Advanced Border Gateway Protocol

Gustavo Sibaja
Systems Architect, CCIE #52660
@GustavoSibaja4

Peter Palúch
Learning with Cisco, CCIE #23527
@Peter_Paluch

BRKENT-3219

Cisco Webex App

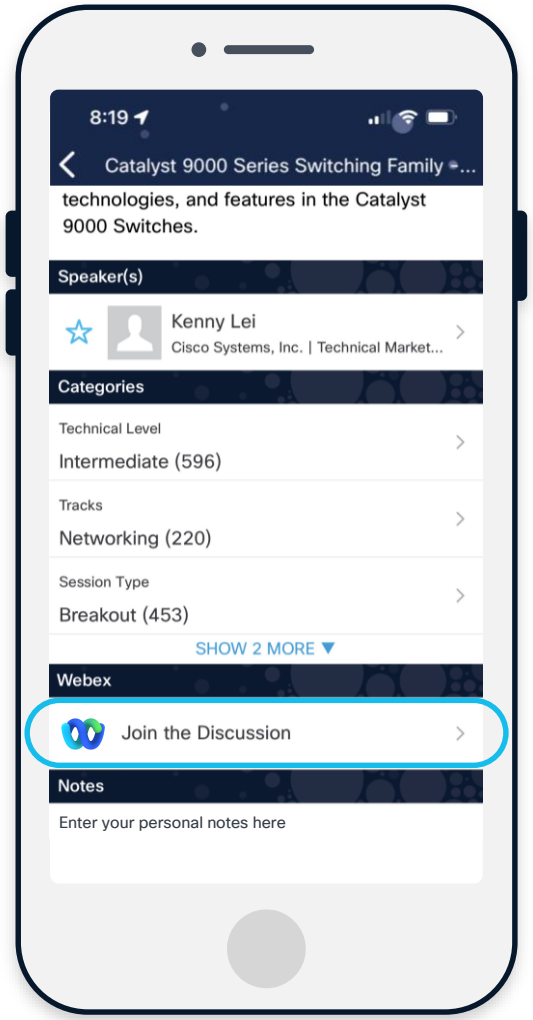
Questions?

Use Cisco Webex App to chat
with the speaker after the session

How

1. Find this session in the Cisco Live Mobile App
2. Click “Join the Discussion”
3. Install the Webex App or go directly to the Webex space
4. Enter messages/questions in the Webex space

Webex spaces will be moderated by the speaker until June 13, 2025.



Agenda

01 Introduction

02 Protocol Extensibility

03 Attributes

04 Security

05 Scalability

06 Conclusion

Introduction

BGP is over 30 years old!

- BGPv4 came around 1994 and stayed with us ever since
- A testament to its flexibility, scalability, robustness, extensibility
- Dr. Yakov Rekhter (father of BGP) called his approach “design by pragmatism”
- This session is a tribute to that pragmatism and the protocol BGP came to be over the 30 years

What we are going to talk about...

- In the session, we will be looking into select advanced BGP topics
- For housekeeping purposes, we split them up into categories
  - Protocol Extensibility
  - Attributes
  - Security
  - Scalability
- These categories should not be taken as absolute
  - Many features could fall into multiple categories at the same time

Protocol Extensibility

Address Families

Multi Protocol BGP and Address Families

- First brought by RFC 2283 (now RFC 4760), BGP was extended with multi-protocol capability
  - Ability to advertise reachability information for different address families
- The address family is identified by two values
  - Address Family Identifier (AFI) – the fundamental address family
  - Subsequent Address Family Identifier (SAFI) – its particular use
- This extension skyrocketed BGP's flexibility and adaptability for various applications and use cases

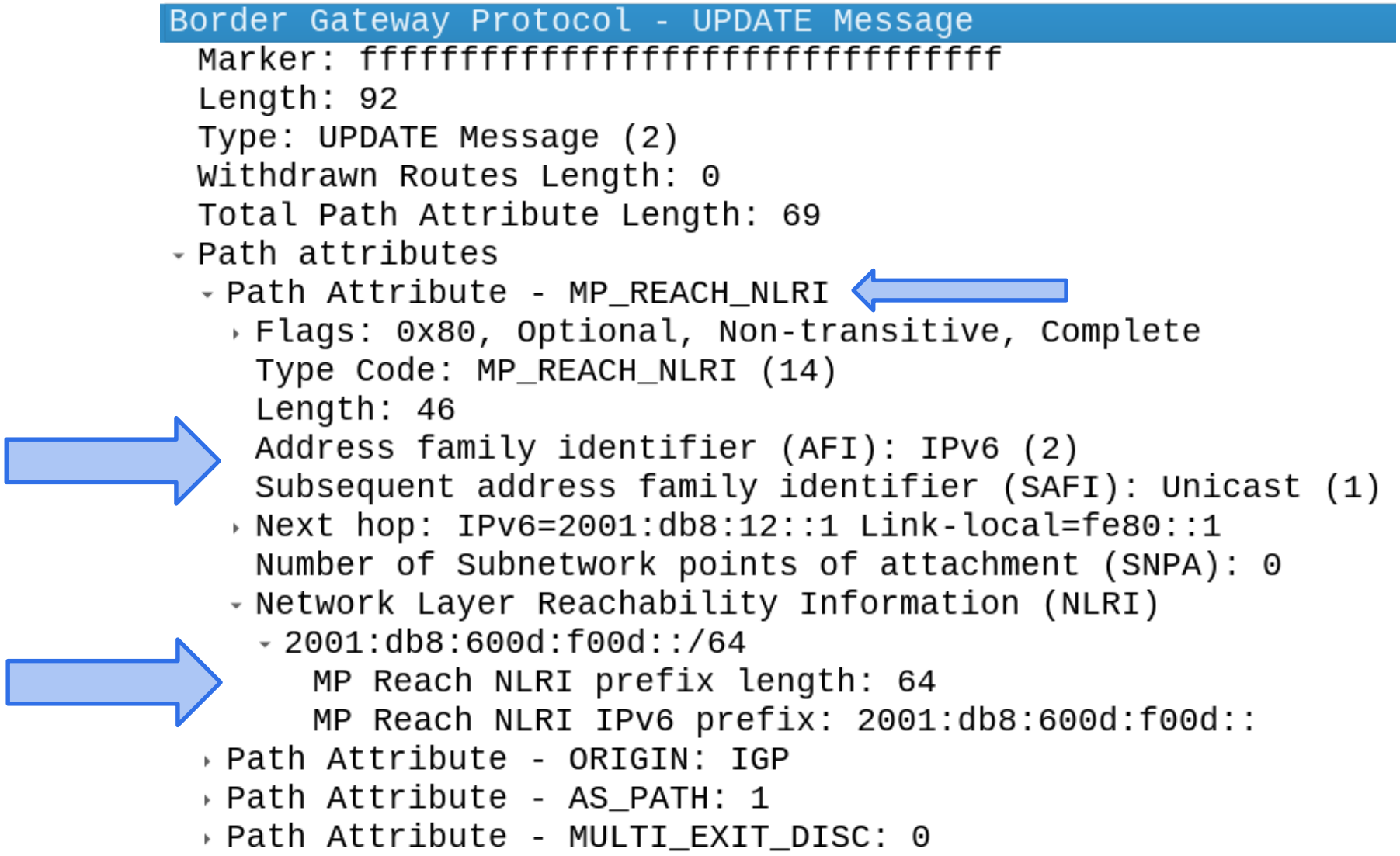
Negotiating supported address families

```
Border Gateway Protocol - OPEN Message
  Marker: ffffffffffffffffffffffffffffffff
  Length: 65
  Type: OPEN Message (1)
  Version: 4
  My AS: 1
  Hold Time: 180
  BGP Identifier: 10.0.12.1
  Optional Parameters Length: 36
▾ Optional Parameters
  ▾ Optional Parameter: Capability
      Parameter Type: Capability (2)
      Parameter Length: 6
    ▾ Capability: Multiprotocol extensions capability
        Type: Multiprotocol extensions capability (1)
        Length: 4
        AFI: IPv4 (1)
        Reserved: 00
        SAFI: Labeled VPN Unicast (128)
  ▾ Optional Parameter: Capability
      Parameter Type: Capability (2)
      Parameter Length: 6
    ▾ Capability: Multiprotocol extensions capability
        Type: Multiprotocol extensions capability (1)
        Length: 4
        AFI: IPv4 (1)
        Reserved: 00
        SAFI: Unicast (1)
```

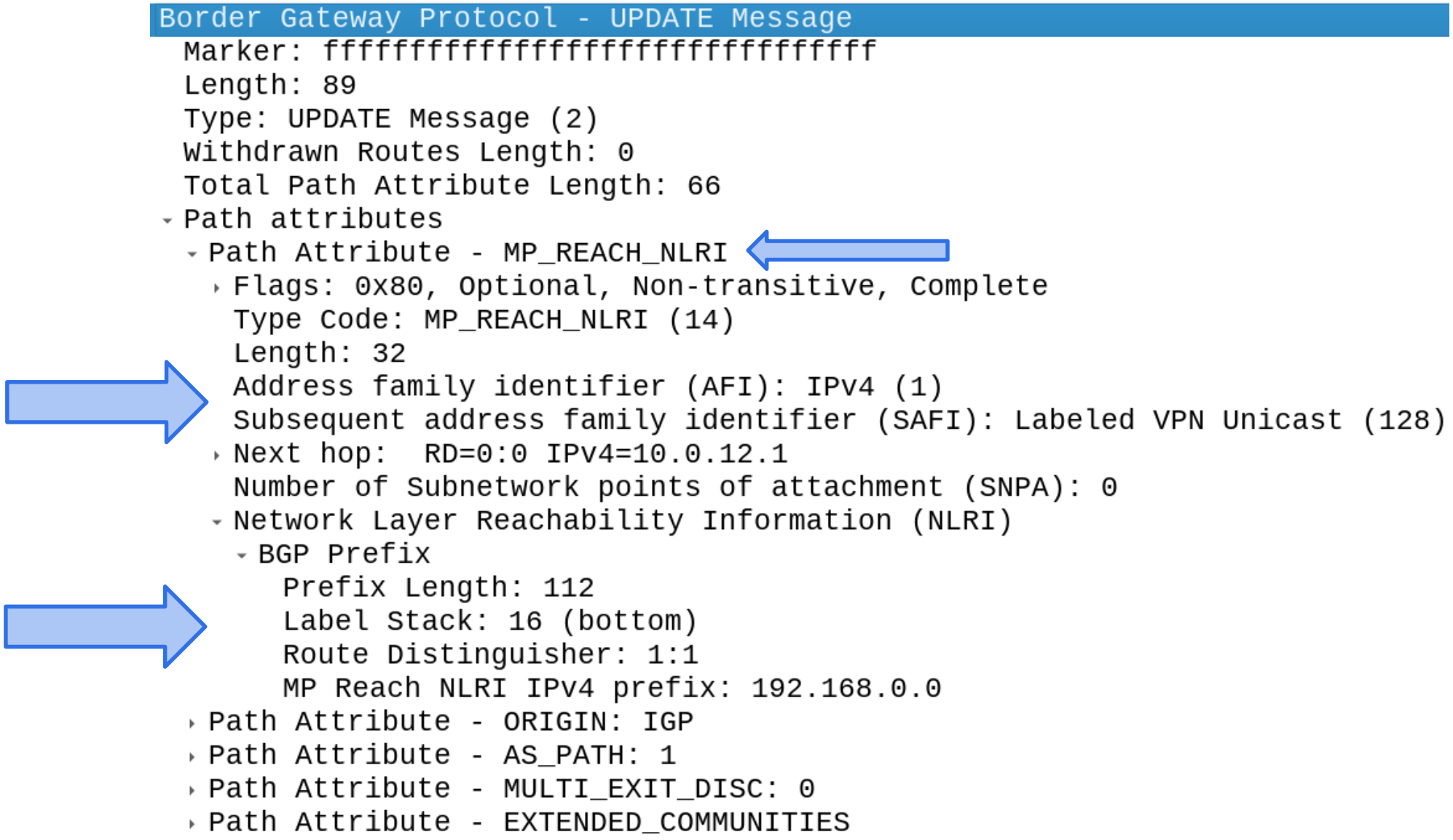
How does a “classic” UPDATE look like...

```
Border Gateway Protocol - UPDATE Message
 Marker: ffffffffffffffffffffffffffffffff
 Length: 67
 Type: UPDATE Message (2)
 Withdrawn Routes Length: 0
 Total Path Attribute Length: 28
▾Path attributes
 ▸Path Attribute - ORIGIN: IGP
 ▸Path Attribute - AS_PATH: empty
 ▸Path Attribute - NEXT_HOP: 10.255.255.1
 ▸Path Attribute - MULTI_EXIT_DISC: 1234
 ▸Path Attribute - LOCAL_PREF: 100
▾Network Layer Reachability Information (NLRI)
 ▸192.168.0.0/24
 ▸192.168.1.0/24
 ▸192.168.2.0/24
 ▸192.168.3.0/24
```

... and how does an MP-BGP UPDATE look like ☺

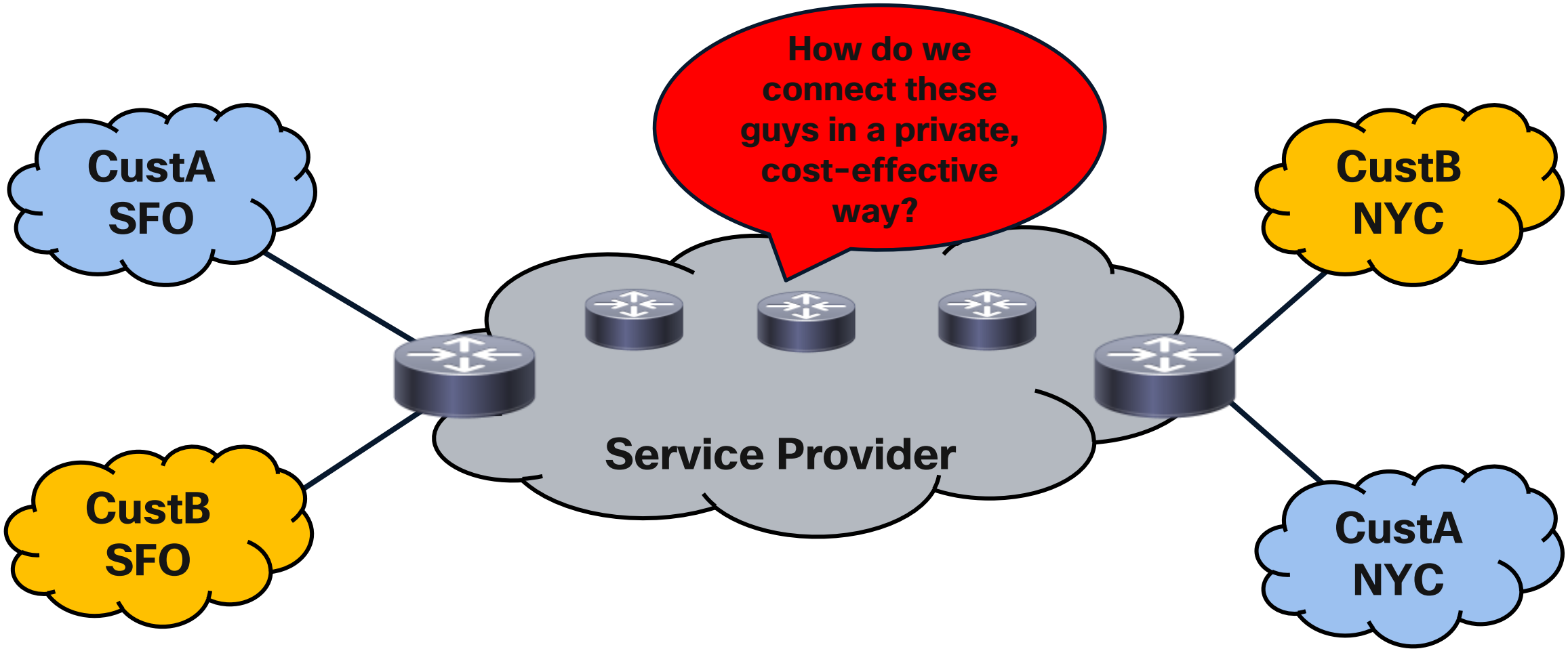


… and how does an MP-BGP UPDATE look like ☺

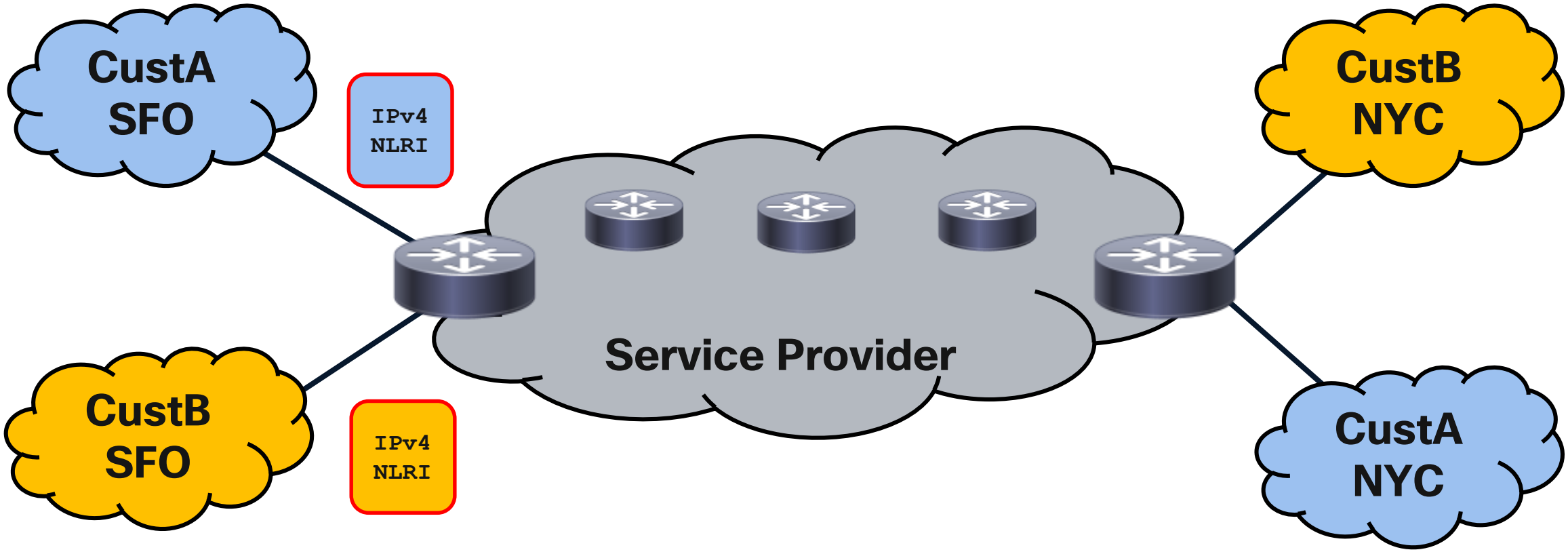


Use of Address Families and Attributes in MPLS L3 VPNs

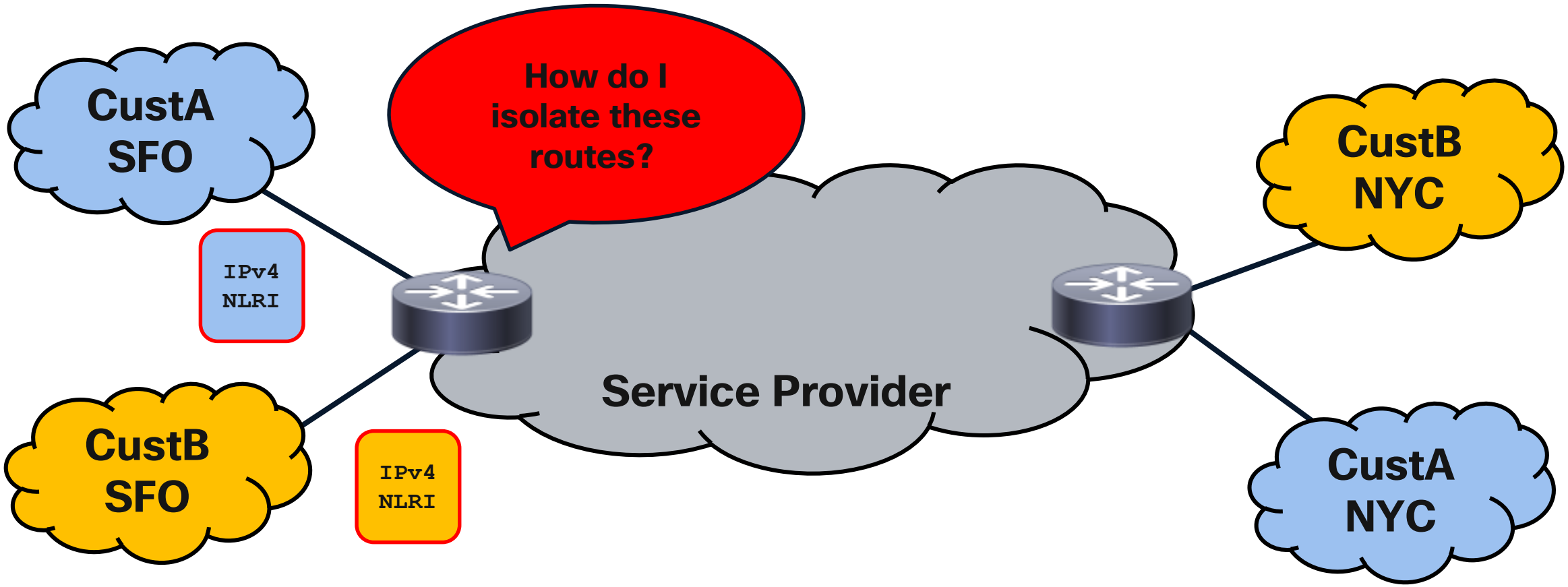
MPLS Layer 3 VPNs (L3VPN)



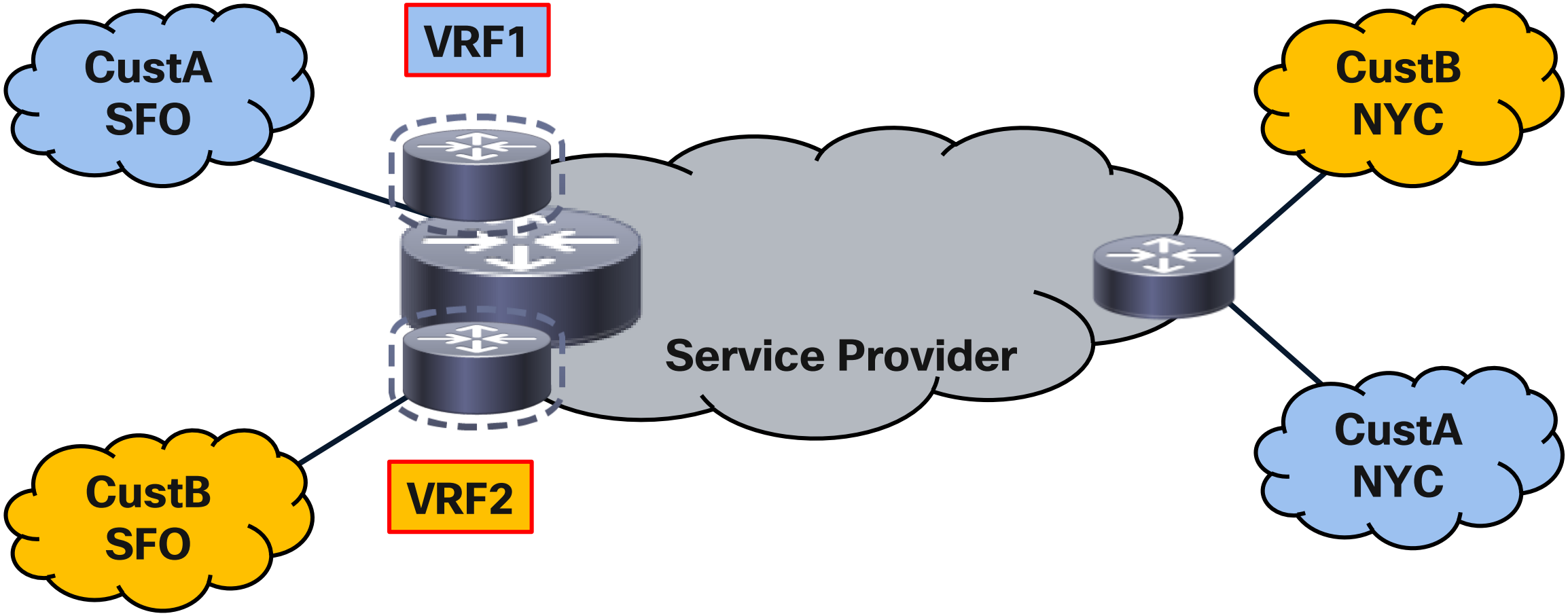
MPLS Layer 3 VPNs (L3VPN)



MPLS Layer 3 VPNs (L3VPN)

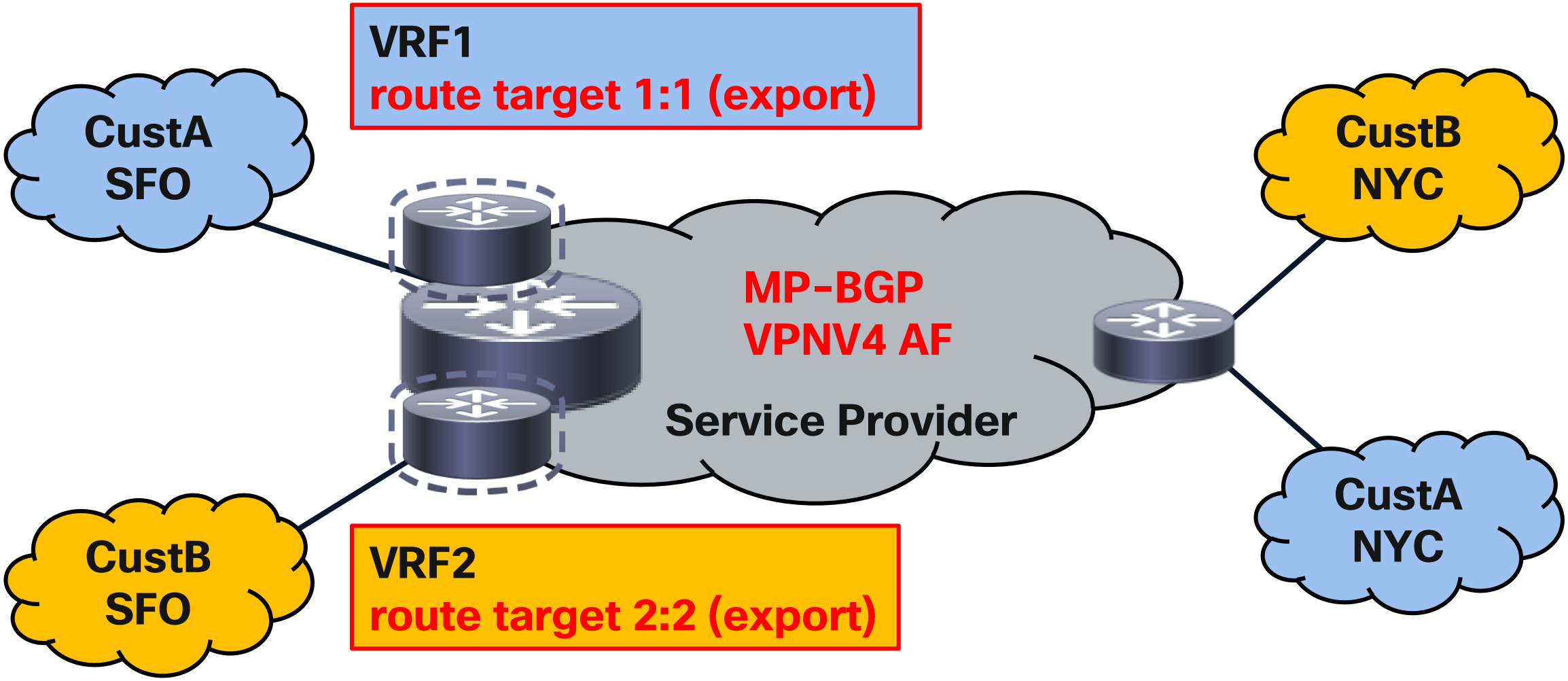


MPLS Layer 3 VPNs (L3VPN)

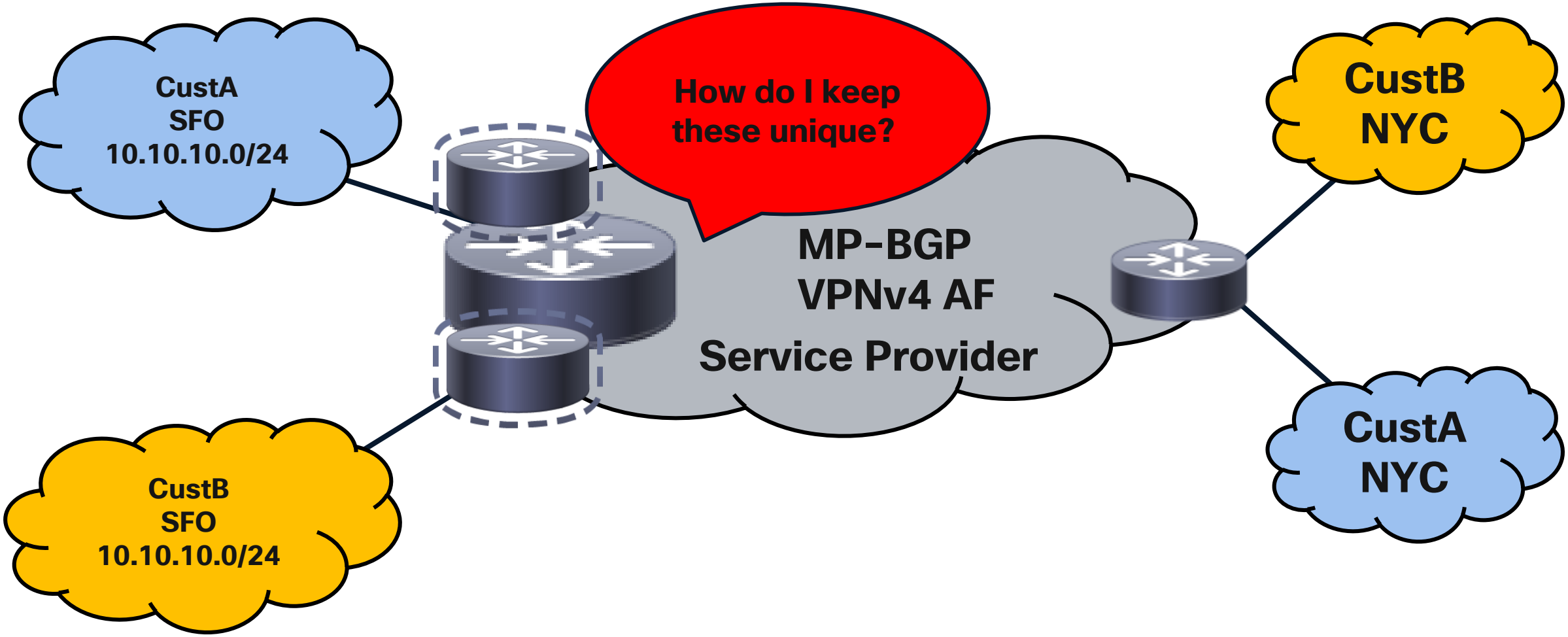


MPLS Layer 3 VPNs (L3VPN)

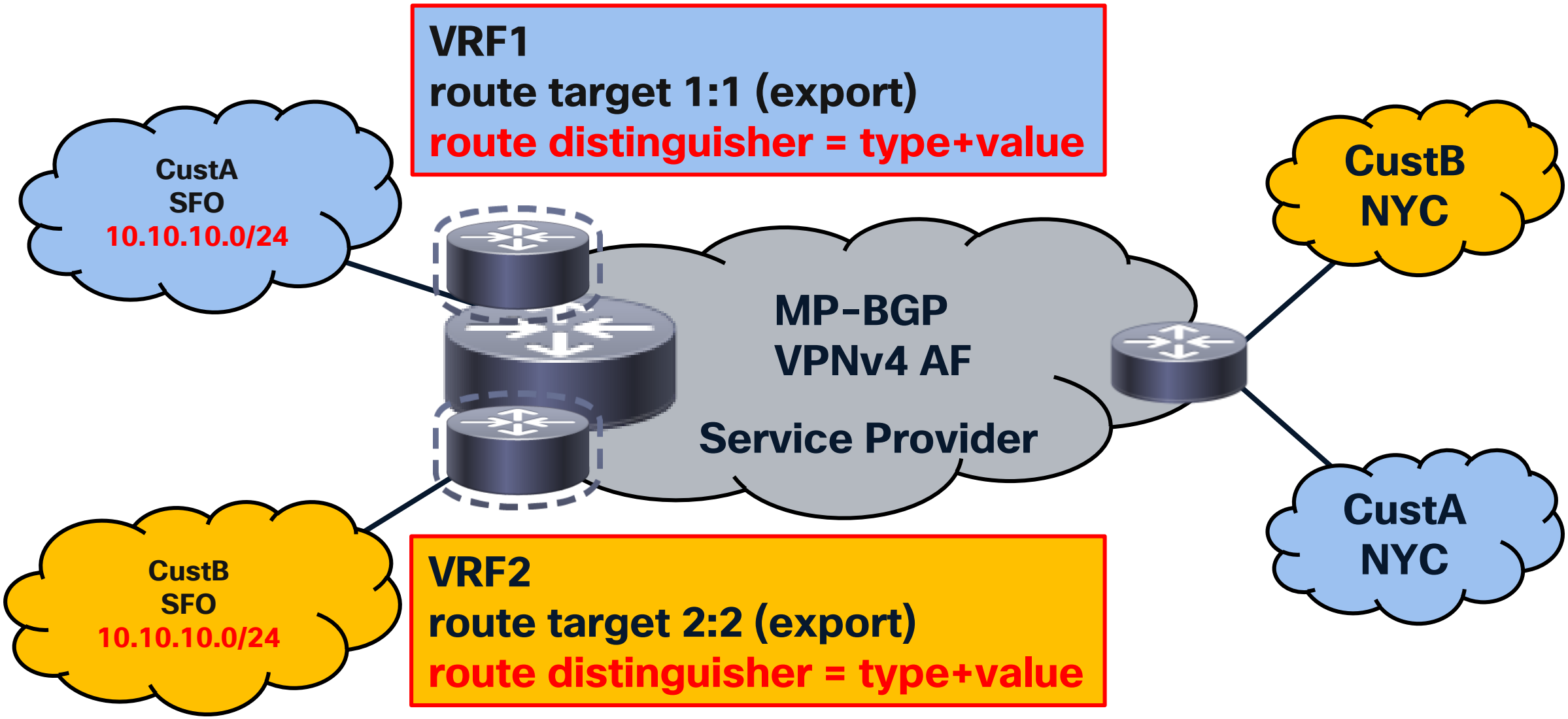
MPLS Layer 3 VPNs (L3VPN)



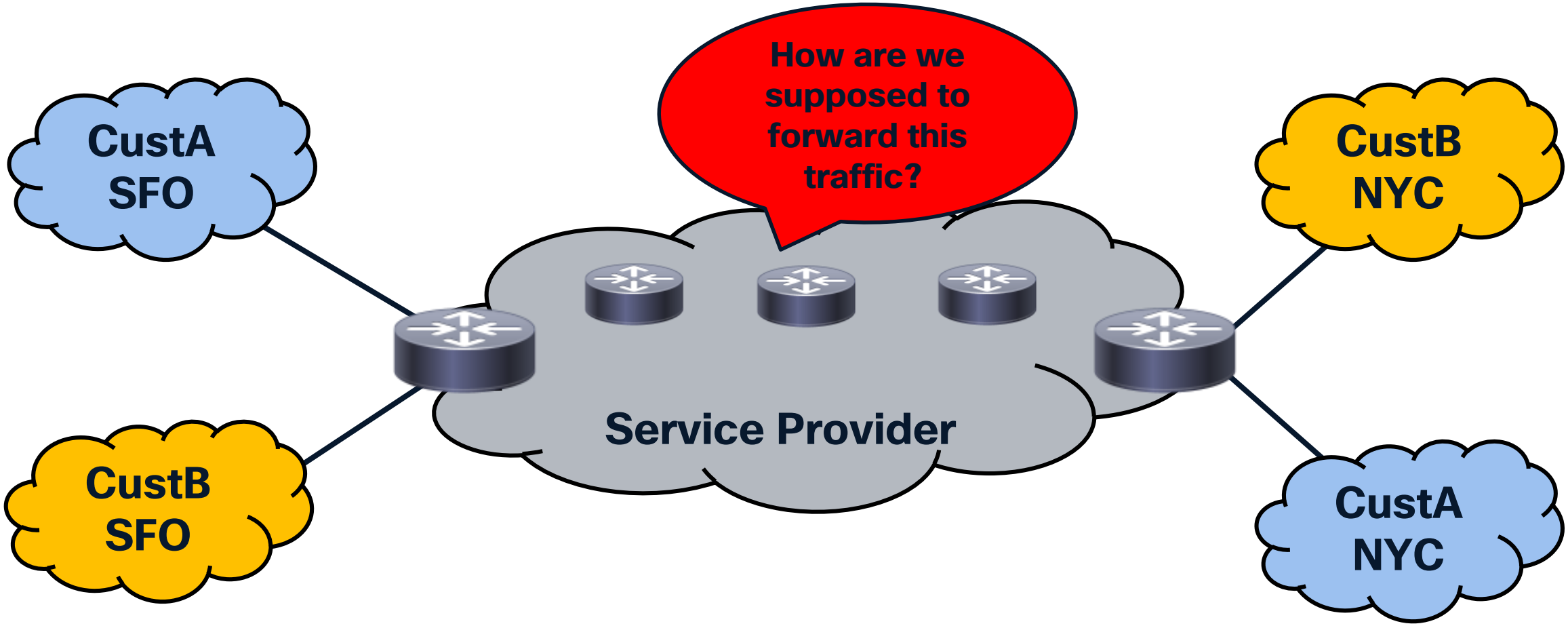
MPLS Layer 3 VPNs (L3VPN)



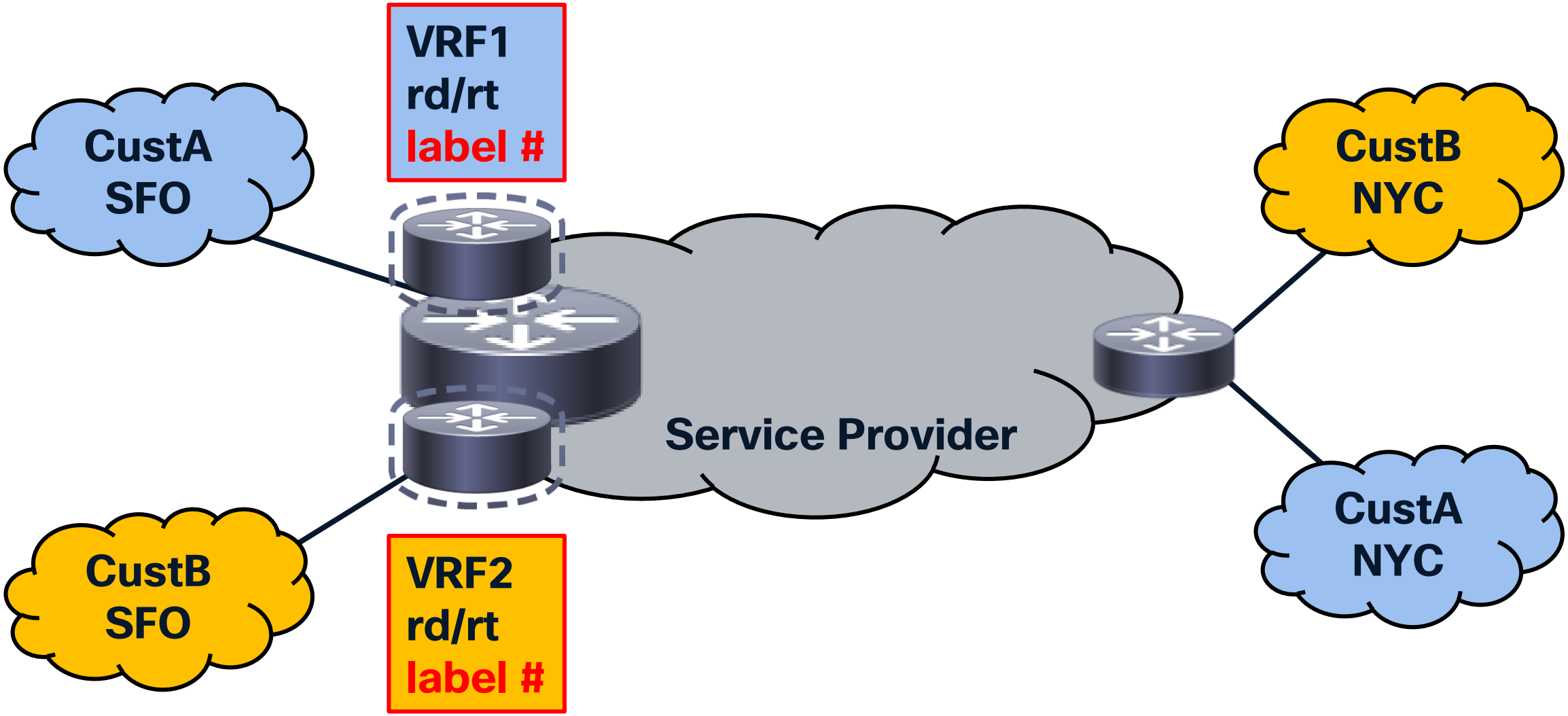
MPLS Layer 3 VPNs (L3VPN)



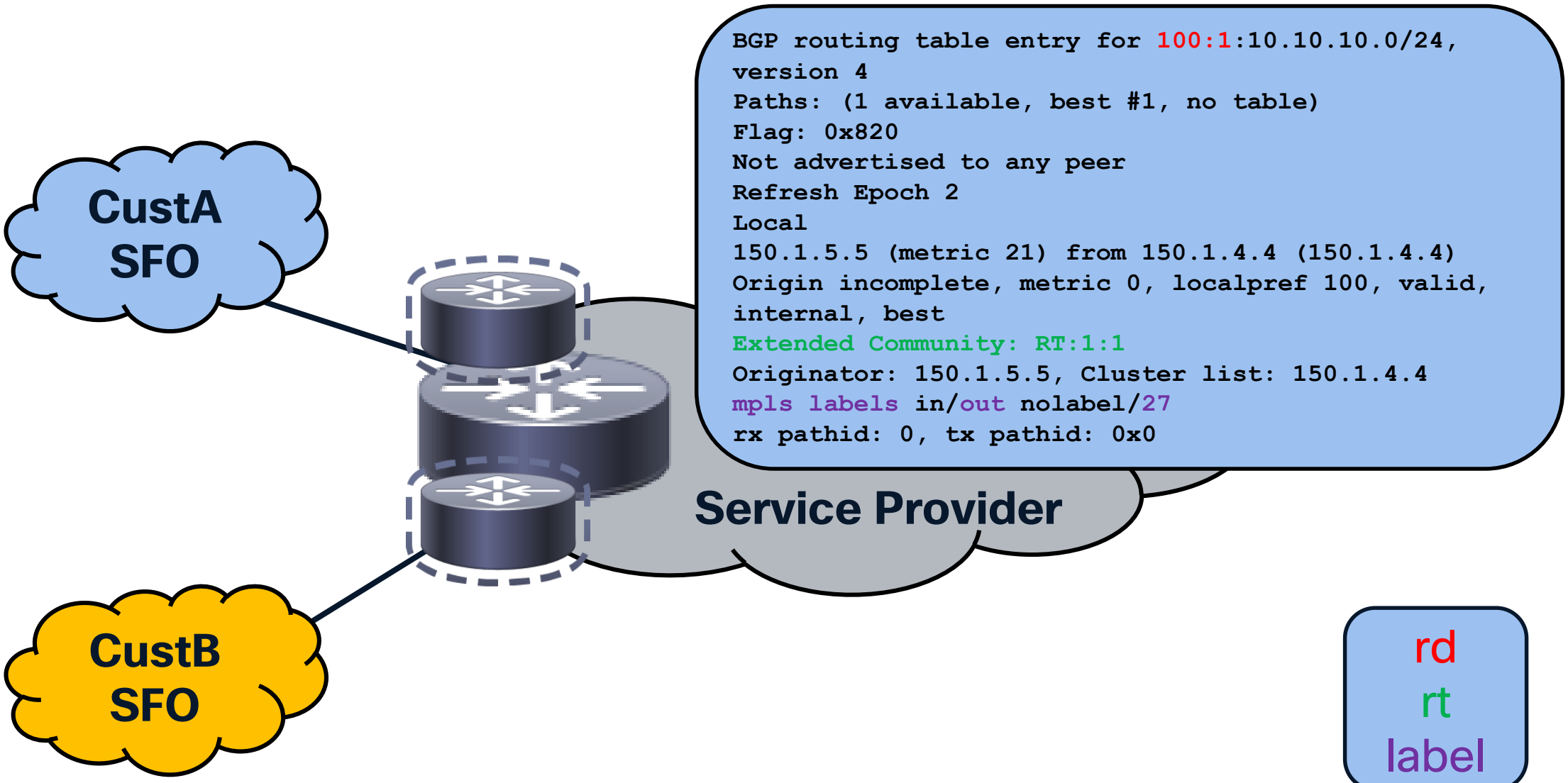
MPLS Layer 3 VPNs (L3VPN)



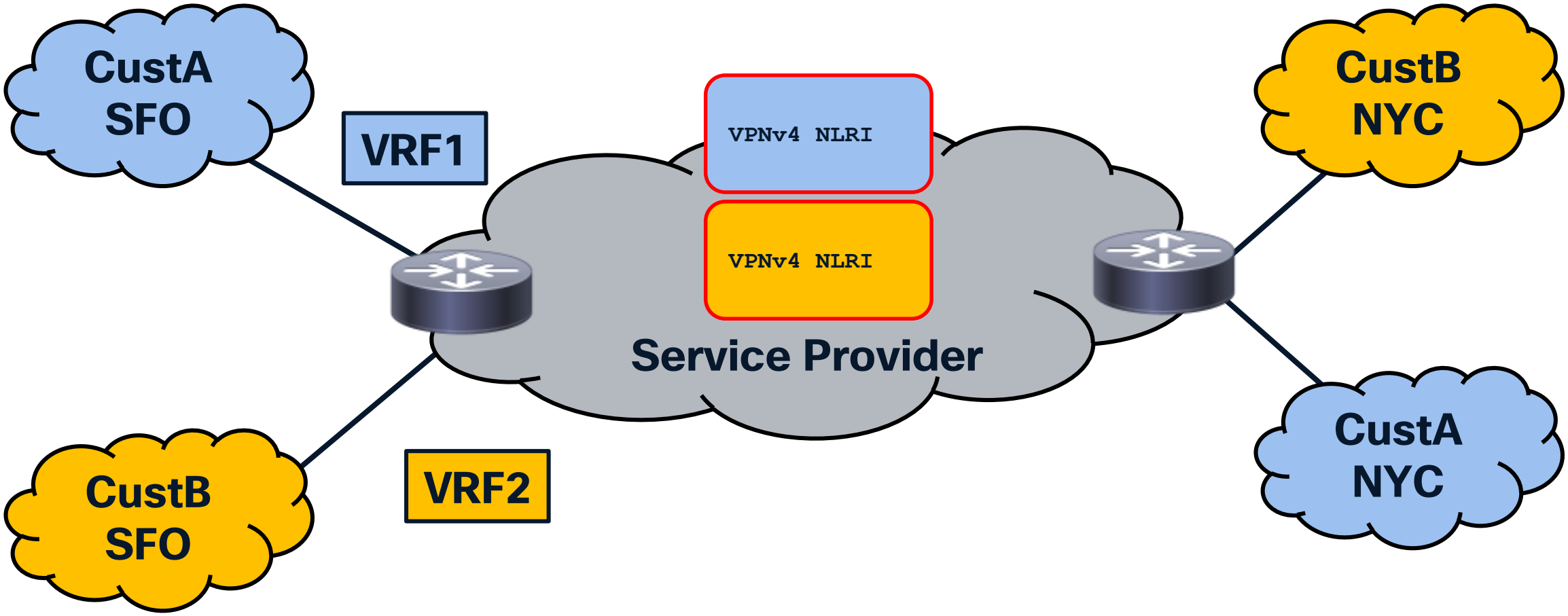
MPLS Layer 3 VPNs (L3VPN)



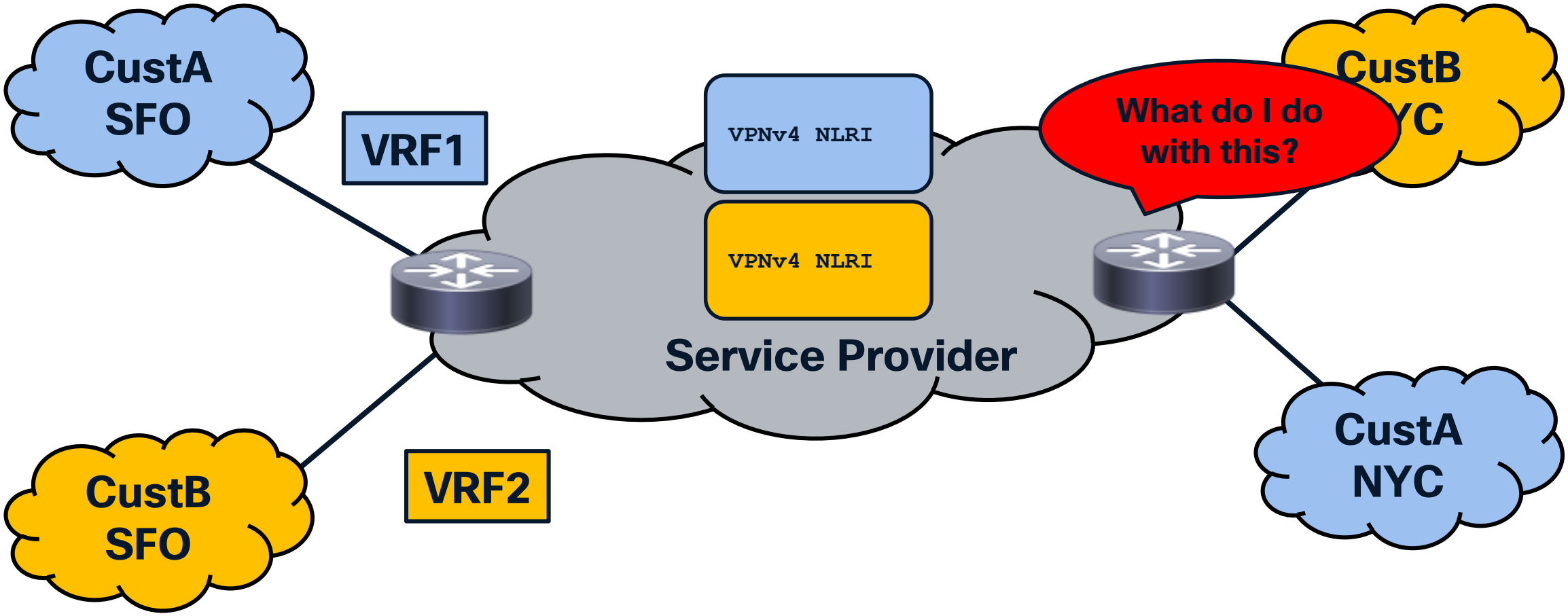
MPLS Layer 3 VPNs (L3VPN)



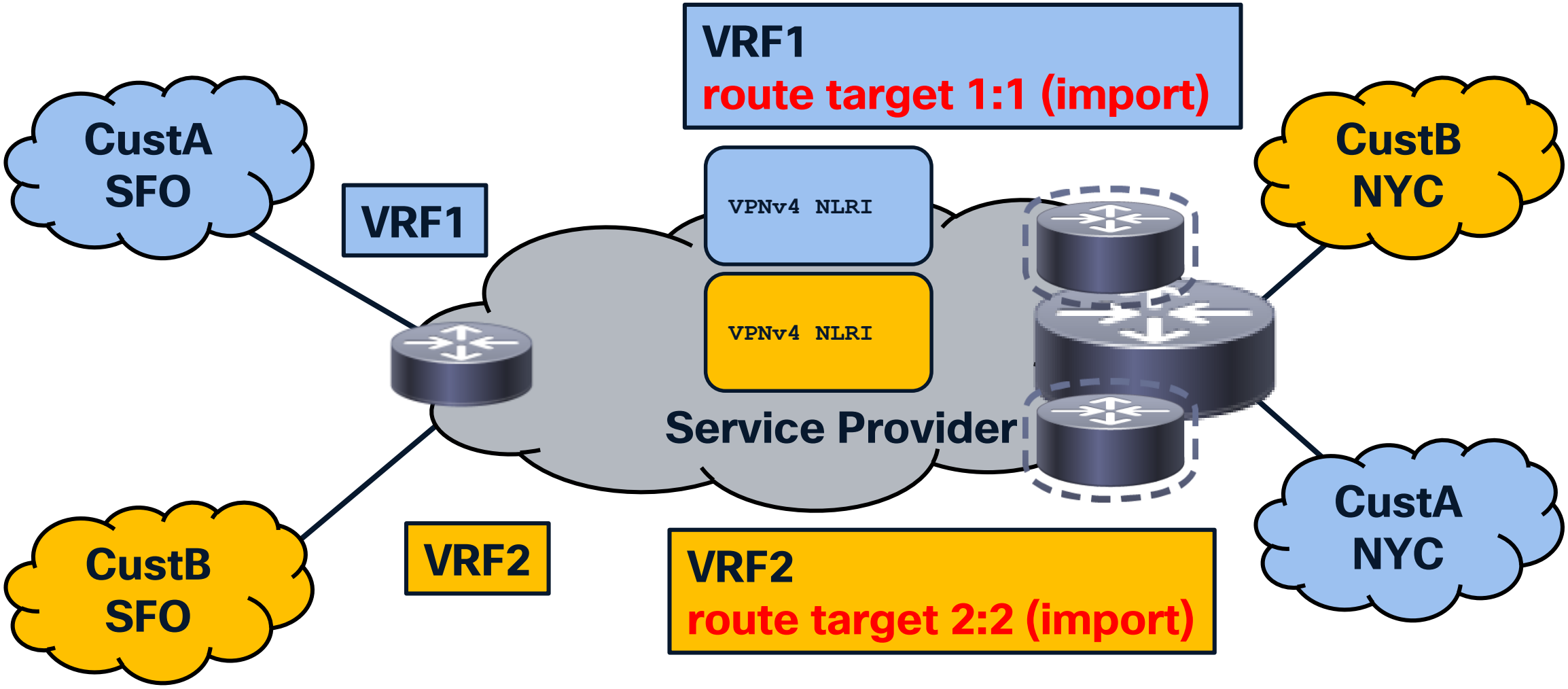
MPLS Layer 3 VPNs (L3VPN)



MPLS Layer 3 VPNs (L3VPN)



MPLS Layer 3 VPNs (L3VPN)

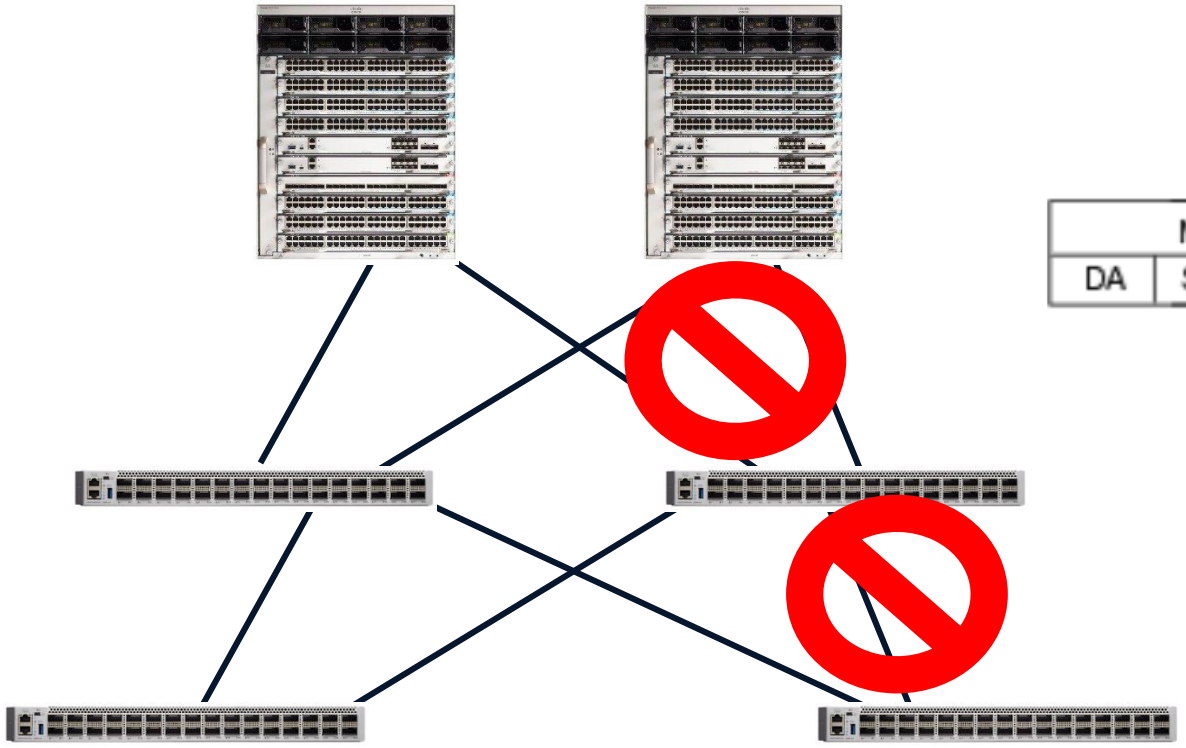


MPLS Layer 3 VPNs (L3VPN)

Use of Address Families in DC Applications

Evolution of the Data Center

Spanning Tree Protocol

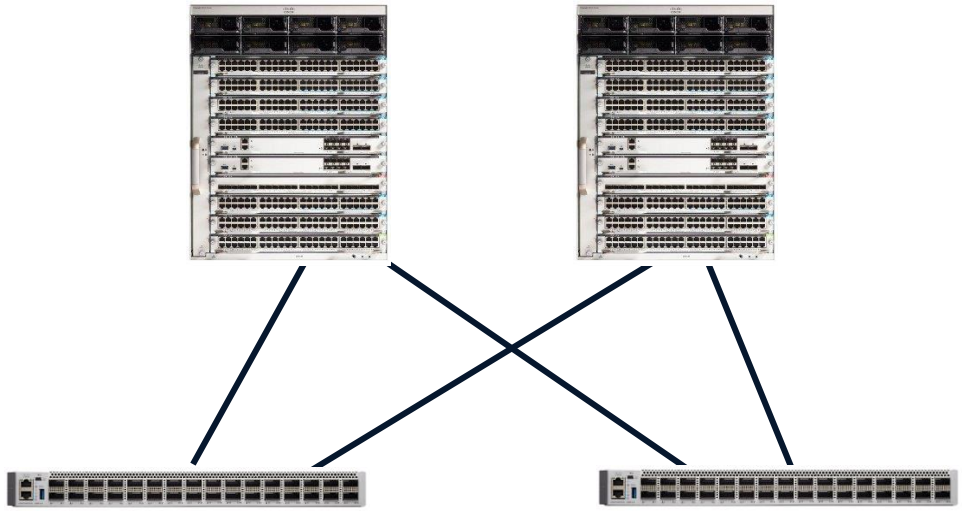


Classical Ethernet frame format

MAC			IP header	Payload
DA	SA	Type		

Evolution of the Data Center

FabricPath



FP frame format

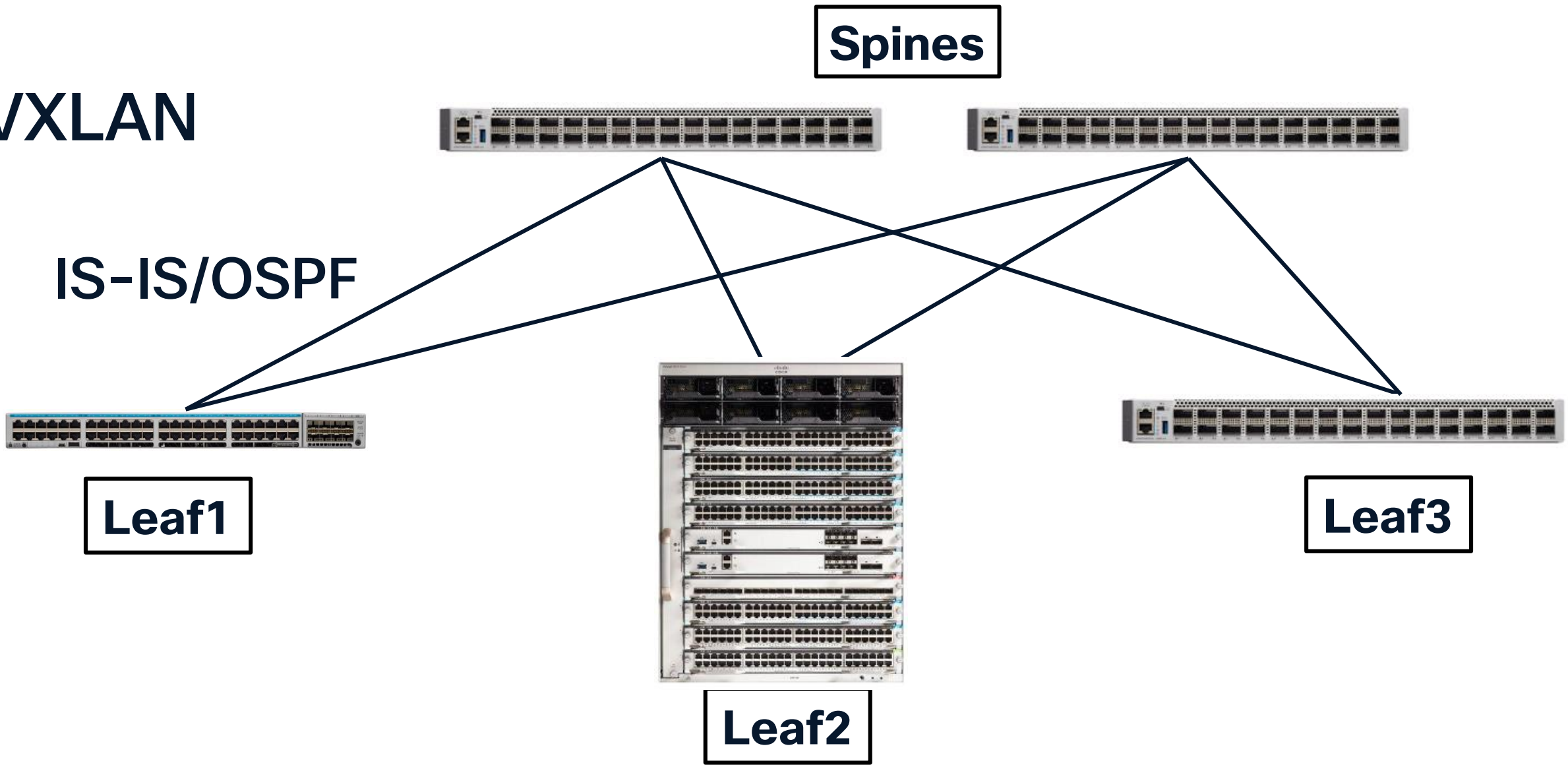
OMAC						iMAC			IP header	Payload
ODA		OSA	FTag			iDA	iSA	Type		
Switch ID	Subswitch ID	Switch ID	Etype 0x8903	FTag	TTL					

FP header = Outer MAC

Evolution of the DC

VXLAN

IS-IS/OSPF



BGP EVPN Route Types

Route Type	Name	Usage
1	Ethernet Auto-Discovery (AD) Route	RFC 7432
2	MAC/IP Advertisement Route	Advertise MAC, address reachability, advertise IP/MAC binding
3	Inclusive Multicast Ethernet Tag Route	RFC7432
4	Ethernet Segment Route	RFC7432
5	IP Prefix Route	RFC7432

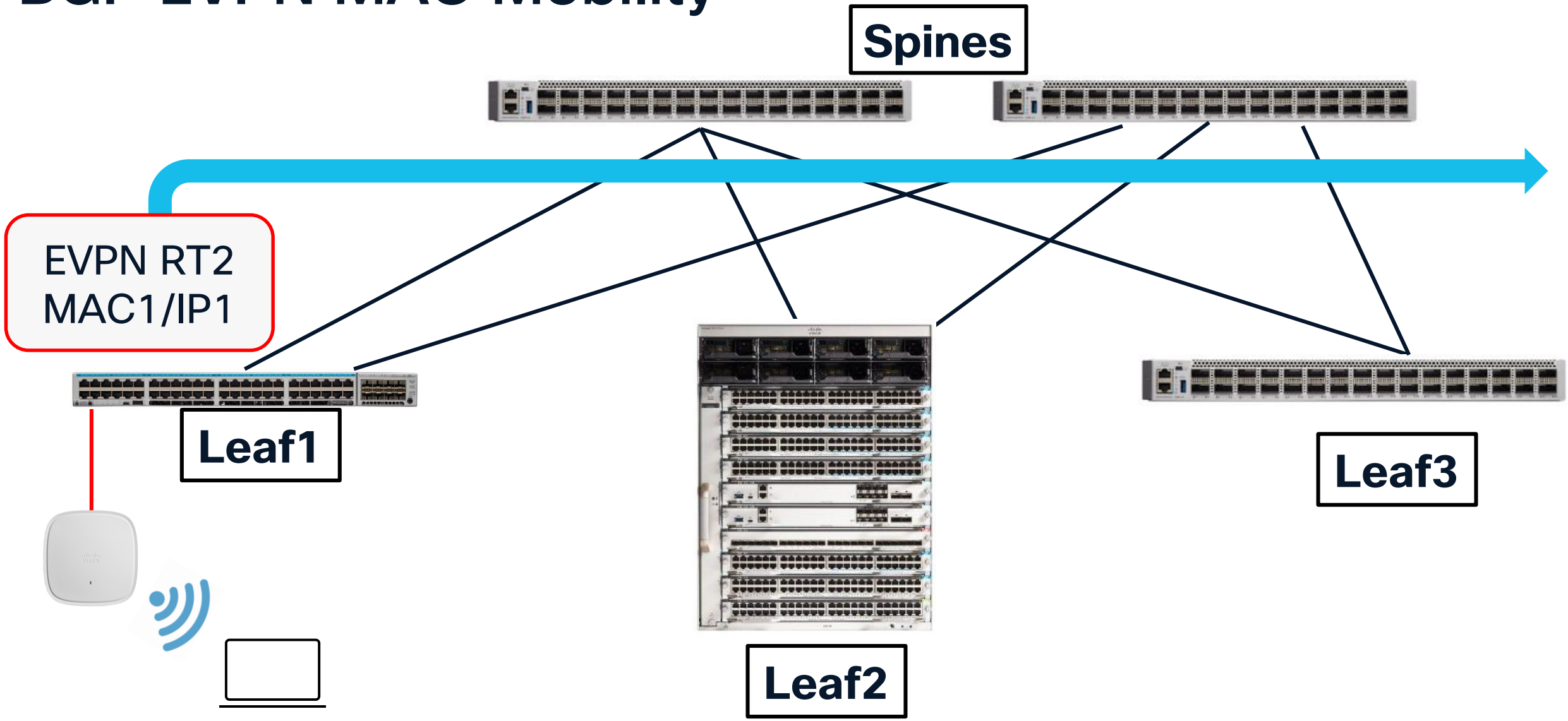
Attributes

BGP Attributes

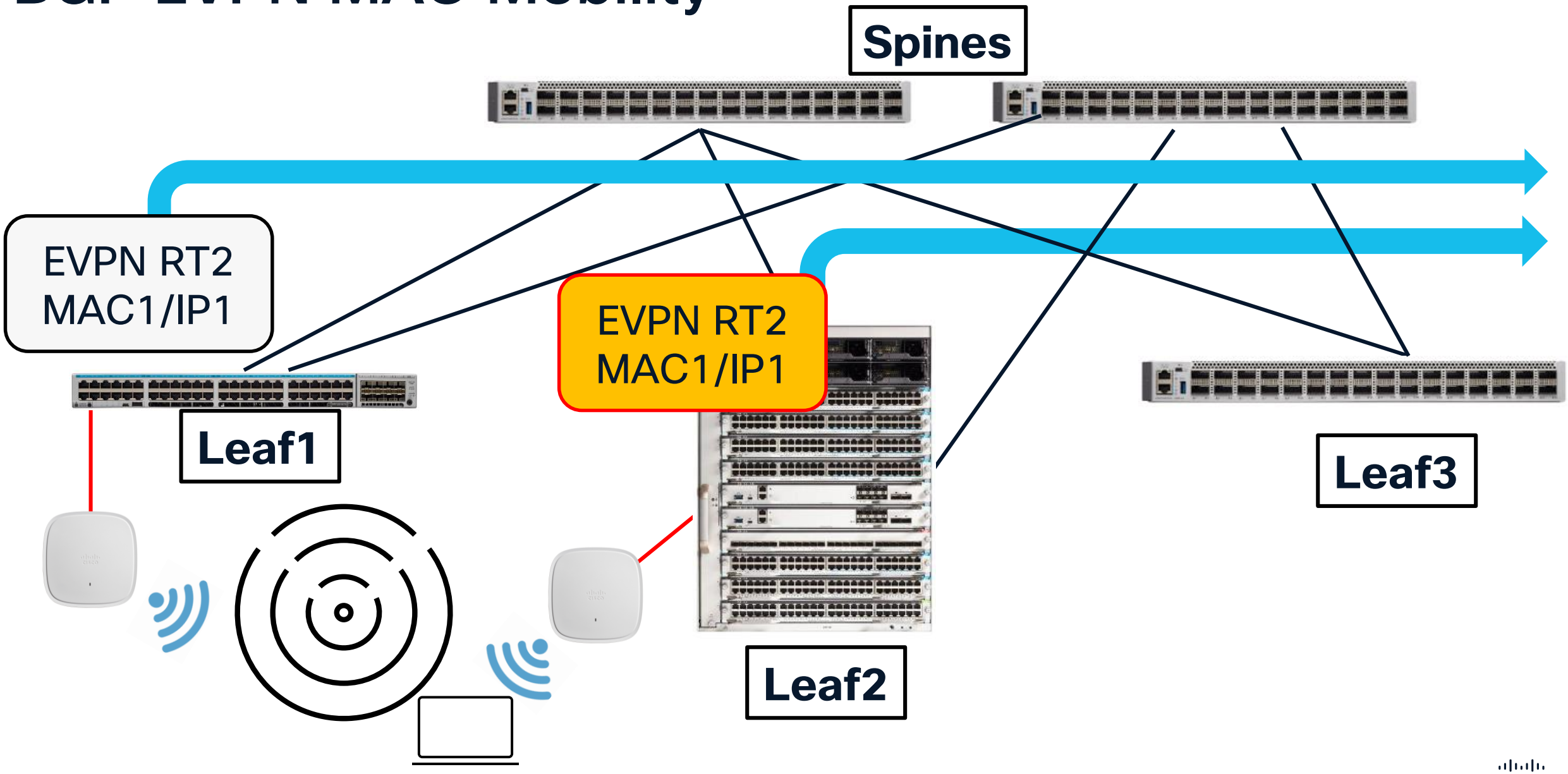
- Well-known mandatory
  - AS_PATH
  - NEXT_HOP
  - ORIGIN
- Well-known discretionary
  - LOCAL_PREF
  - ATOMIC_AGGREGATE
- Optional transitive
  - AGGREGATOR
  - COMMUNITIES
  - EXTENDED_COMMUNITIES
- Optional non-transitive
  - MULTI_EXIT_DISC
  - CLUSTER_LIST
  - MP_REACH_NLRI / MP_UNREACH_NLRI
  - AIGP

Example of Community Attribute use in DC

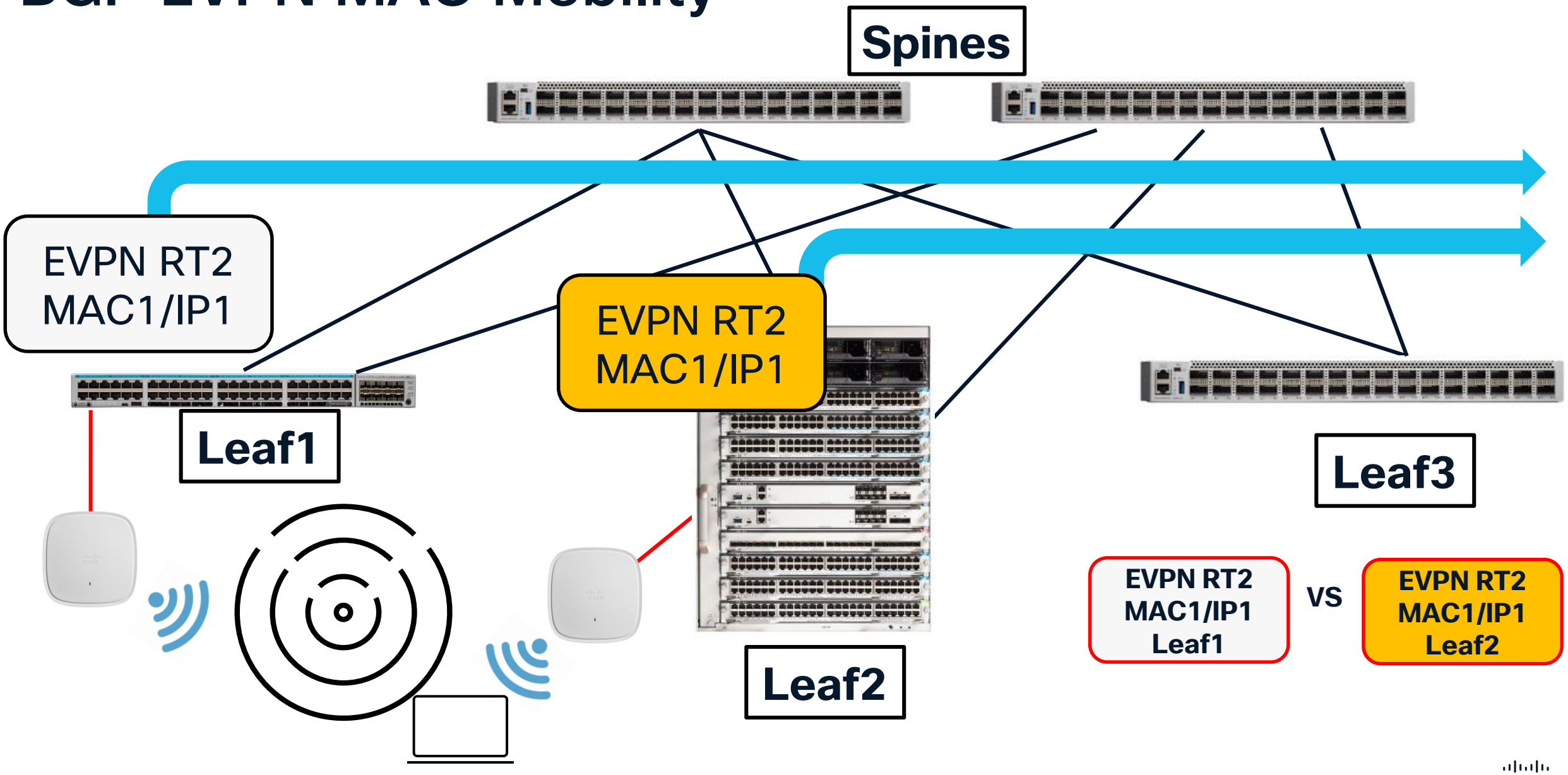
BGP EVPN MAC Mobility



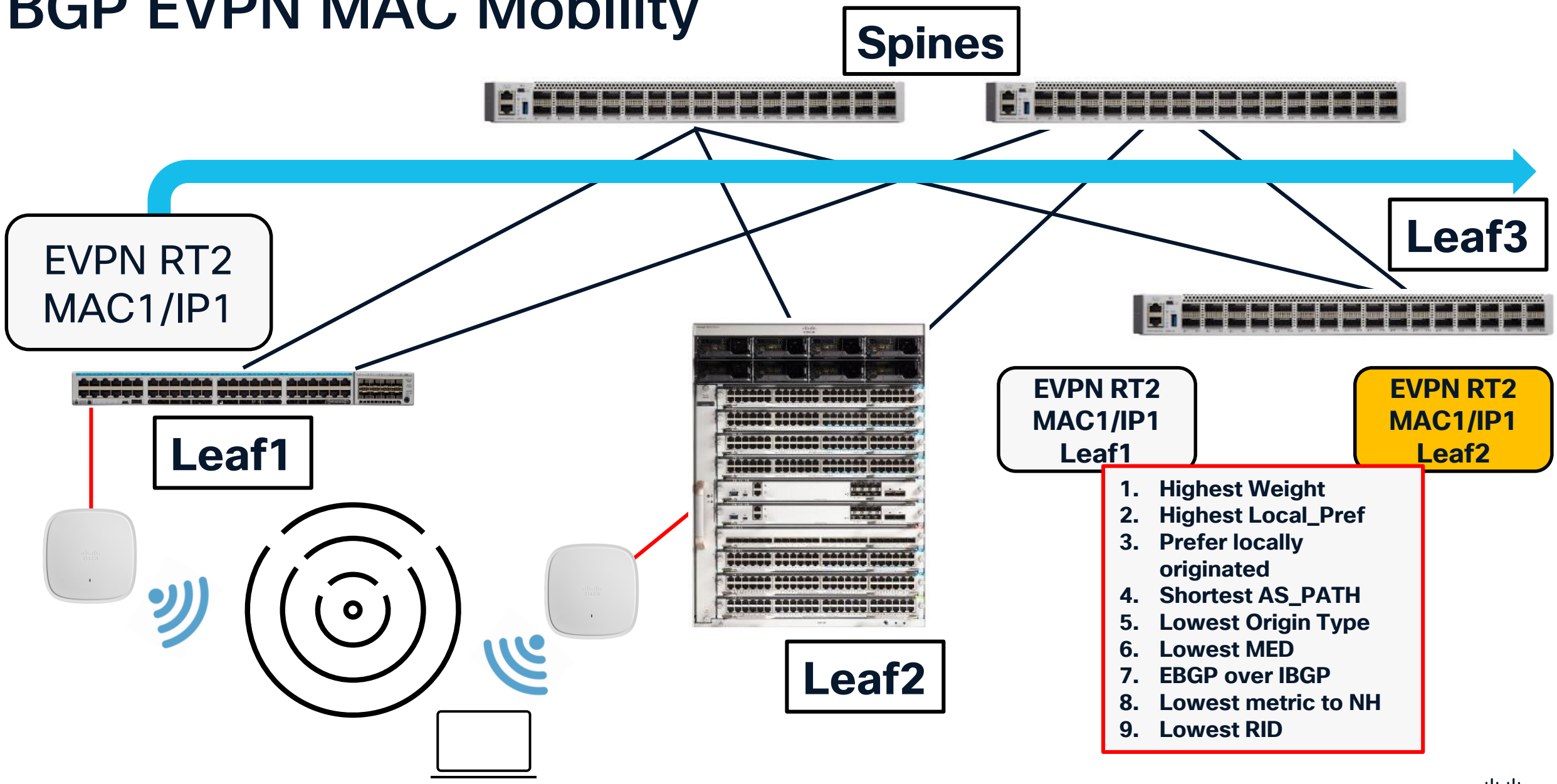
BGP EVPN MAC Mobility



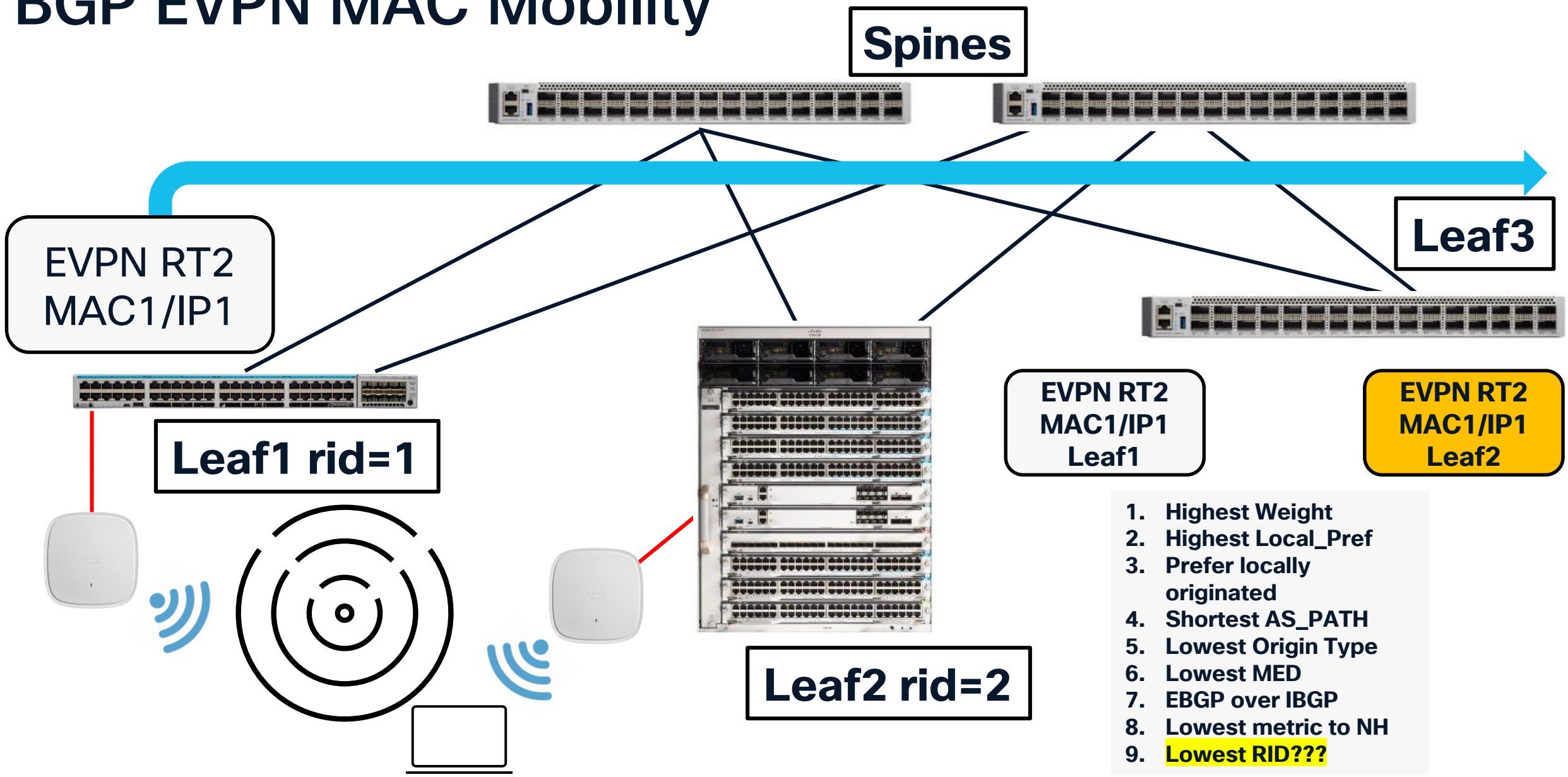
BGP EVPN MAC Mobility



BGP EVPN MAC Mobility

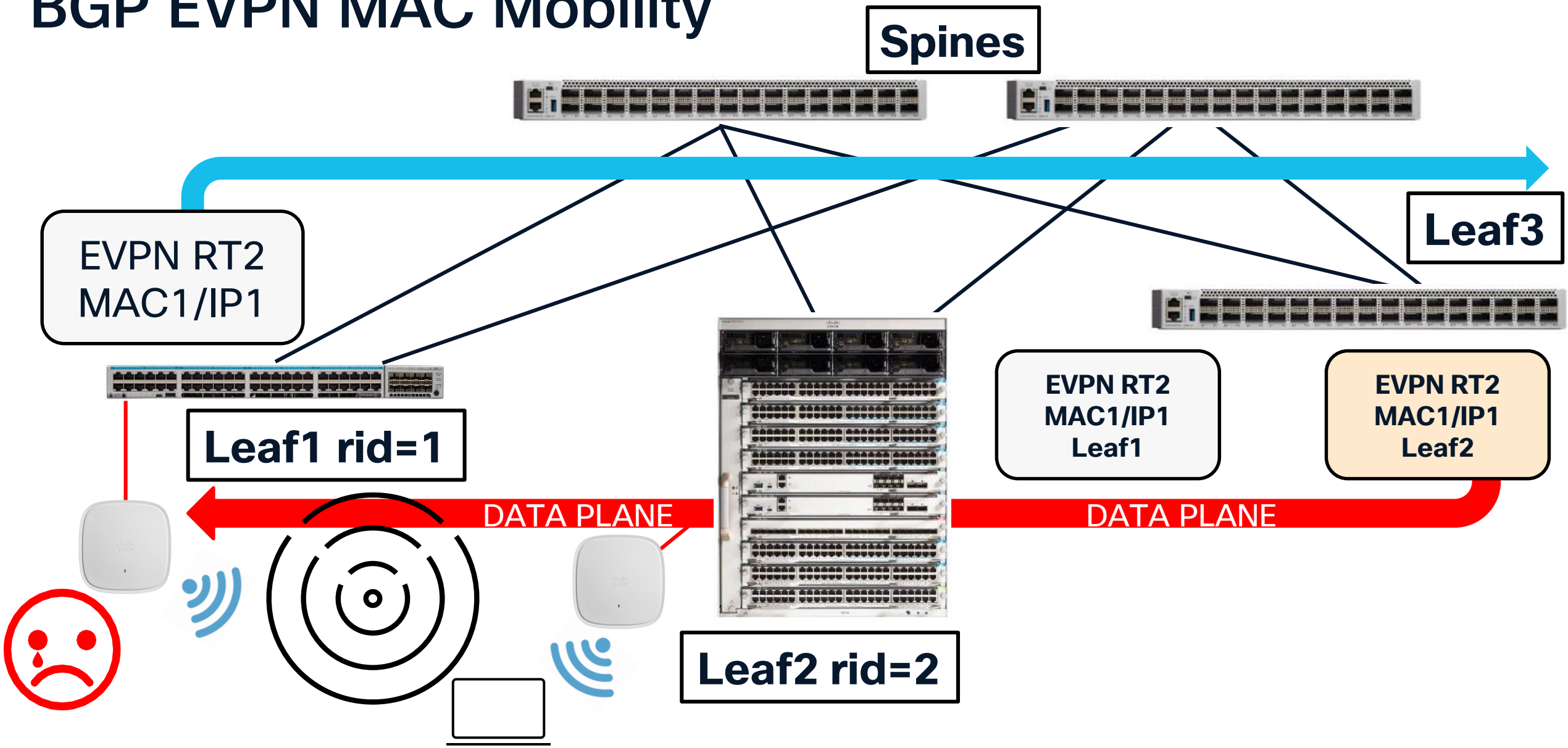


BGP EVPN MAC Mobility

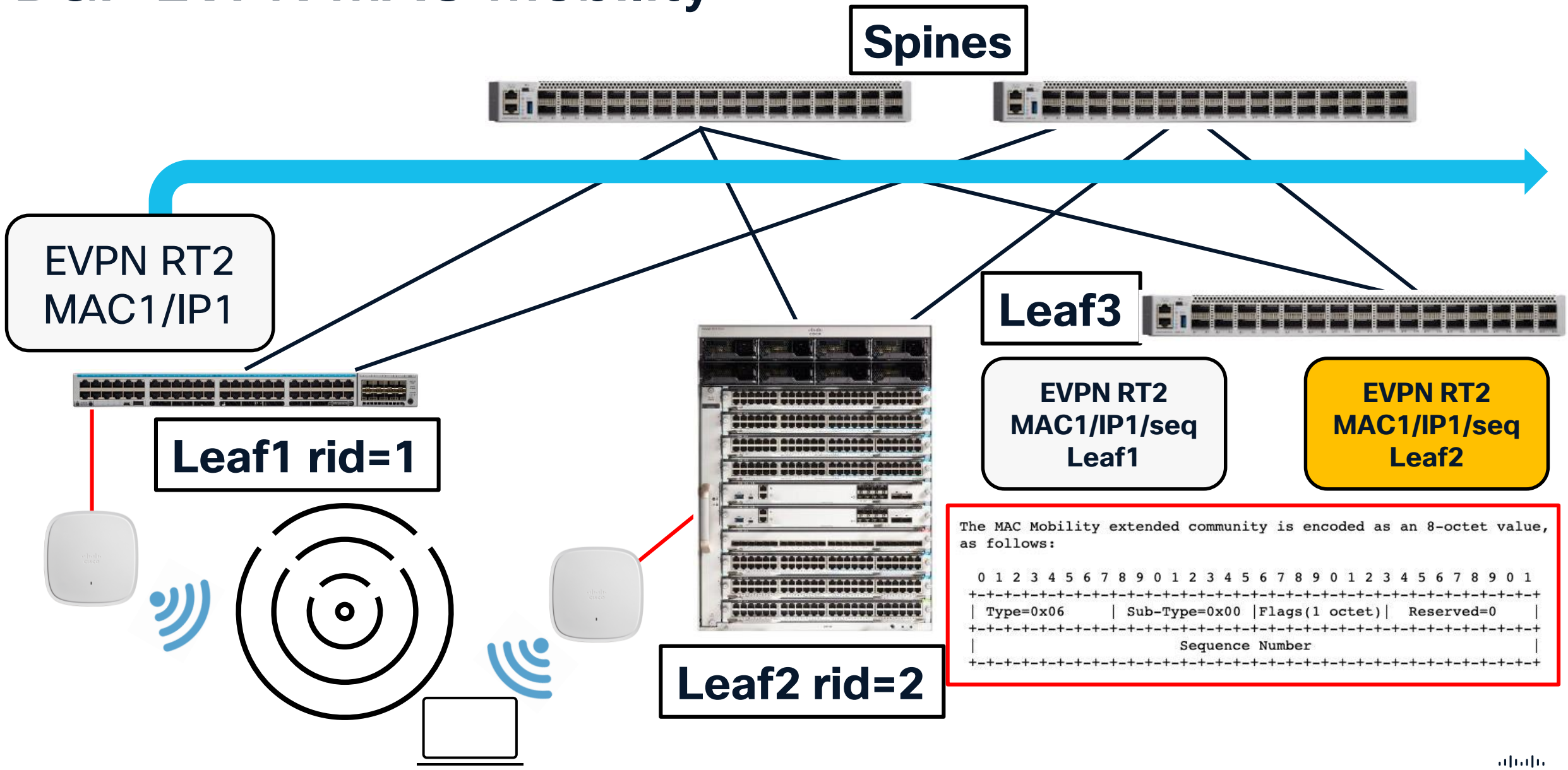


1. Highest Weight
2. Highest Local_Pref
3. Prefer locally originated
4. Shortest AS_PATH
5. Lowest Origin Type
6. Lowest MED
7. EBGP over IBGP
8. Lowest metric to NH
9. Lowest RID???

BGP EVPN MAC Mobility



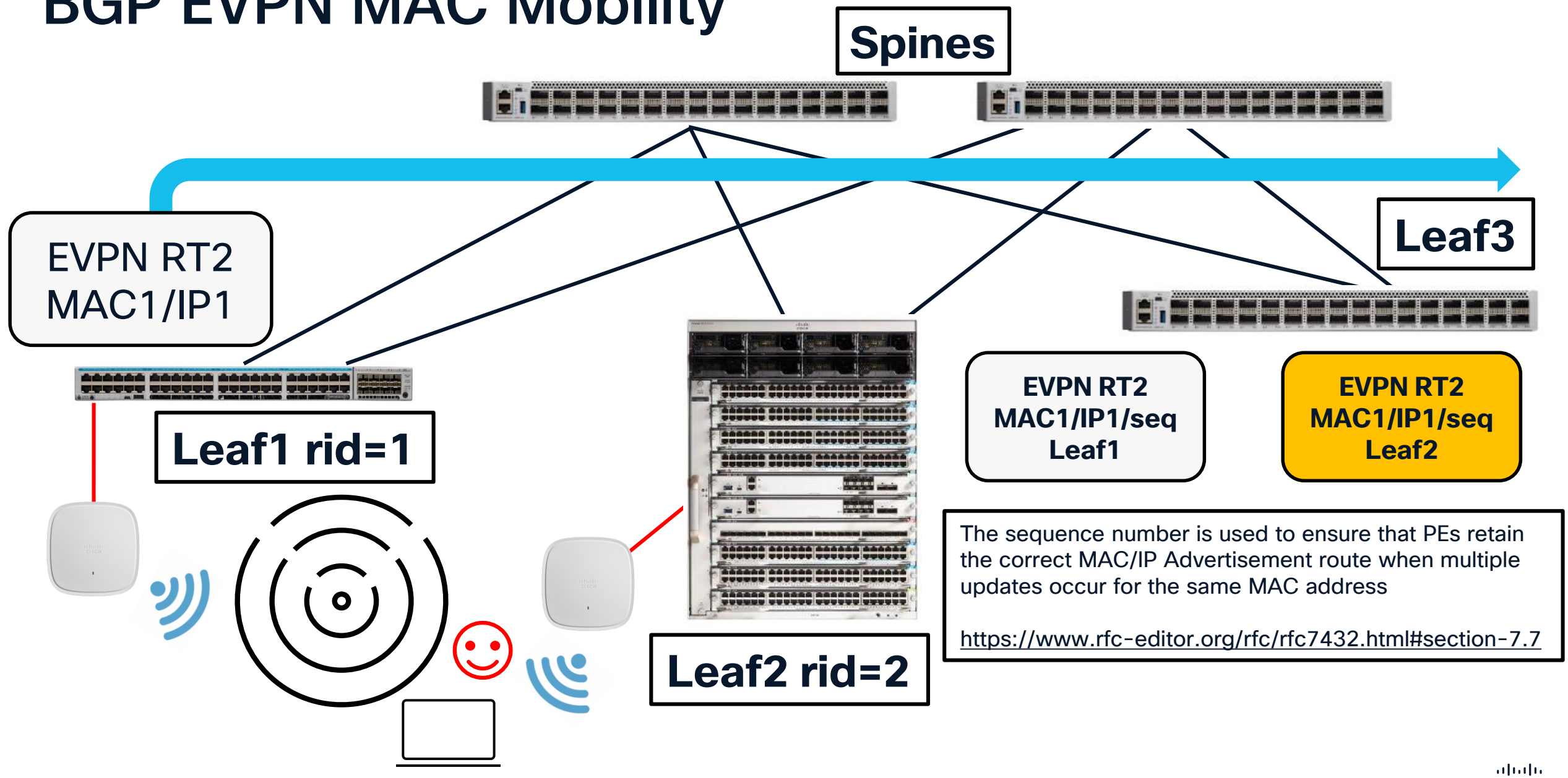
BGP EVPN MAC Mobility



The MAC Mobility extended community is encoded as an 8-octet value, as follows:

```
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Type=0x06     | Sub-Type=0x00 |Flags(1 octet)|   Reserved=0    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          Sequence Number                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

BGP EVPN MAC Mobility



The sequence number is used to ensure that PEs retain the correct MAC/IP Advertisement route when multiple updates occur for the same MAC address

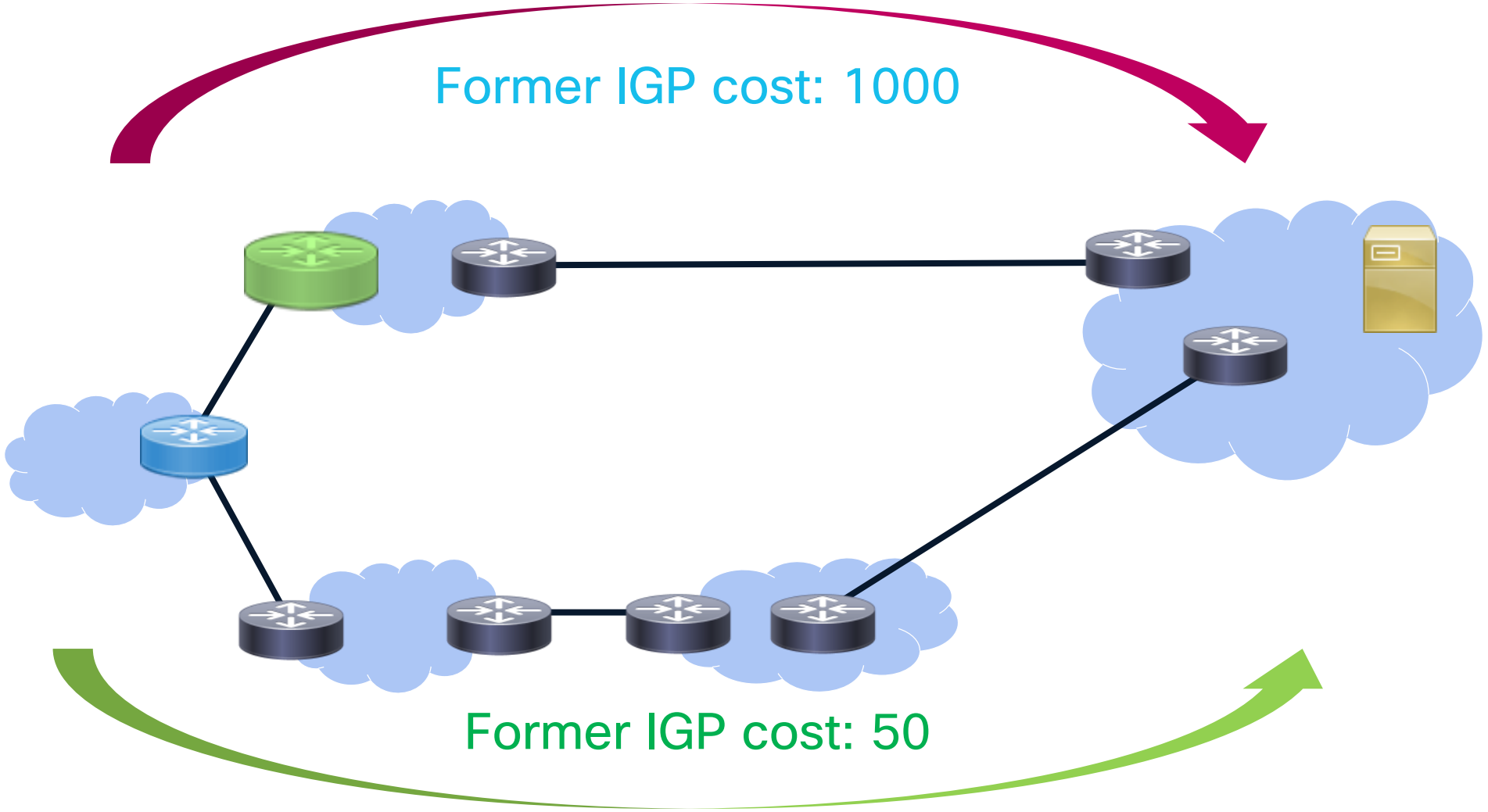
https://www.rfc-editor.org/rfc/rfc7432.html#section-7.7

Accumulated IGP Cost Attribute

Predictable Routing Between Autonomous Systems

- Very large networks can be split into separate IGP domains and treated as multiple autonomous systems running BGP
- Even after such split, there may be a requirement for BGP to select best paths similarly to how the IGP would if the whole network ran it
  - Have BGP select best paths on the total IGP metric to the destination
- To accomplish this, we need
  - An attribute accumulating the total IGP metric to the destination
  - A modification to the best path selection algorithm to consider the cumulative IGP metric early in the process

Motivation for AIGP attribute



Accumulated IGP Metric Attribute

- RFC 7311 introduces the Accumulated IGP Metric attribute
  - Optional Nontransitive attribute carrying the total IGP metric to the NLRI
- When injecting a route into BGP…
  - AIGP is set to the route's IGP metric (or to an arbitrary value)
- When advertising a route in BGP…
  - A BGP router that does not modify the NEXT_HOP of the route does not modify the AIGP attribute either;
  - A BGP router that modifies the NEXT_HOP of the route to itself adds its own IGP metric to the former next hop to the AIGP attribute value

Accumulated IGP Metric Attribute

- When selecting the best path...
  - The AIGP comparison is performed after Step 3 (prefer locally originated routes) and before Step 4 (prefer shortest AS_PATH)
- The overall result
  - AIGP contains the (almost) total IGP distance to the destination
  - AIGP is considered right after Weight, Local Preference and self-originated routes, beating the AS_PATH
  - BGP's best path selection with AIGP very closely resembles the path selection a single network-wide IGP would do

AIGP Deployment Considerations

- Using AIGP only makes sense if the individual autonomous systems run IGPs with comparable metric calculations
  - Summing together, say, EIGRP distance with OSPF cost is meaningless
- To utilize AIGP, peers must be mutually configured
- IOS XE implements RFC 7311 with a few quirks
  - When injecting routes to BGP, they must be assigned the initial AIGP value explicitly
  - Between directly connected eBGP peers with on-link addresses, AIGP is not incremented (CSCut48797)

Security

TCP Authentication Option

TCP Authentication Option

- BGP relies on TCP support to provide optional authentication
- Initially, TCP only supported MD5-based hash (RFC 2385)
- RFC 5925 brings in TCP Authentication Option (TCP-AO)
  - Cryptographically stronger protection based on hash message authentication codes; does not prescribe any particular hash function
- For TCP-AO, IOS XE supports
  - hmac-sha-1
  - hmac-sha-256
  - aes-128-cmac

Configuring BGP authentication with TCP-AO

- Parameters of the TCP Authentication Option need to be configured in a specialized key chain of the “tcp” type

```
key chain bgpkeys tcp
 key 1
  send-id 123
  recv-id 123
  cryptographic-algorithm aes-128-cmac
  key-string s3cr3tP4ss
```

- The key chain then must be applied to the neighbor

```
router bgp 64512
 neighbor ... ao bgpkeys
```

Secure eBGP Default Policy

Secure eBGP Default Policy

- RFC 8212 stipulates a strict default behavior for eBGP
  - If no explicit inbound policy for a peer is configured, ignore any received routes from this peer
  - If no explicit outbound policy for a peer is configured, do not advertise any routes to this peer
- This serves multiple purposes
  - Preventing a router from inadvertently becoming a transit router
  - Preventing a router from advertising unintended routes
  - Preventing a router from having its existing routing inadvertently changed
  - Preventing a router from crashing due to receiving too many routes

Secure eBGP policy in IOS-XR and IOS XE

- This has been the default behavior in IOS-XR since day one
- In IOS XE, this behavior can be activated by

```
router bgp 64512
 bgp safe-ebgp-policy
```

- Both IOS-XR and IOS XE implement RFC 8212 with a quirk
  - An eBGP neighbor must be explicitly configured with both inbound and outbound policy, otherwise no routes will be exchanged with it

BGP Multi Session Capability

BGP Multi Session Capability

- BGP is a true multi-address-family routing protocol
  - A single BGP session can carry multiple disparate address families
- There are pros and cons to running multiple address families over a single BGP session
- A major argument against it is fate sharing
  - If the session goes down for any reason, all address families advertised over it are affected
- Multi Session Capability allows a BGP speaker to open a new session to the same peer for every enabled address family

Enabling BGP Multi Session Capability

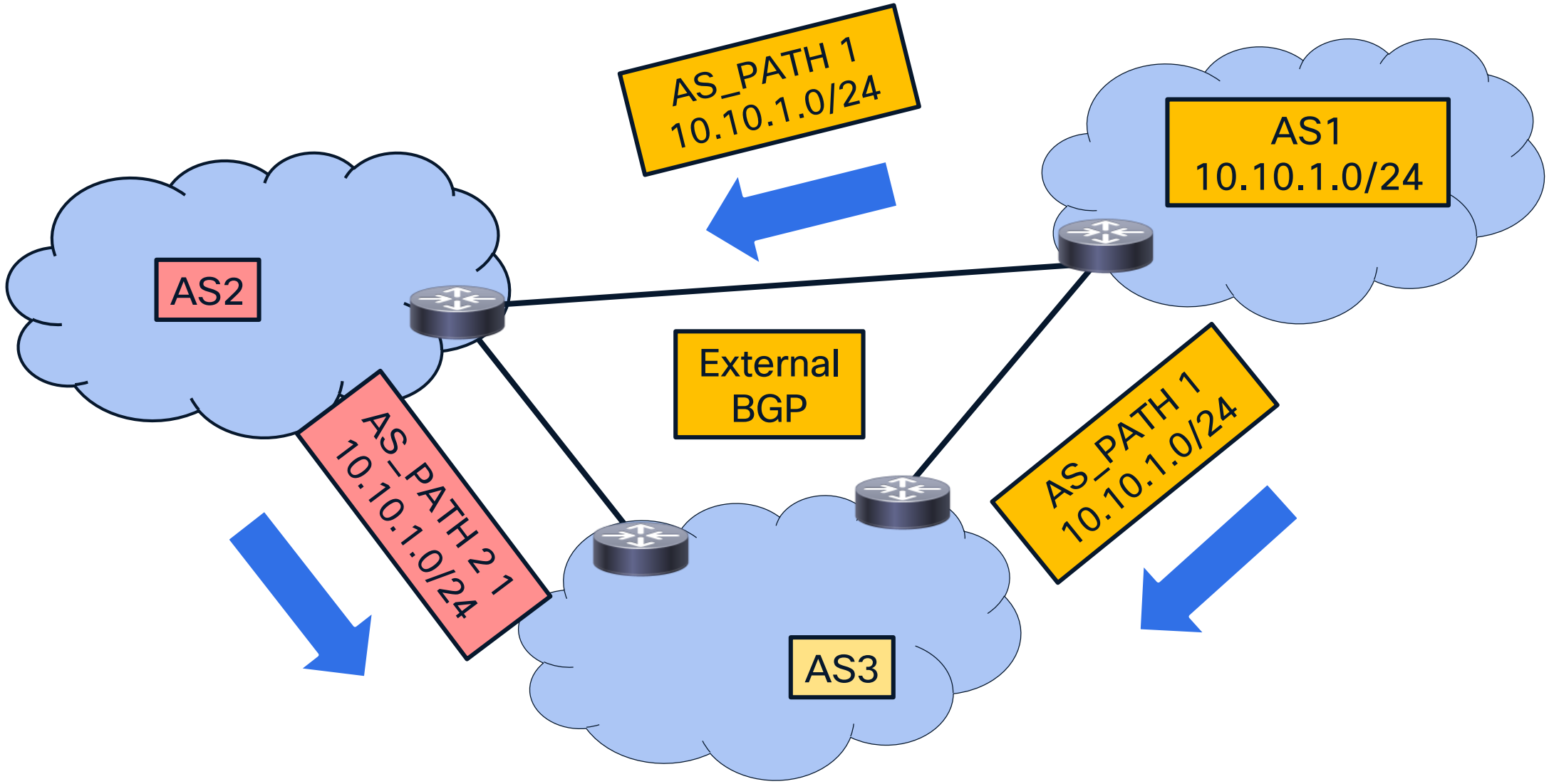
- Multi Session Capability is supported by IOS XE
- Both peers must be mutually configured for multi session, otherwise they will not establish a peering

```
router bgp 64512
 neighbor ... transport multi-session
```

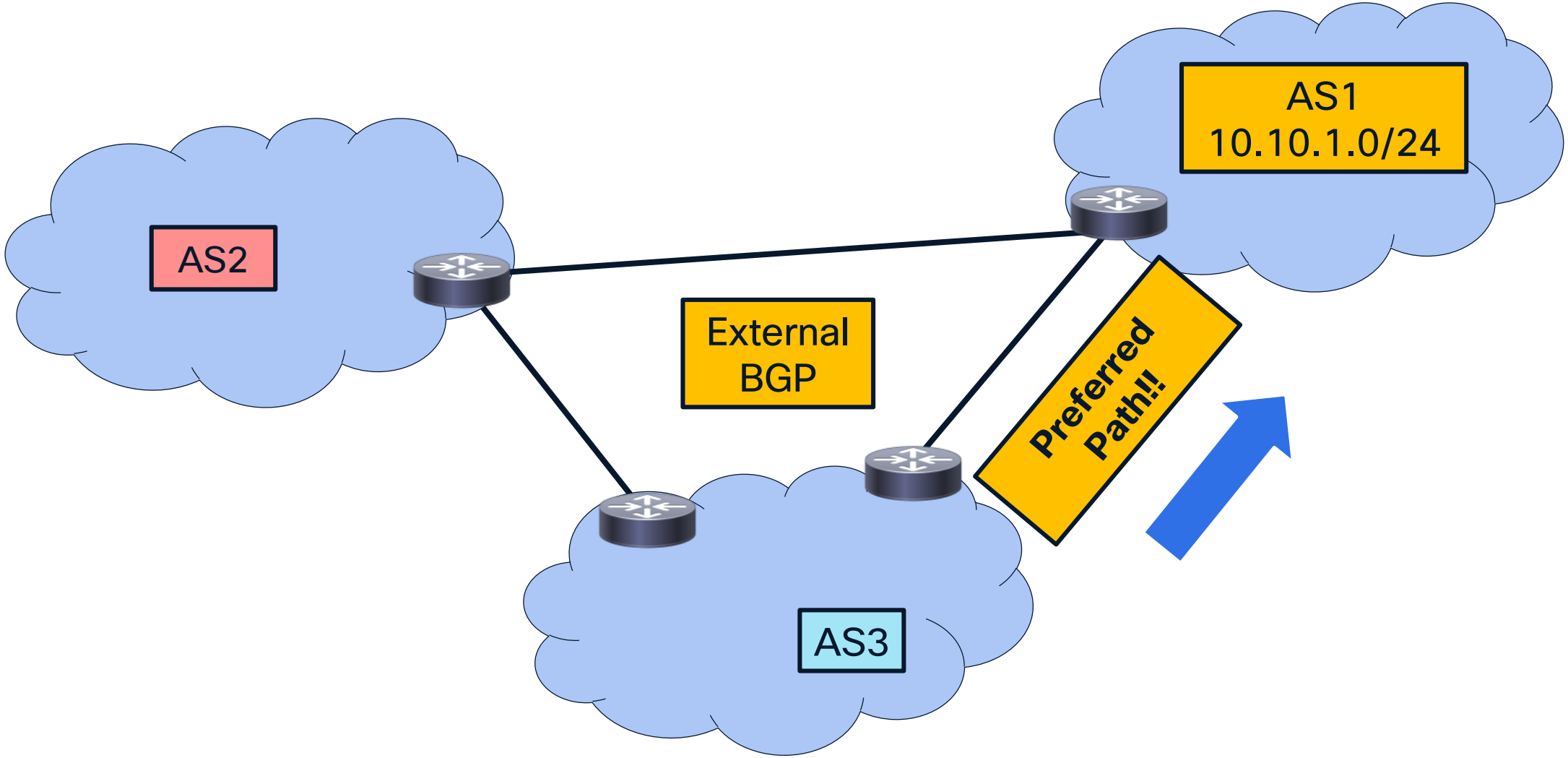
- Once enabled on both peers, every address family toward the same peer's IP address will be carried in a dedicated TCP session

Resource Public Key Infrastructure

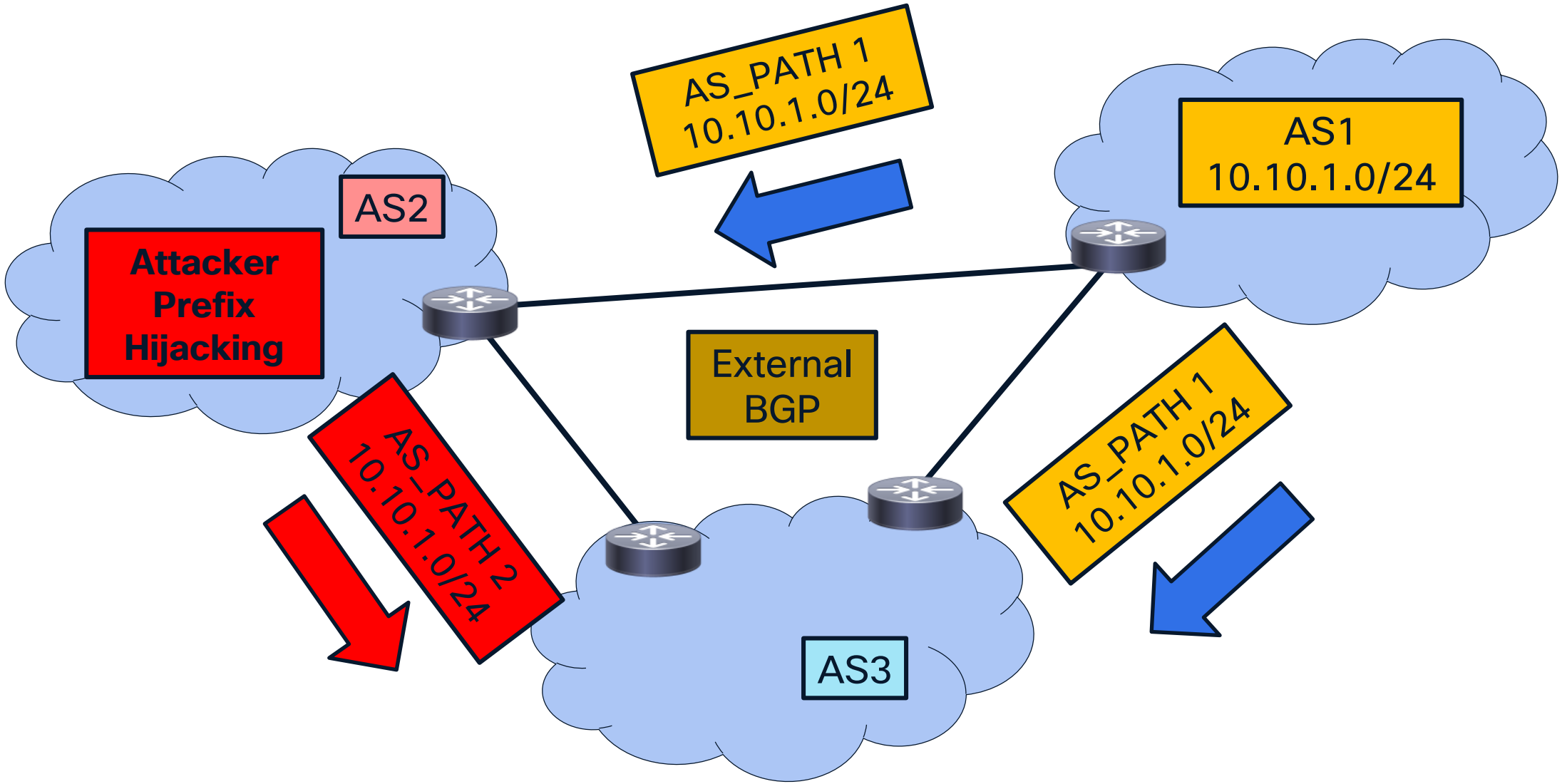
BGP Resource Public Key Infrastructure (RPKI)



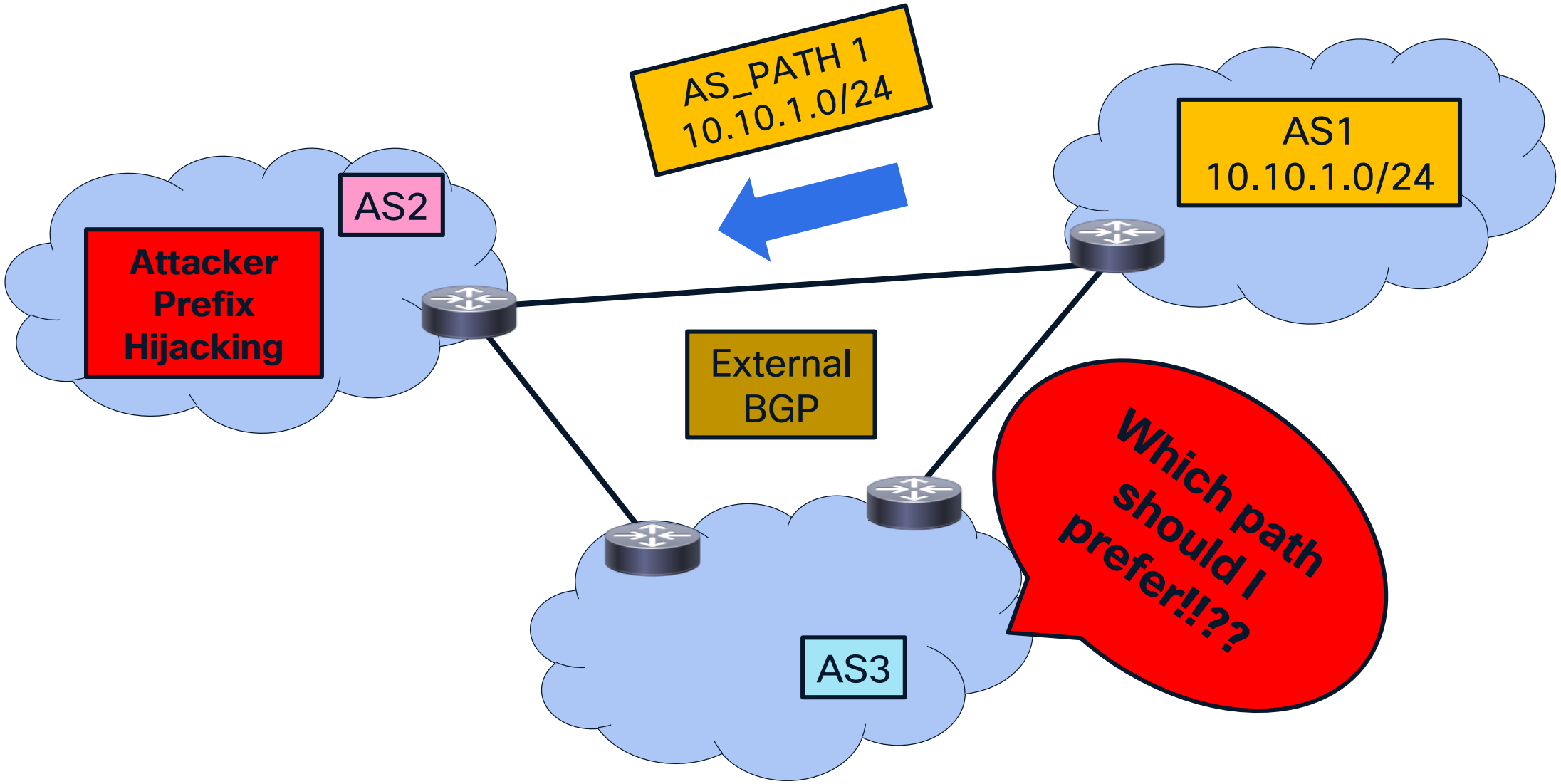
BGP Resource Public Key Infrastructure (RPKI)



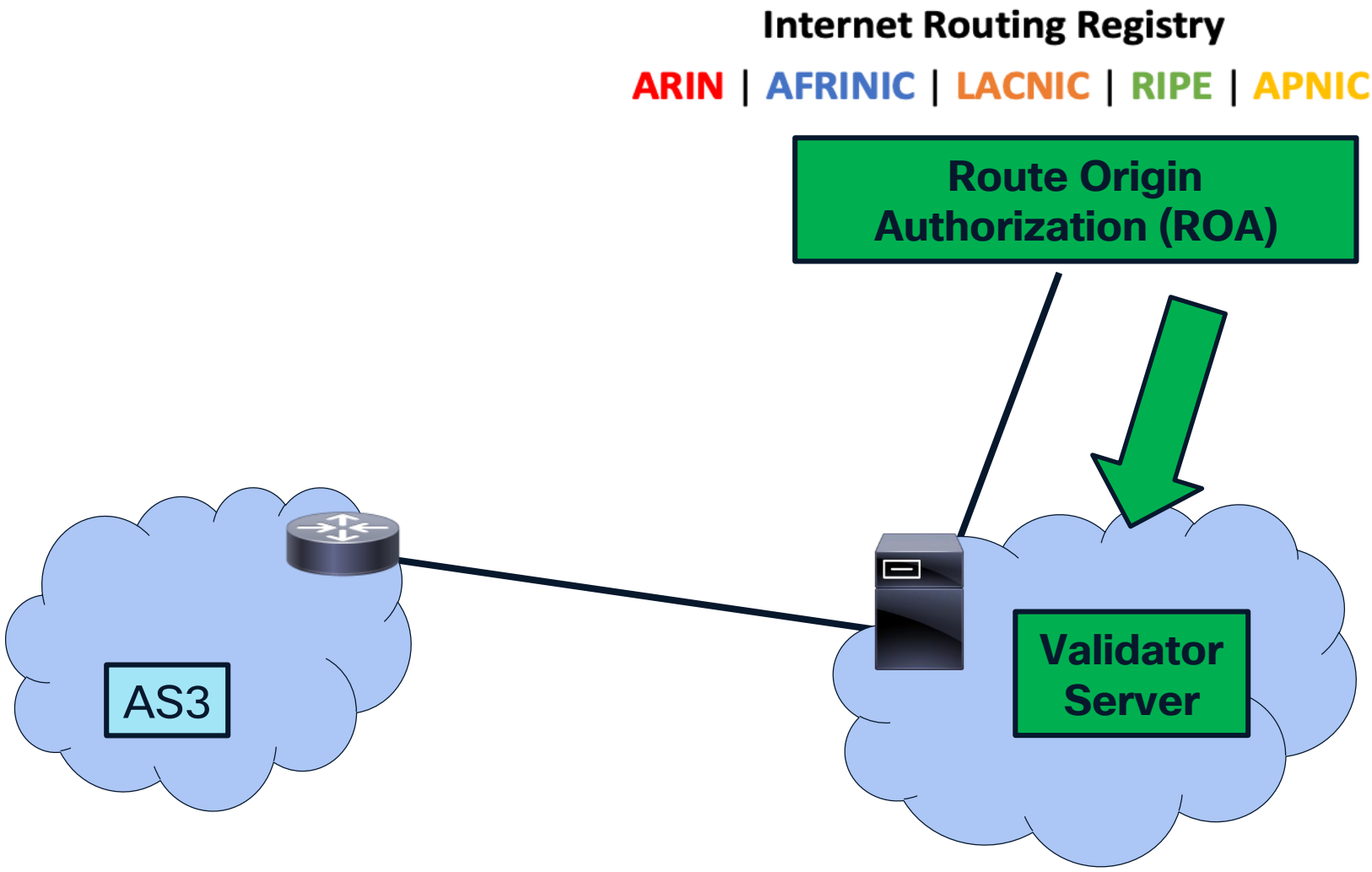
BGP Resource Public Key Infrastructure (RPKI)



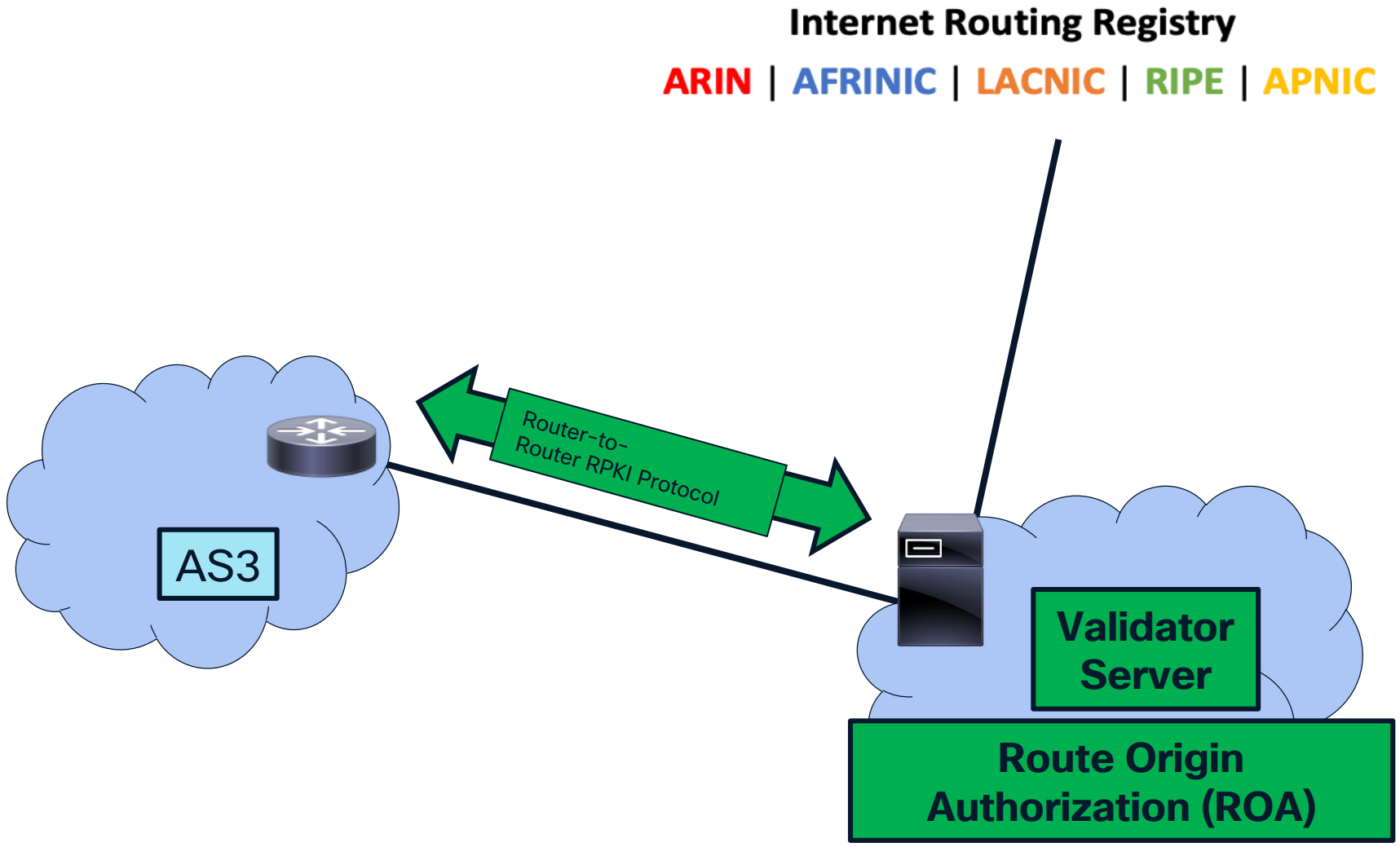
BGP Resource Public Key Infrastructure (RPKI)



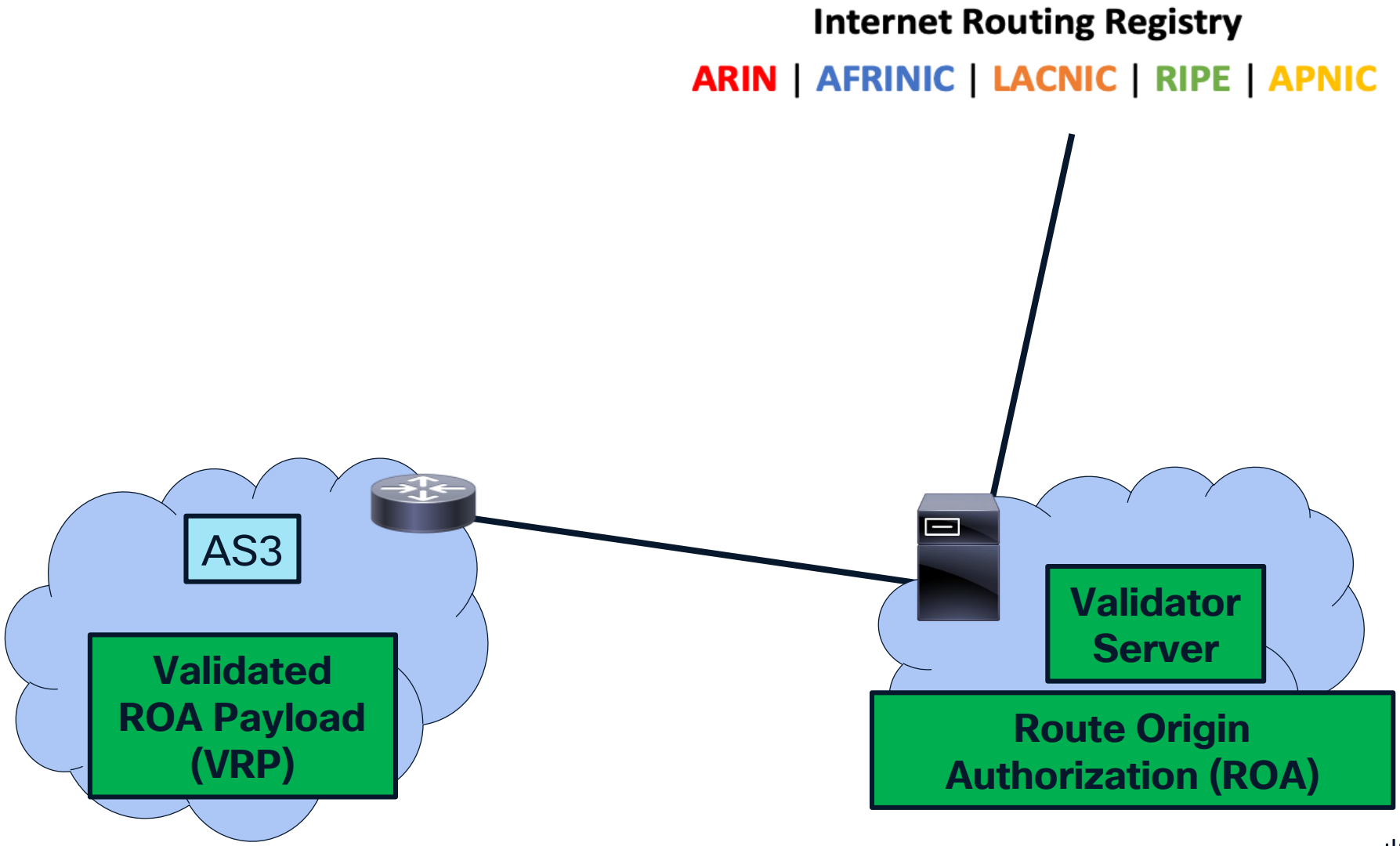
BGP Resource Public Key Infrastructure (RPKI)



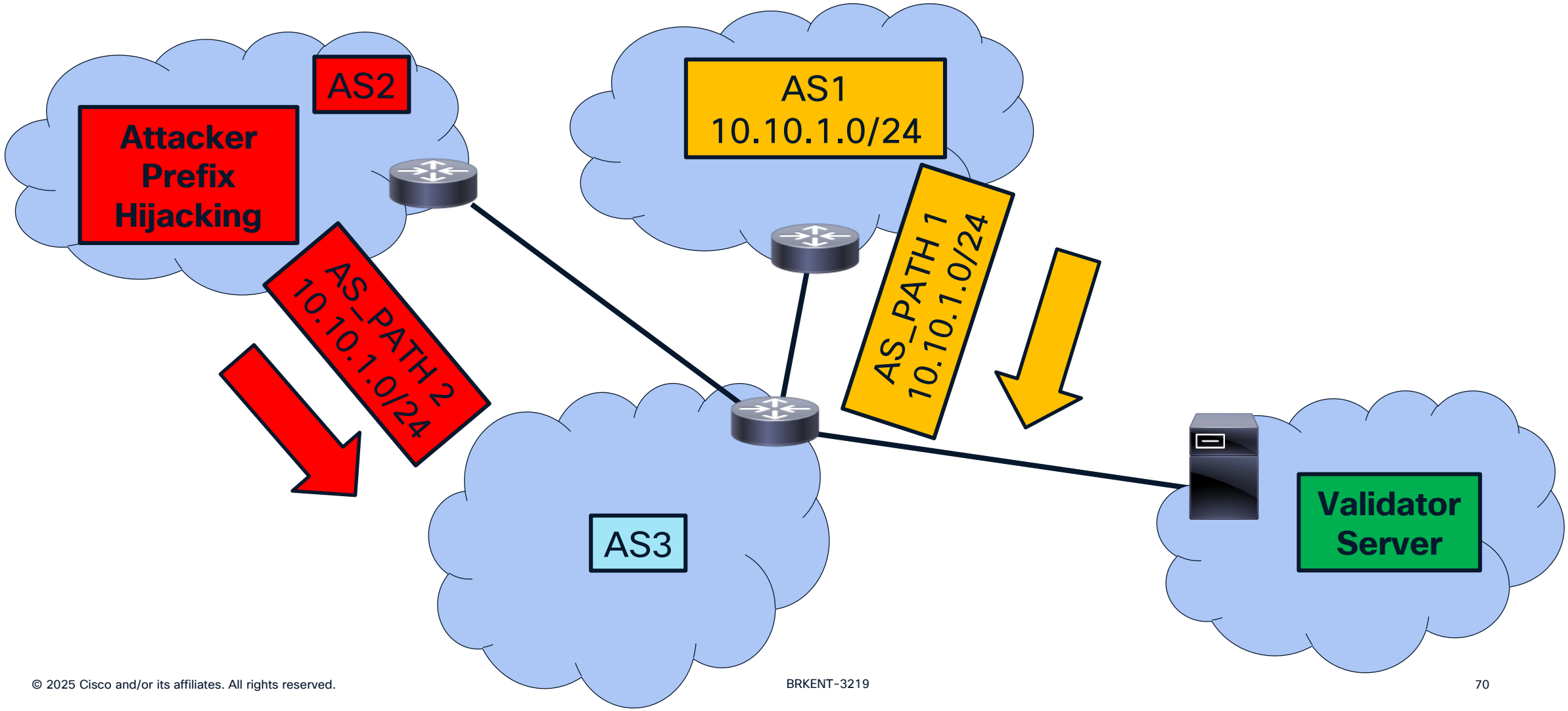
BGP Resource Public Key Infrastructure (RPKI)



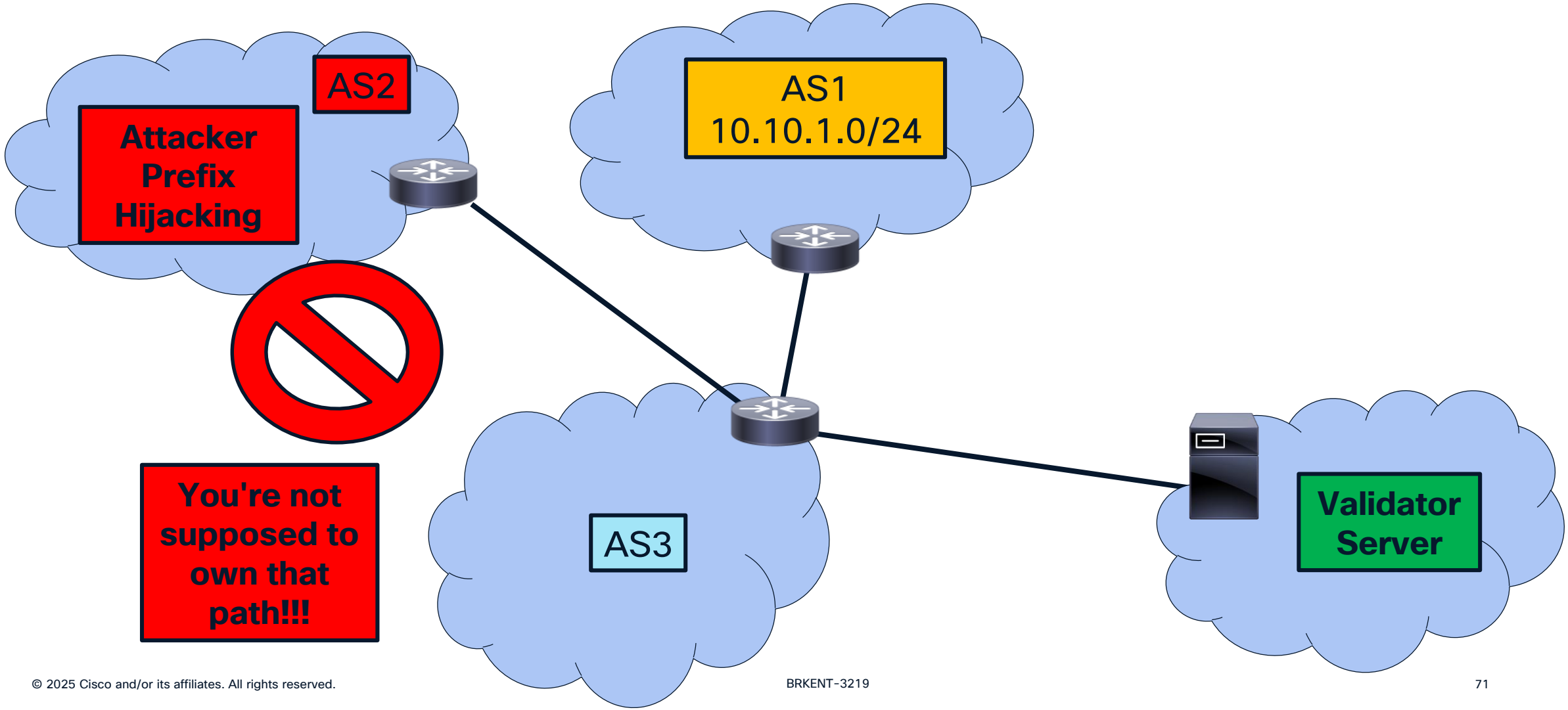
BGP Resource Public Key Infrastructure (RPKI)



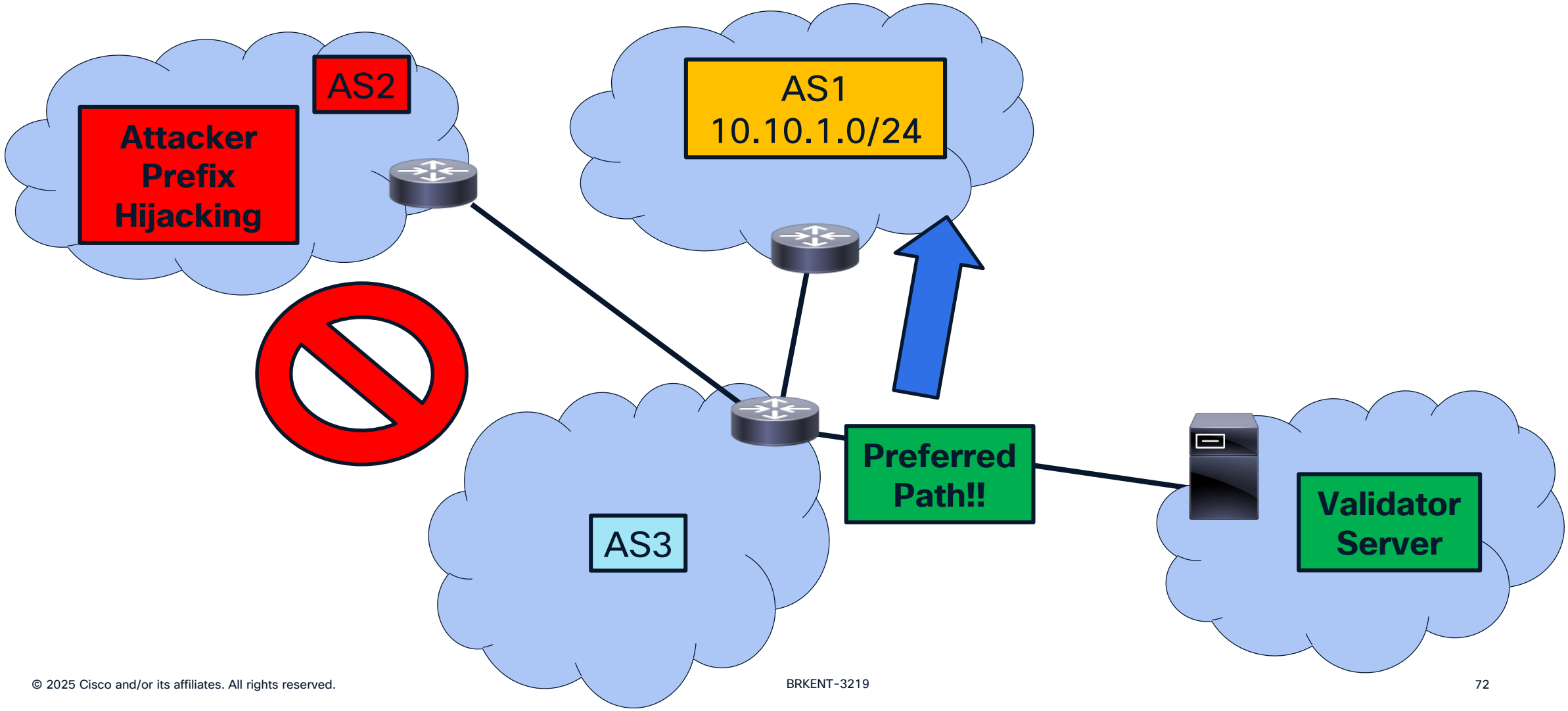
BGP Resource Public Key Infrastructure (RPKI)



BGP Resource Public Key Infrastructure (RPKI)



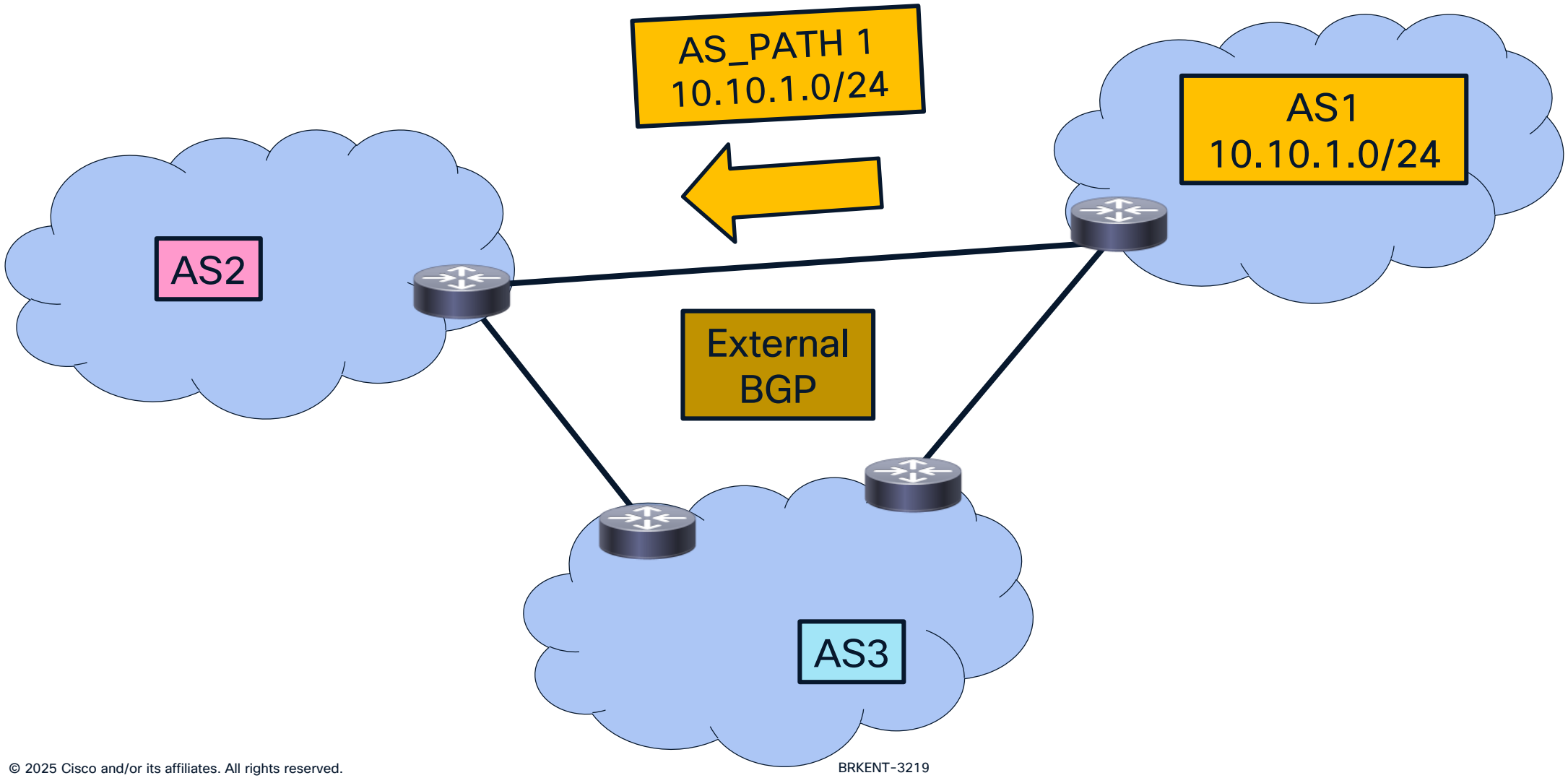
BGP Resource Public Key Infrastructure (RPKI)



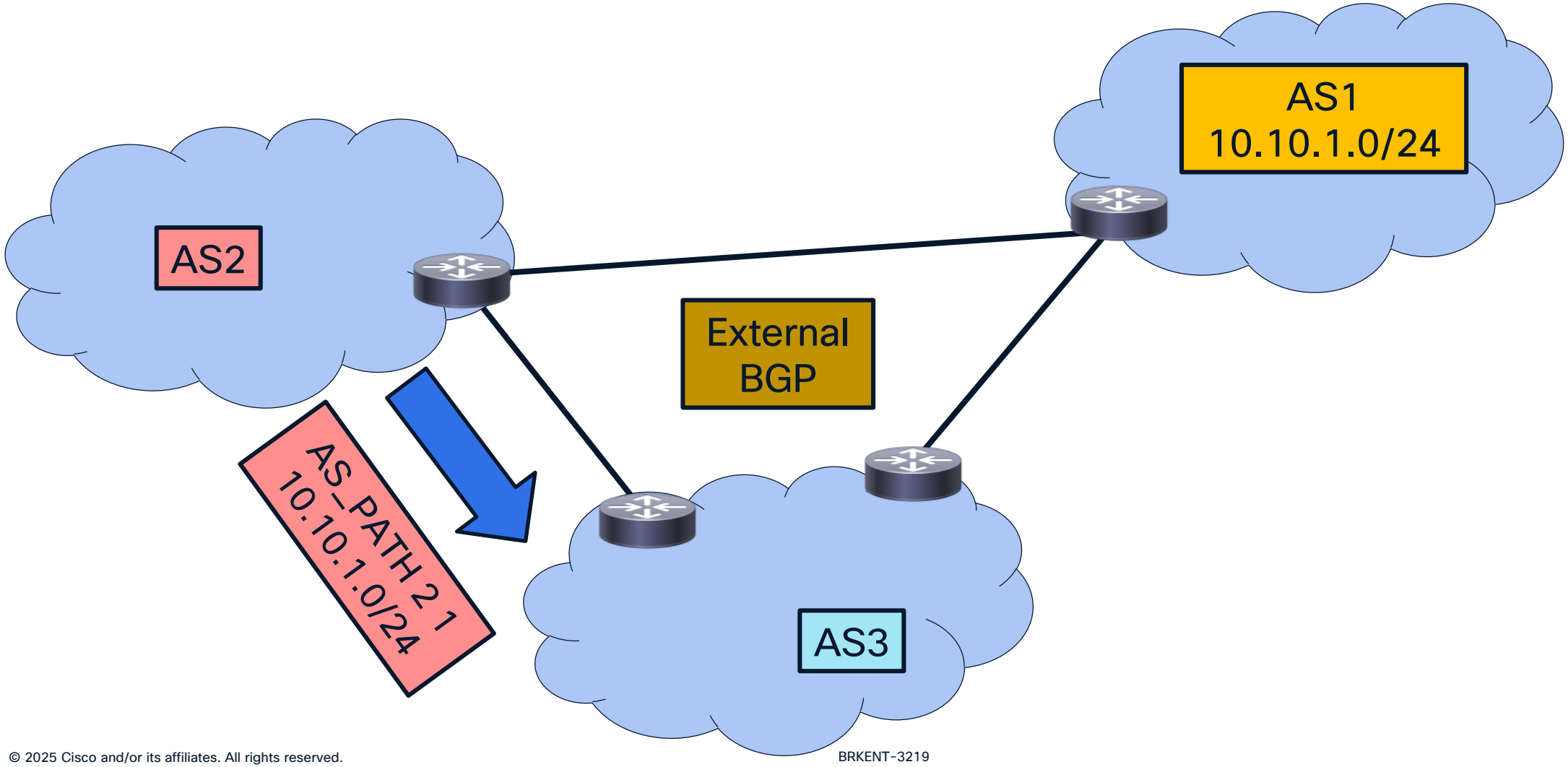
Scalability

Loop Prevention and Route Reflection

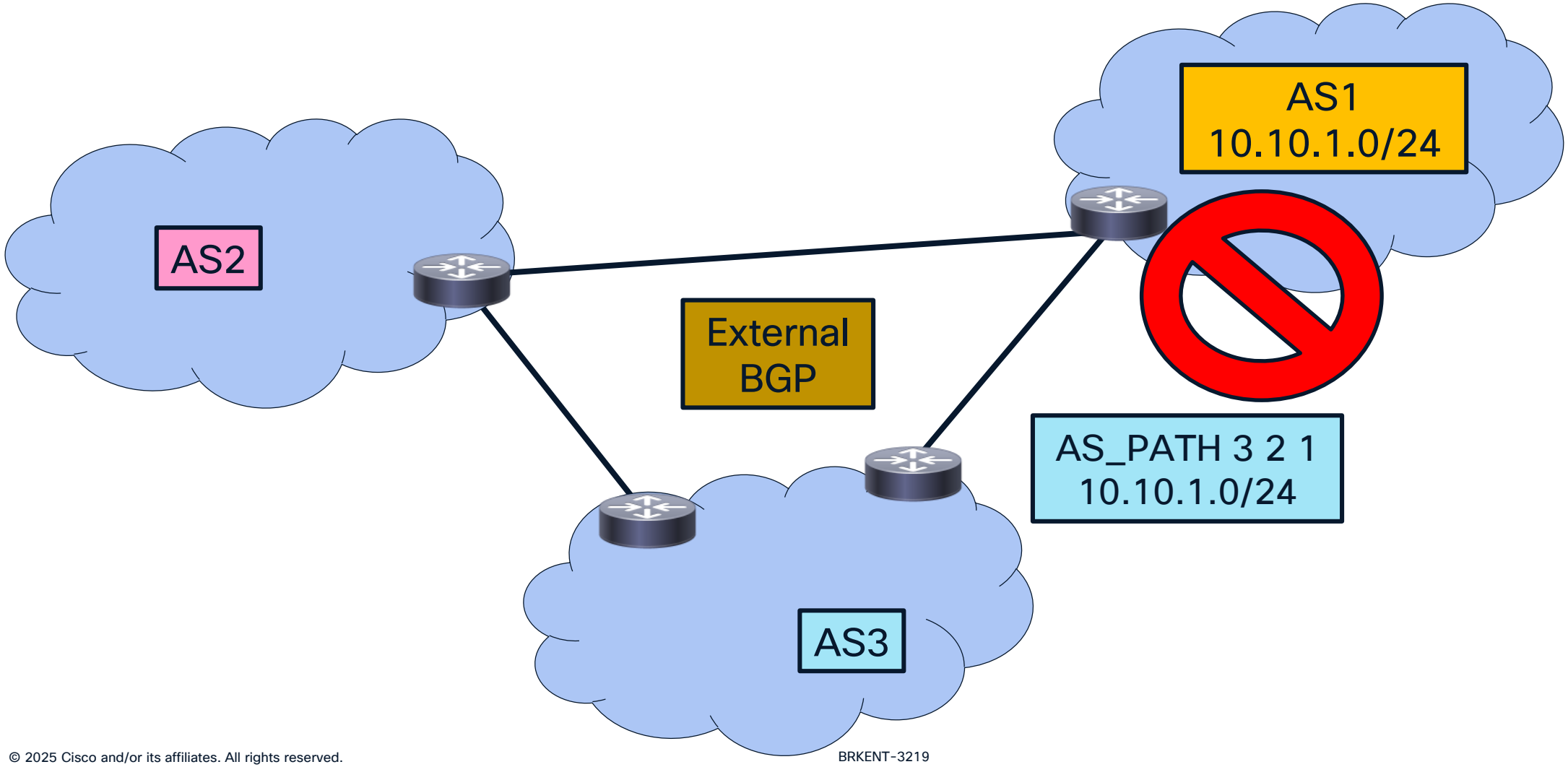
Loop prevention in BGP



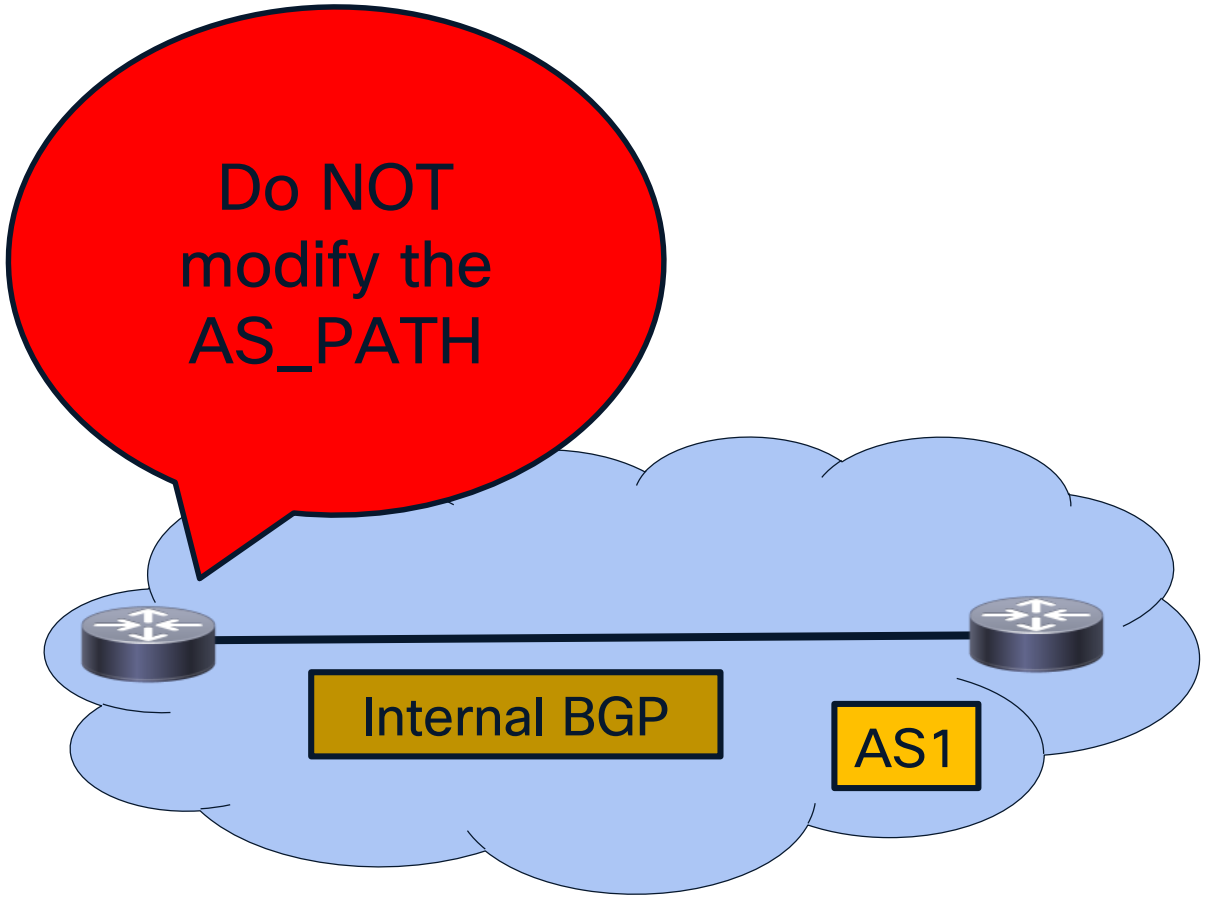
Loop prevention in BGP



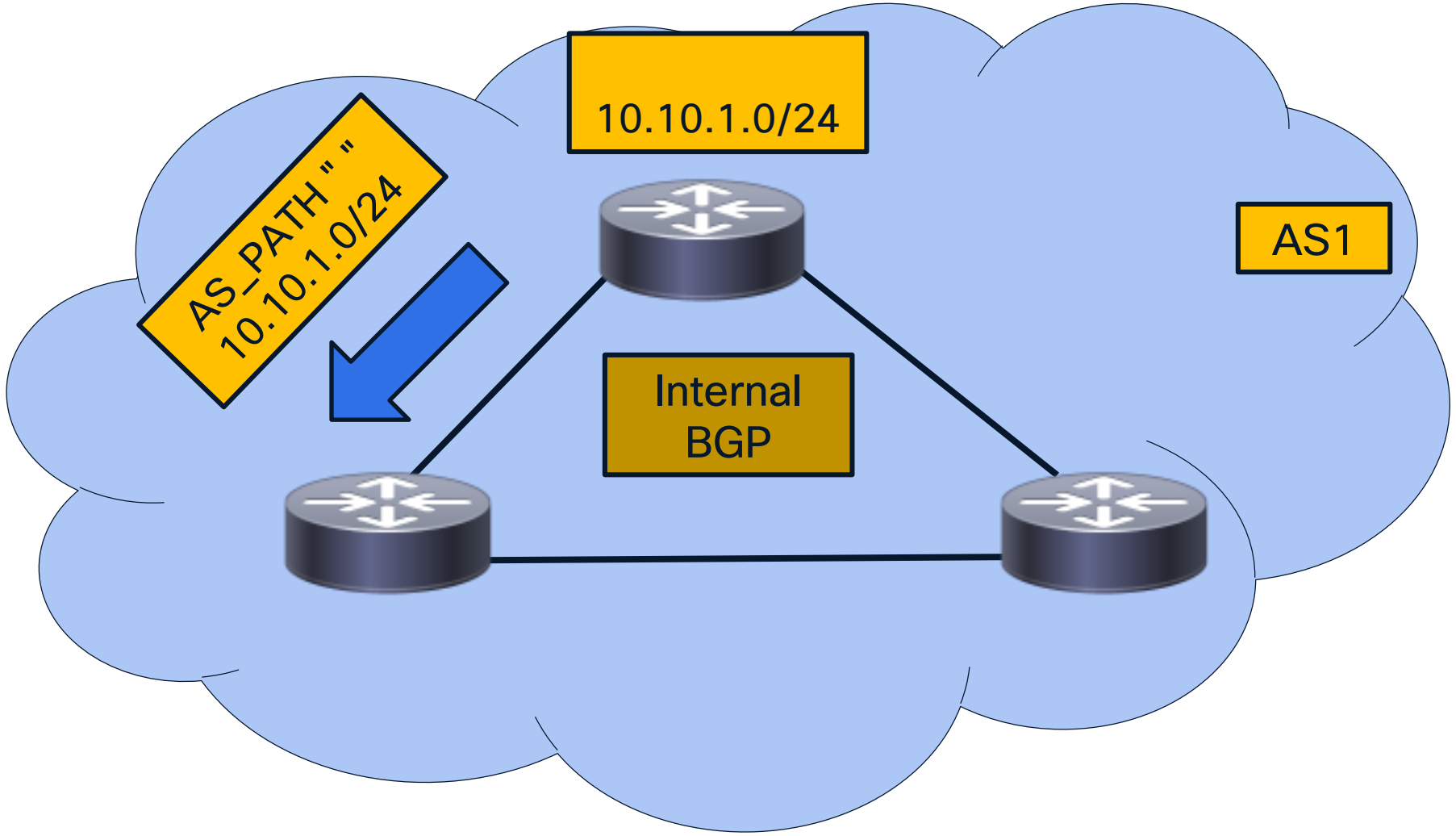
Loop prevention in BGP



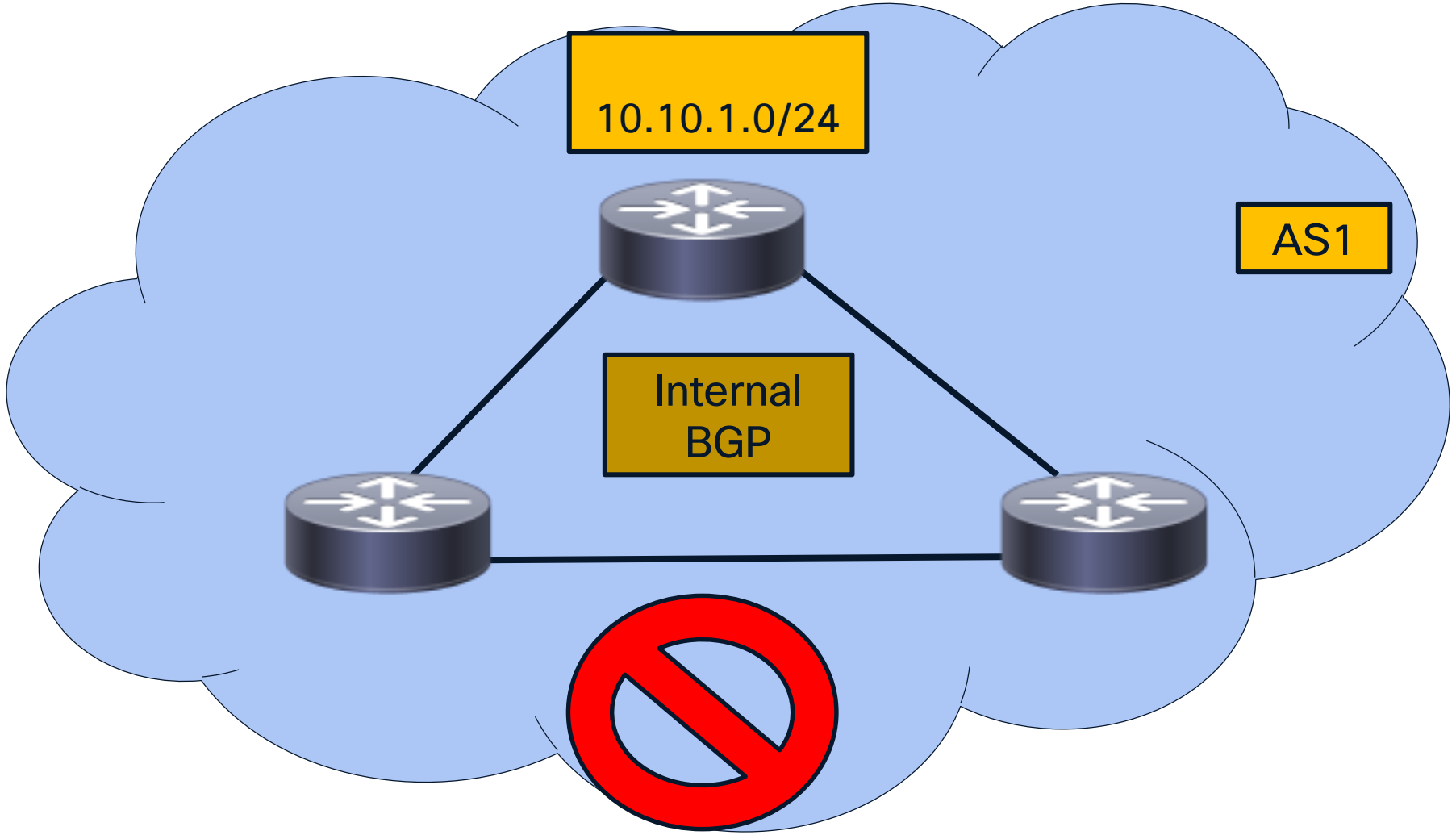
Loop prevention in BGP



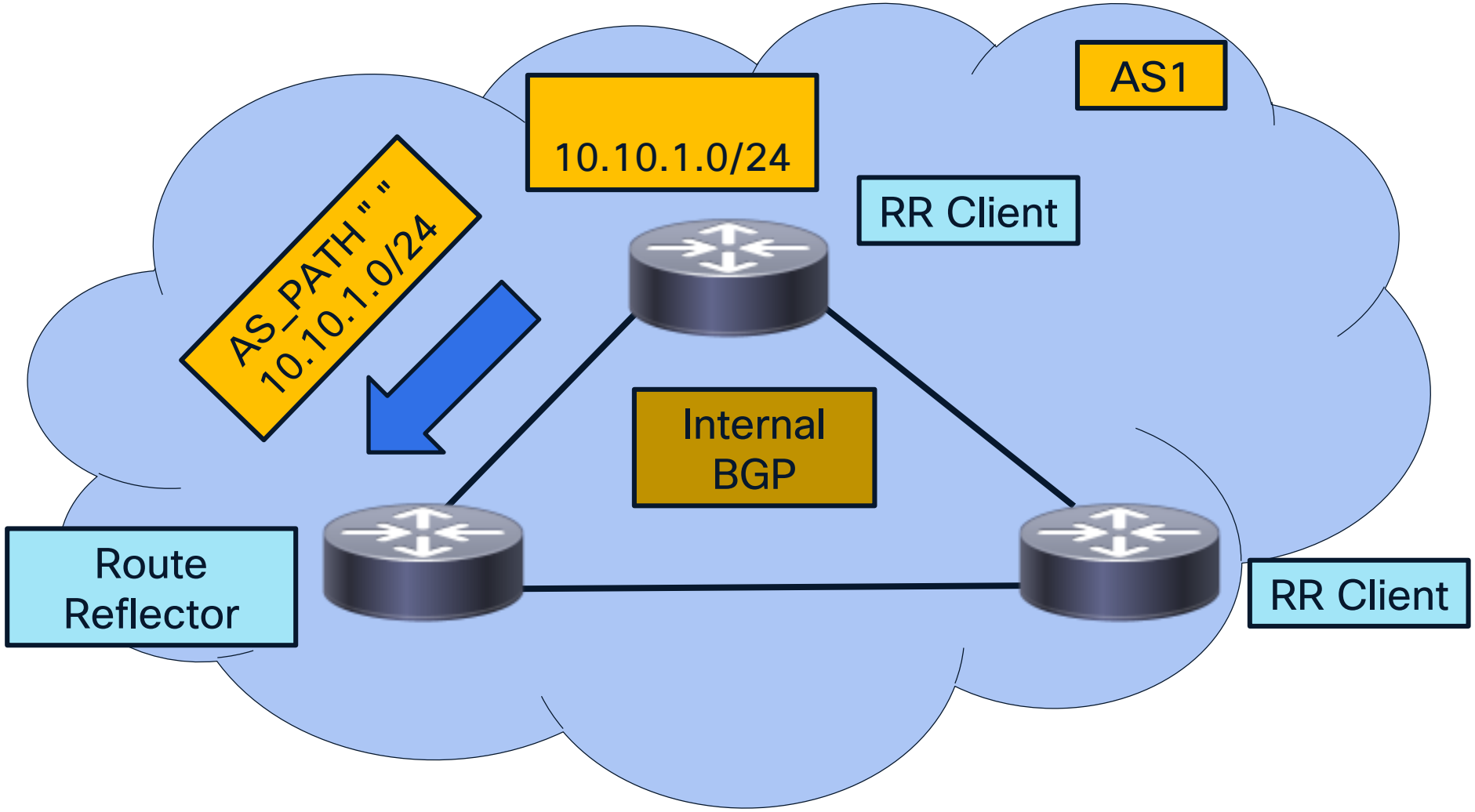
Loop prevention in BGP



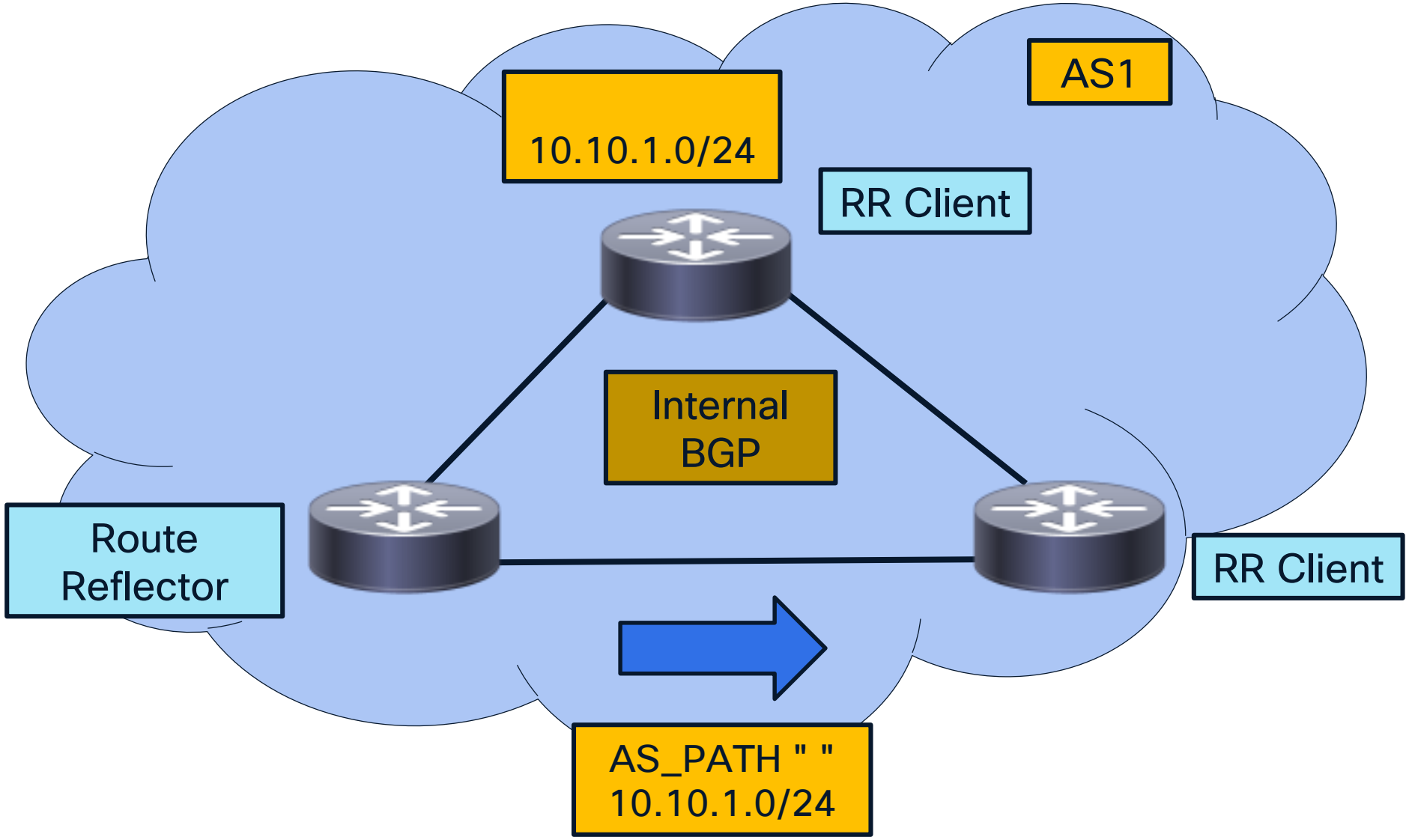
Loop prevention in BGP



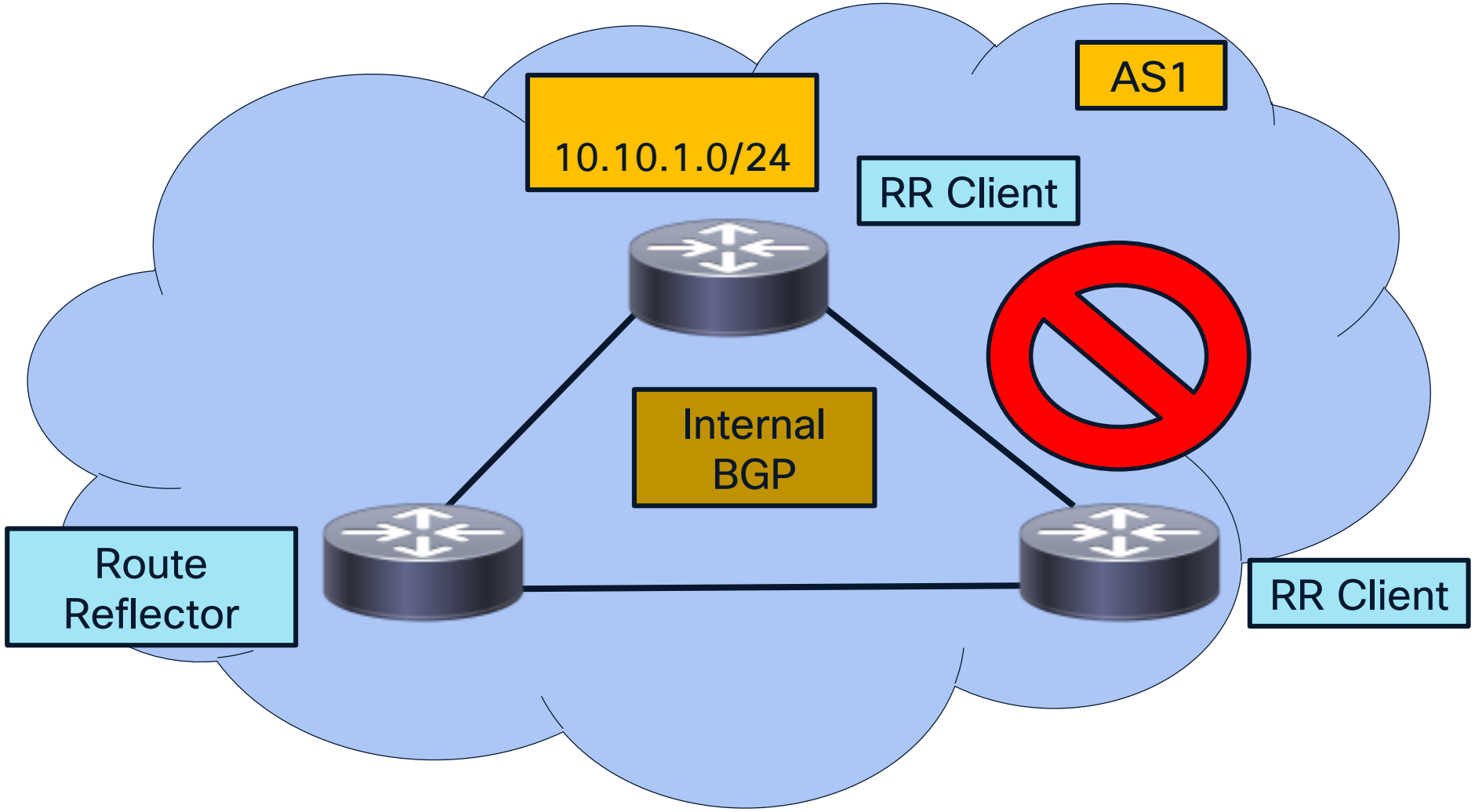
Route Reflection



Route Reflection



Route Reflection



Path Diversity

Advertising additional paths in BGP

- By default, BGP selects and advertises only a single best path
  - This is true even if configured to install multiple paths into local RIB
- This is a problem if we need our peers to be also aware of multiple paths to the same destination
  - Also impacts the speed of reconvergence in case one path fails
- Simply having a BGP speaker announce multiple paths to a peer would not solve the issue
  - Repeated advertisement of the same NLRI is processed as an update of the previous advertisement – the latter replaces the former

BGP Additional Paths

- RFC 7911 (AddPath) brings an extension to BGP allowing it to advertise multiple paths to a neighbor
  - This is accomplished by extending the NLRI with an additional field called the Path ID
- BGP speaker announcing multiple paths to a neighbor assigns a locally unique Path ID to every path
  - Since the Path ID is a part of NLRI, every path has a different compound NLRI and so does not appear as an update of the previous advertisement

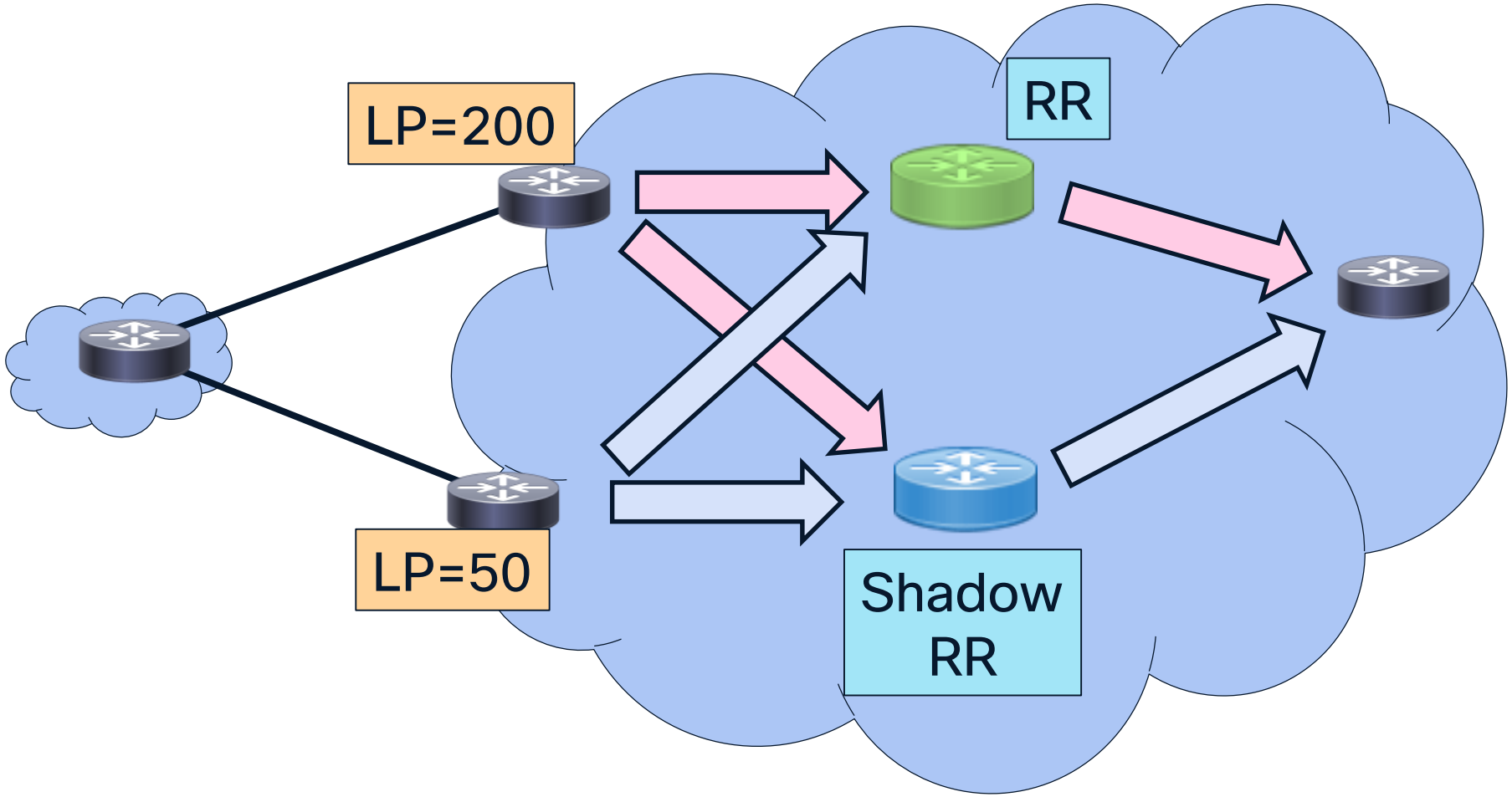
Selecting Best Additional Paths

- RFC 7911 does not dictate how the best paths should be selected
- Different BGP implementations may use vastly different algorithms
- IOS / IOS XE
  - First two / First three best paths
  - All paths with unique next hops
  - Optionally selected per-neighbor-AS

BGP Diverse Path

- As simple as it appears, BGP Additional Path is a fairly complex mechanism to implement well
- As a simplified approach specifically targeting route reflectors, RFC 6774 brings the Diverse Path facility
- BGP Diverse Path idea:
  - Have an additional (shadow) route reflector alongside the regular one, and have it advertise a different route than the best one
  - There is no change to BGP messaging, only to the best path selection and advertisement policy on the shadow route reflector

BGP Diverse Path Operation



Conclusion

“BGP only gets better with age”

Complete your session evaluations

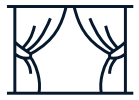
Complete a minimum of 4 session surveys and the Overall Event Survey to be entered in a drawing to win 1 of 5 full conference passes to Cisco Live 2026.

Earn 100 points per survey completed and compete on the Cisco Live Challenge leaderboard.

Level up and earn exclusive prizes!

Complete your surveys in the Cisco Live mobile app.

Continue your education



Visit the Cisco Showcase for related demos

Book your one-on-one Meet the Engineer meeting

Attend the interactive education with DevNet, Capture the Flag, and Walk-in Labs

Visit the On-Demand Library for more sessions at www.CiscoLive.com/on-demand

Contact us at: gsibaja@cisco.com / ppaluch@cisco.com

Thank you


CISCO