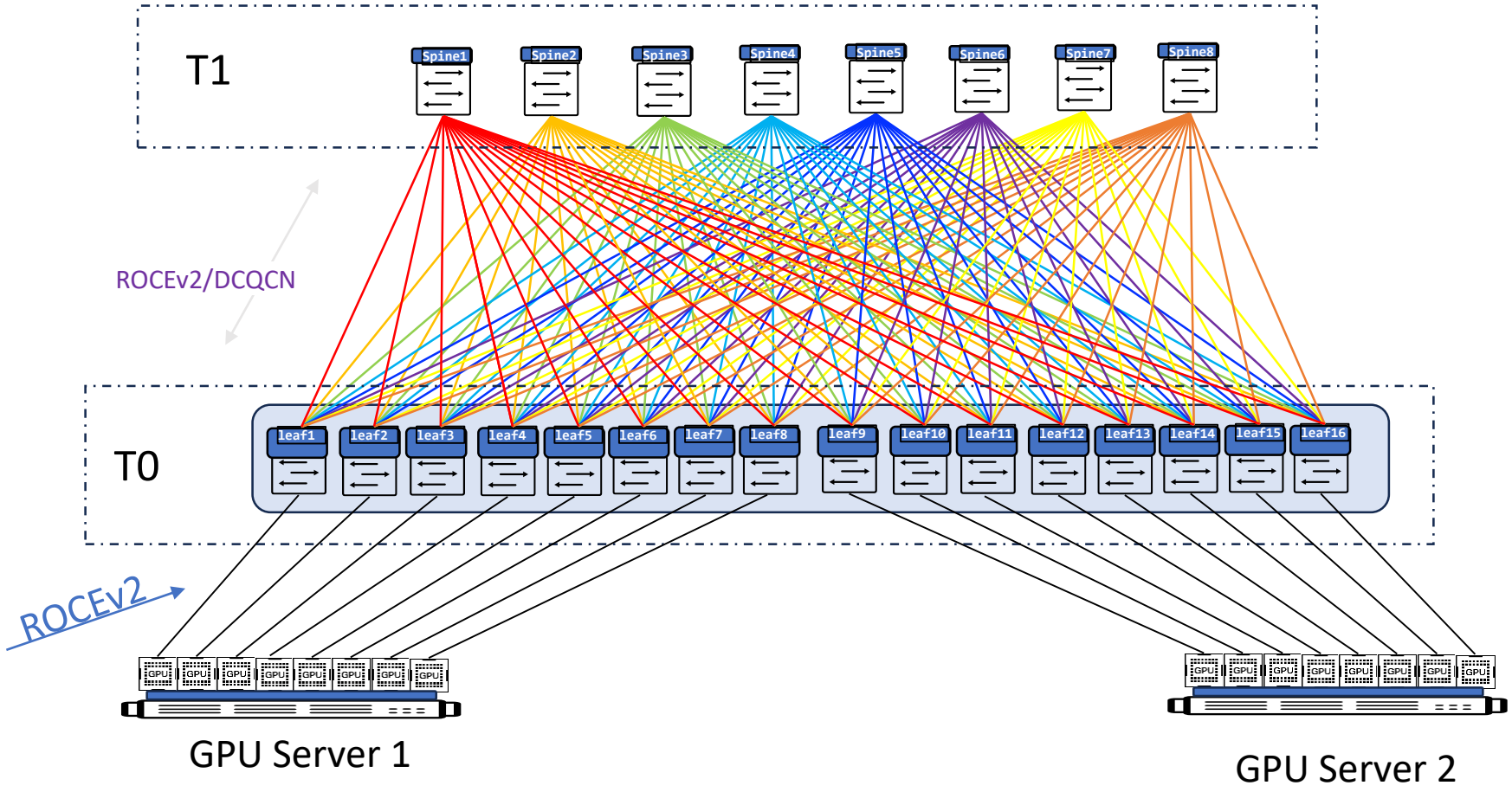
- Separate a physical IP fabric into multiple logical fabrics
- Flows can be mapped to different logical fabrics for:
  - Load balancing
  - Differentiated SLAs
  - No fate sharing
- A lightweight traffic engineering for IP fabrics



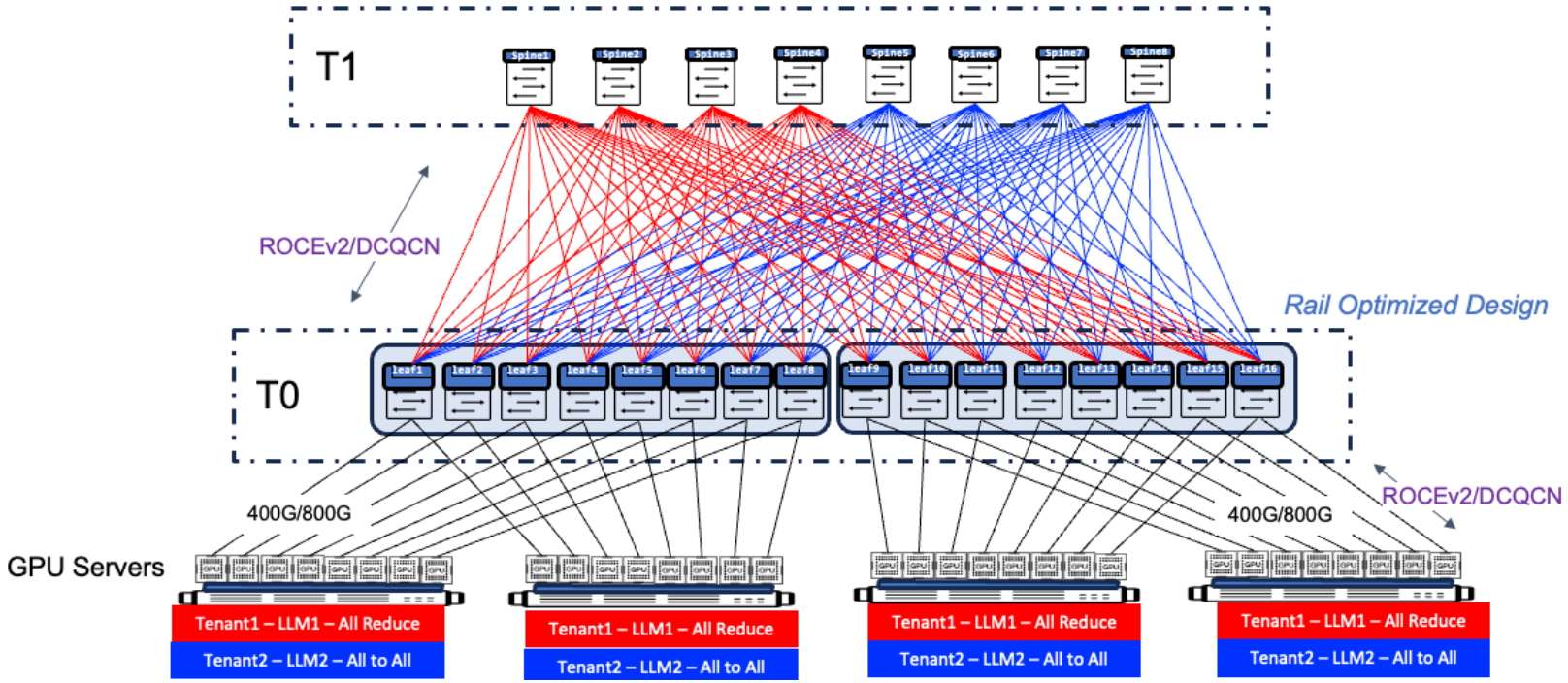
Use Case 1: Queue Pair Pinning (QPP)

- ROCEv2 BTH Header QPAIRs are pinned to different paths.

```
Ethernet II, Src: Performa_00:00:02 (00:10:94:00:00:02),
Internet Protocol Version 4, Src: 10.1.1.2, Dst: 5.1.1.1
User Datagram Protocol, Src Port: 1024, Dst Port: 4791
InfiniBand
  Base Transport Header
    Opcode: Reliable Connection (RC) - SEND First (0)
    0... .... = Solicited Event: False
    .0.. .... = MigReq: False
    ..00 .... = Pad Count: 0
    .... 0000 = Header Version: 0
    Partition Key: 65535
    Reserved: 00
    Destination Queue Pair: 0x000064
    0... .... = Acknowledge Request: False
    .000 0000 = Reserved (7 bits): 0
    Packet Sequence Number: 1
  Invariant CRC: 0x289532e9
Data (20 bytes)
```



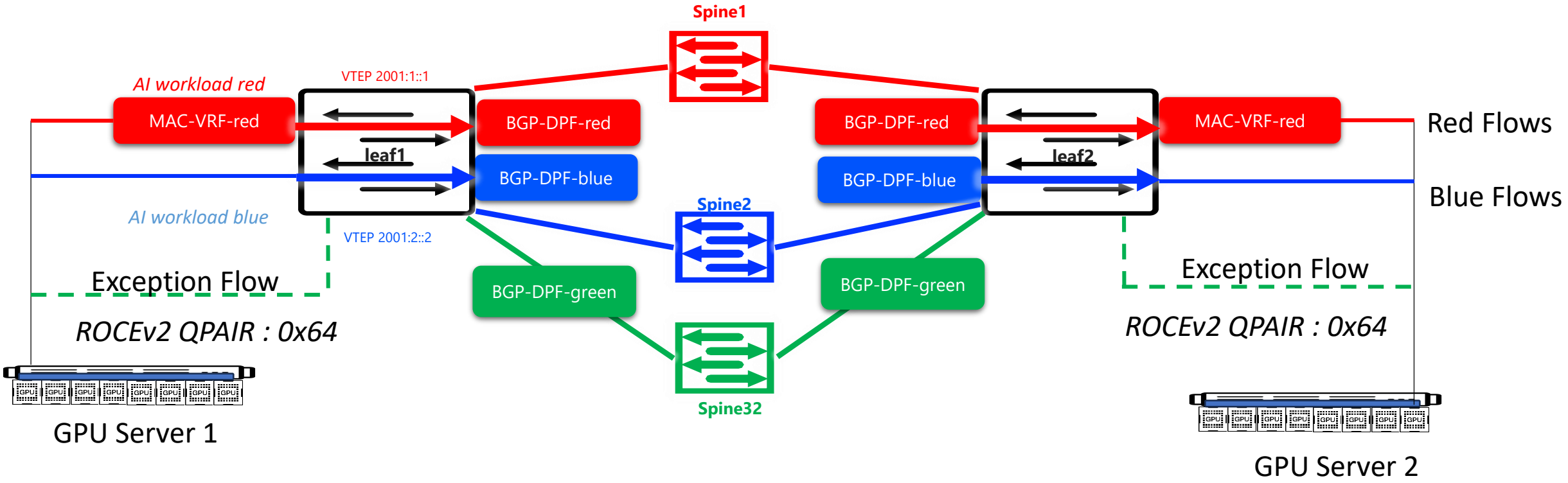
Use Case 2: Multi-tenancy for GPU as a Service



- Blue customer and red customer might have different SLA requirements
- Mapping them to different logical fabrics could help to separate their flows and allow them to have different SLAs

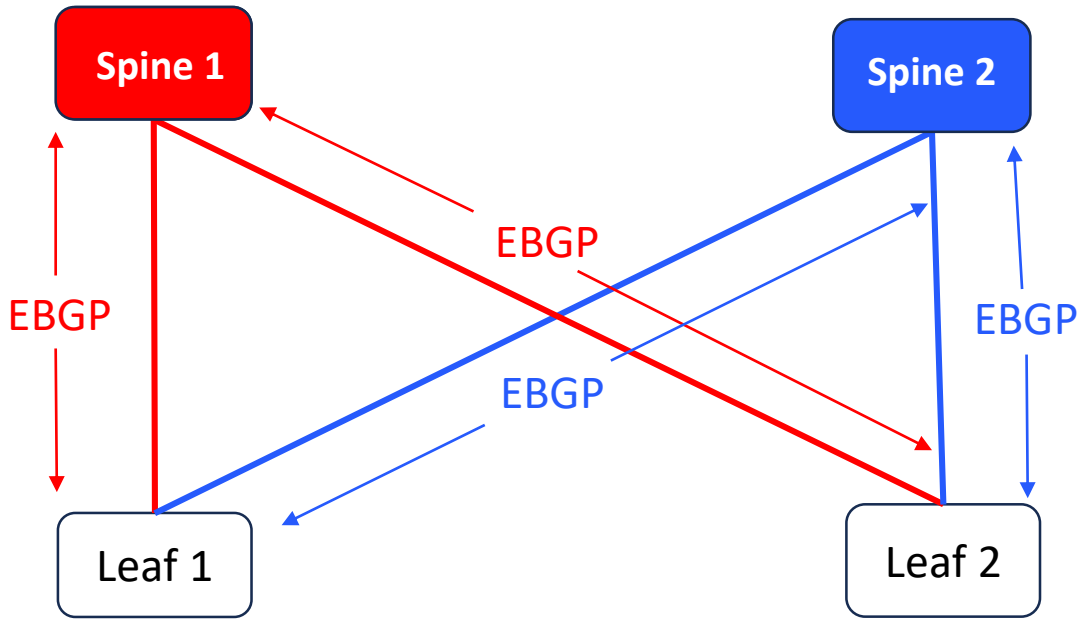
Use Case 3: Avoid Fate Sharing

- Exception flows can be mapped differently
- Redundant flows (PRP) can be mapped differently



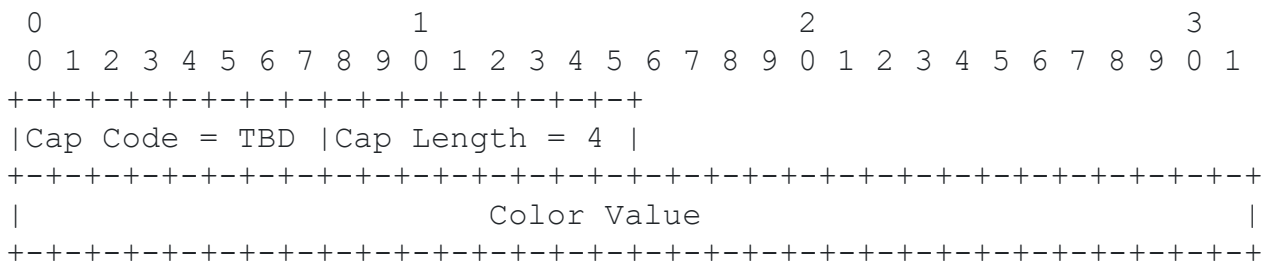
Color EBGP Sessions

- Each EBGP session is colored with a color community to represent the underlying link color
- A route can only be advertised/received over a colored session if it has the same color
- Uncolored routes match any colored sessions



Colored Session Modes

- Strict mode: uses SESSION-COLOR capability in OPEN message to negotiate session color
  - Pros: detects misconfiguration early
  - Cons: session flaps on color change
- Loose mode: don’t negotiate session color; routes will be rejected when received over a session with mismatched colors
  - Pros: allows color change without session flap
  - Cons: detects misconfiguration late



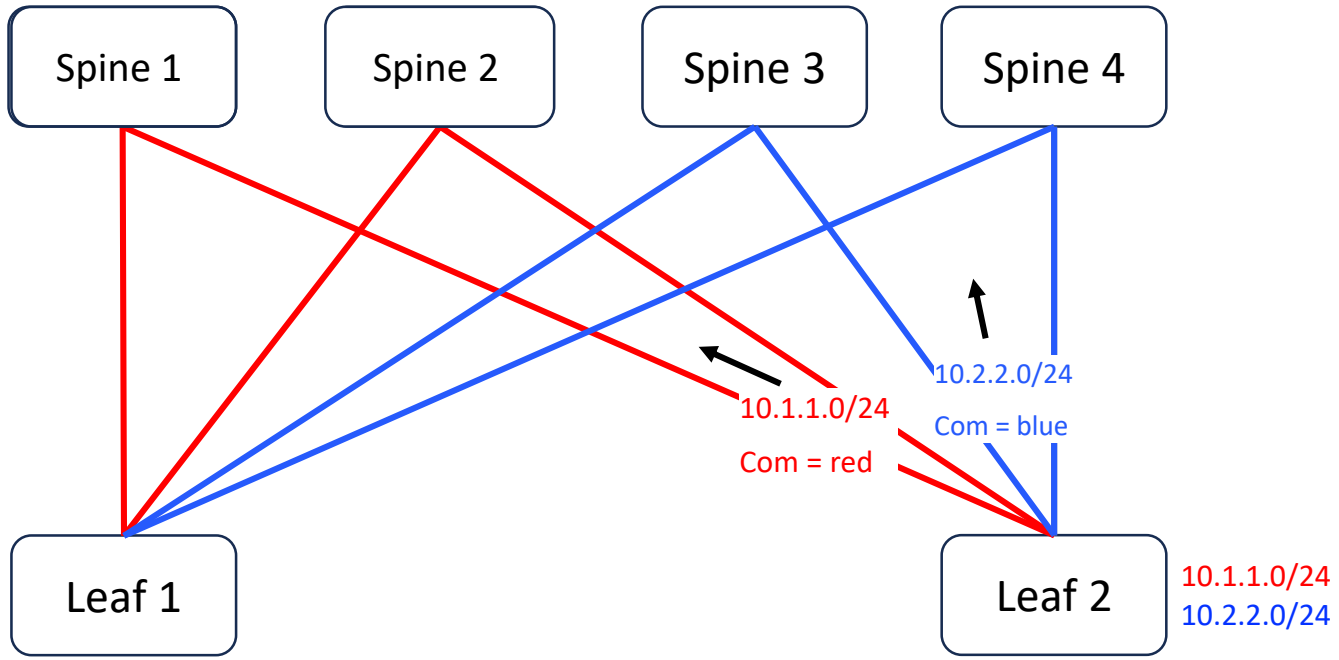
SESSION-COLOR Capability

Route Coloring: Primary Color Only

Leaf 1:

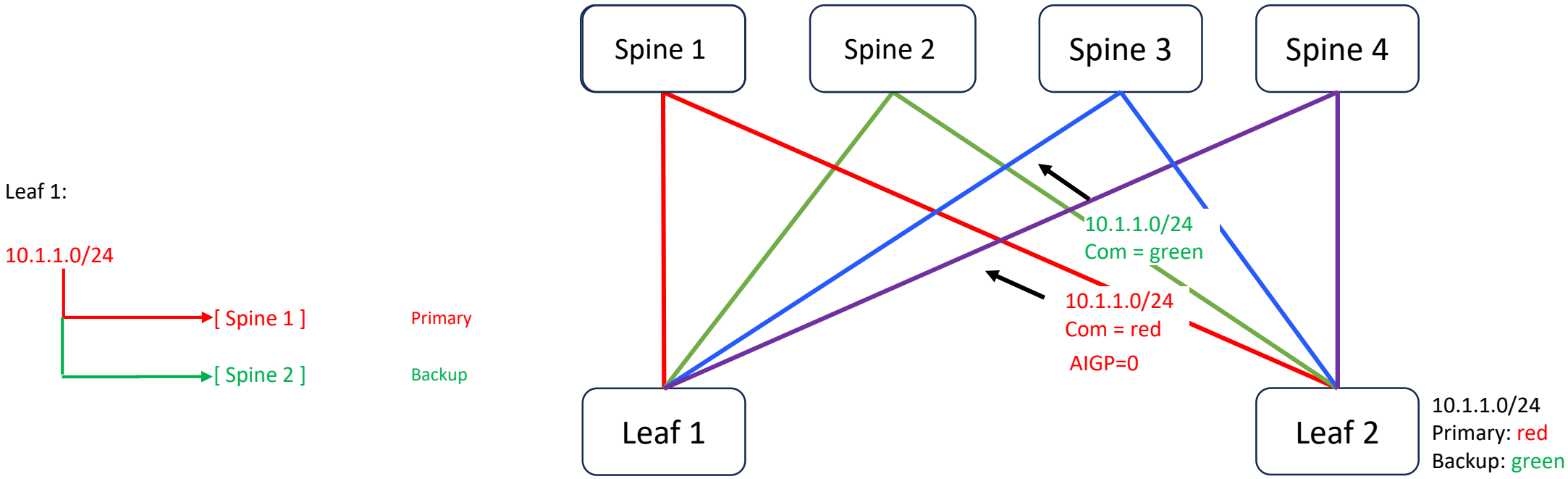
10.1.1.0/24 → [Spine 1, Spine 2]

10.2.2.0/24 → [Spine 3, Spine 4]



- Leaf2 colors 10.1.1.0/24 with red community and 10.2.2.0/24 with blue community
- Red route is routed exclusively in the red fabric and blue route is routed exclusively in the blue fabric
- Leaf1 has ECMP of 2 red paths for 10.1.1.0/24 and ECMP of 2 blue paths for 10.2.2.0/24

Route Coloring: Primary + Backup Colors



- Leaf2 colors 10.1.1.0/24 with red as primary color and green as backup color
- Route 10.1.1.0/24 is advertised to the red fabric with red community and to the green fabric with green community
- AIGP is attached when advertising over red fabric
- Leaf1 has red path as primary and green path as backup

Route Coloring: Primary + All Color Backups

Leaf 1:

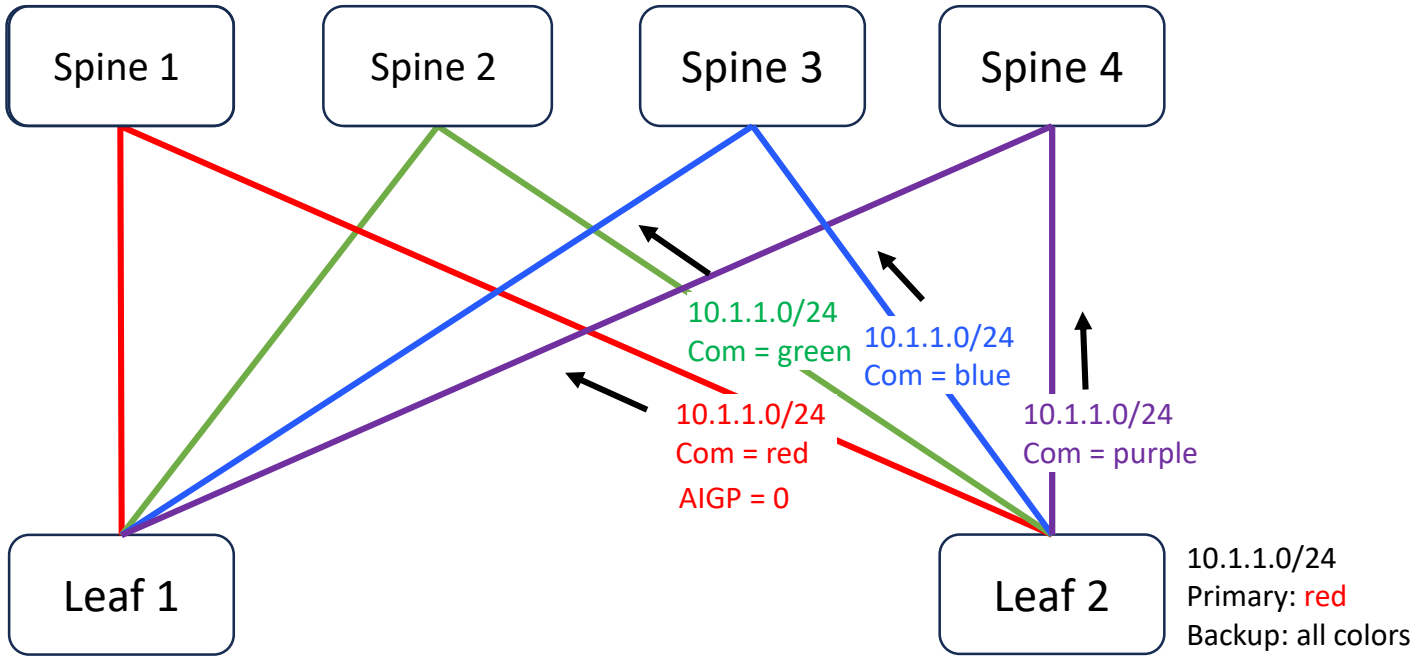
10.1.1.0/24

[Spine 1] Primary

[Spine 2, Backup

Spine 3,

Spine 4]



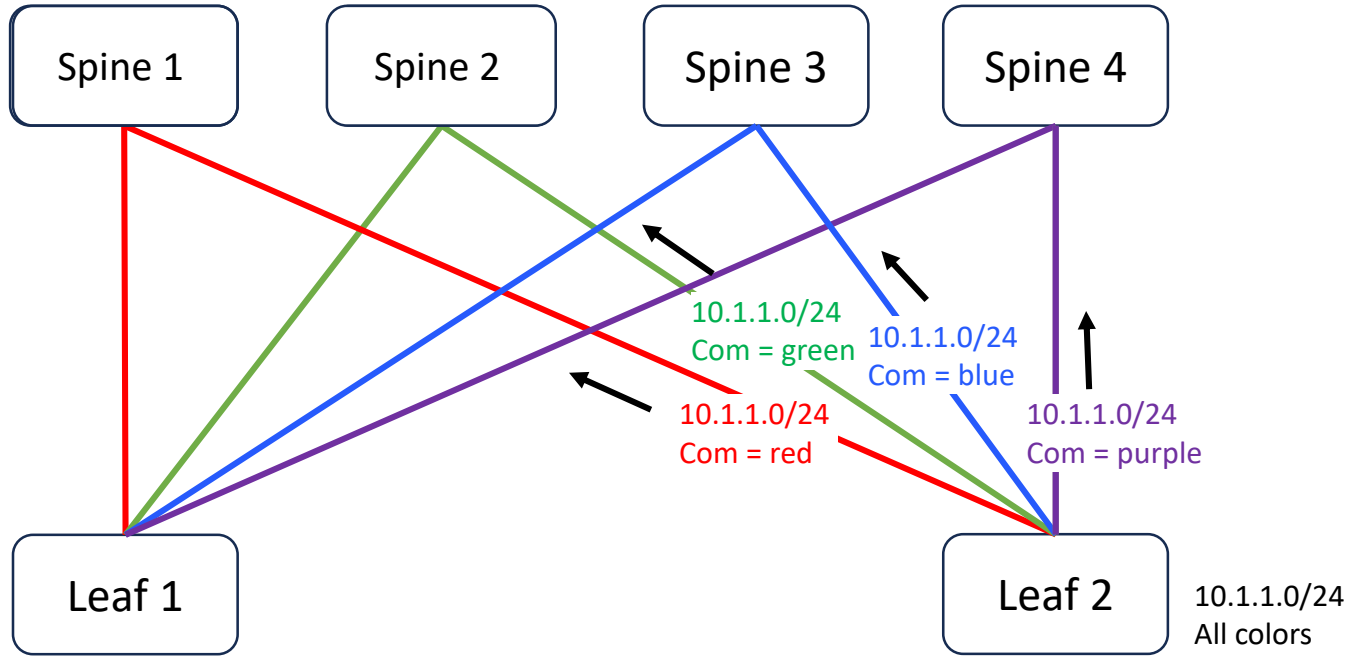
- Leaf2 colors 10.1.1.0/24 with red as primary color and all other colors as backup
- Route 10.1.1.0/24 is advertised to each fabric with the corresponding color community
- AIGP is attached when advertising over red fabric
- Leaf1 has red as primary and ECMP of green, blue, purple as backup. This mode can be used for QPP

Route Coloring: All Colors

Leaf 1:

10.1.1.0/24

[Spine 1, Spine2, Spine3, Spine4]



- Leaf2 colors 10.1.1.0/24 with all colors
- Route 10.1.1.0/24 is advertised to each fabric with the corresponding color community
- ECMP is formed for all colors at Leaf1 and firewall filter is used to map flows to different colored next-hops

Thank you!