Segment Routing (SR) Introduction and Tutorial

Adrian Farrel (afarrel@juniper.net)

Take-Aways

- Some historical context
- What Segment Routing is trying to achieve
- Basic building blocks
- How it works in different SR environments
- Minimal control plane
- What you might do with SR and how you might deploy it

- Do not expect:
  - Too many details of how it works
  - Every possible use case or future application
  - A sales pitch (for the technology or for a vendor's solution)

TRAFFIC ENGINEERING

Purpose of TE

- TE is concerned with performance optimization of operational networks
- Control of how traffic flows through a network
- Optimise:
  - Amount of traffic the network carries
    - Traffic is money
  - Utilisation of resources
    - Resources cost money
  - The quality of service delivered
    - Bandwidth, latency,…
    - Reliability
- Avoid specific issues:
  - Planned maintenance
  - Suspect resources
  - Untrusted parts of the network

Strict Paths in TE

- Encode path information in the packet
  - Packet header enumerates every node in the path
  - No path information stored in the network
  - Example: IPv4 with Strict Source Routing Option
    - Not much used
- Store path information in the network
  - Packet header contains exactly one path identifier
  - No further path information is encoded in the packet
  - Example: RSVP-signaled MPLS

Loose Paths in TE

- Path is divided into segments
  - Segment contains one or more router hops
- Packet header lists each segment that the packet traverses
  - But it does not necessarily enumerate every node
- Network contains enough state to forward the packet through multi-node segments
- Examples
  - IPv4 Loose Source Routing Option
    - Not much used
  - IPv6 Routing Extension Header
    - Not much used

Segment Routing: A New Approach to TE

- Path information is placed in the packet header
- No control plane signaling or state
- History
  - Ideas first brought to the IETF in 2013
  - Source Packet Routing in Networking (SPRING) working group
    - Chartered October 2013
    - 500 members of the mailing list
  - Only one RFC on Segment Routing so far
    - RFC 7855 – Problem Statement
  - Around 50 Internet-Drafts in progress
    - Input from all the big vendors and a lot of the big operators

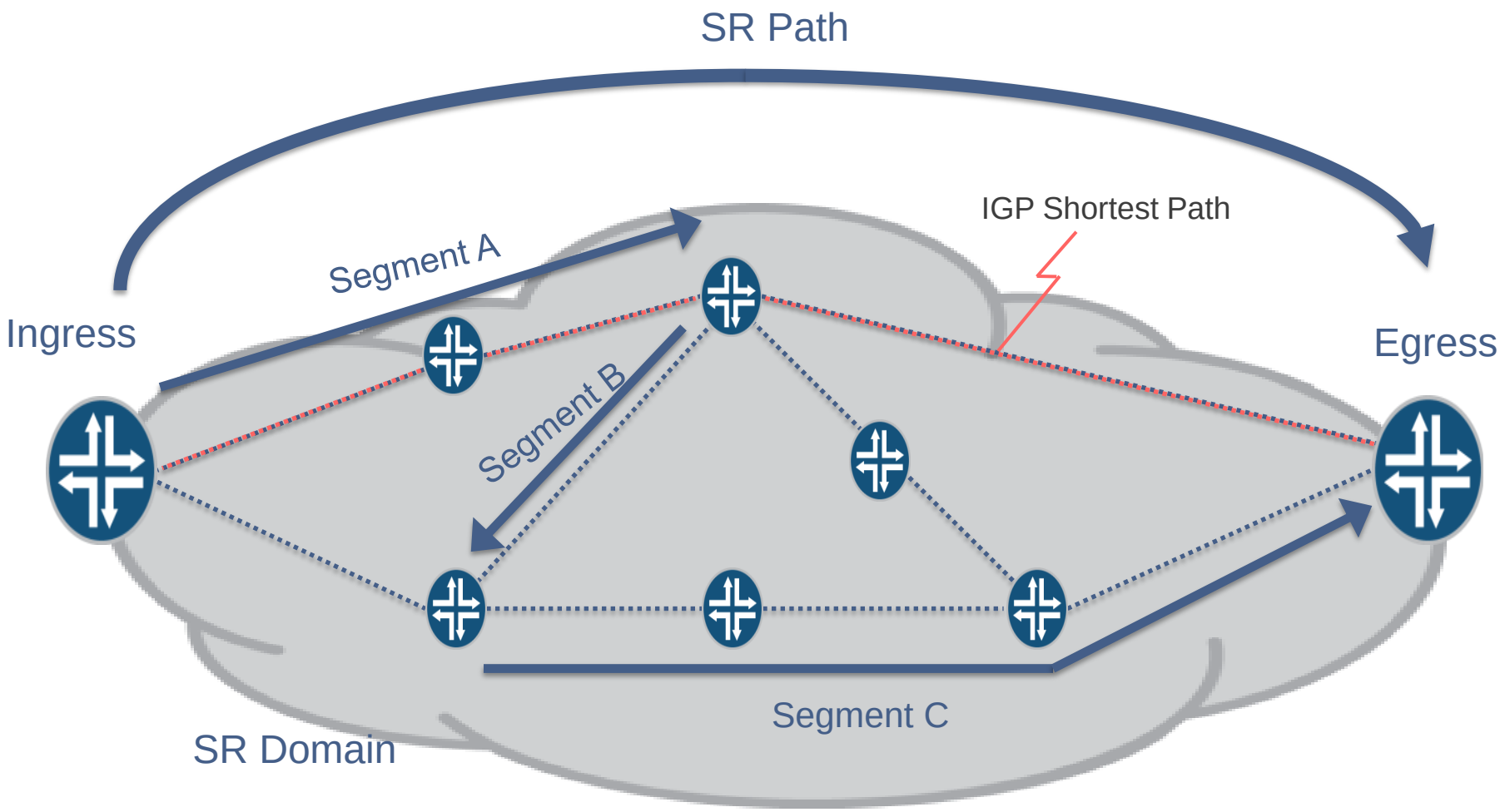
Segment Routing Objectives

- Deliver simple TE in packet networks
- Leverage shortest path forwarding
- Steer packets away from shortest paths for TE reasons
  - Load balance in the network
  - Create disjoint end-to-end paths
  - Repair after failure
- Achieve this without complexity in the network
  - Remove signaling protocols and associated state
  - Leverage existing forwarding paradigms (IP and MPLS)
  - Leverage existing routing protocols (IGPs and BGP)

Terminology

- *SR Domain* - A collection of SR capable devices
  - Roles: Ingress, transit, egress
- *SR Path* - Can be different from least cost path
  - Contains one or more *SR Segments*
- *SR Segment* - Connects two points in SR domain
  - Can traverse one or more router hops
  - Is represented by a Segment Identifier (SID)
- *Segment Identifier (SID)*
  - Node-local or domain-wide (a.k.a., global) significance

Pictorial Terminology



Basic Segment Types

- Adjacency (single router hop)
  - Represents an IGP adjacency
- Prefix (one or more hops)
  - Represents IGP least cost path to a prefix
- Anycast (one or more hops)
  - Represents IGP least cost path to a non-unique prefix
- Binding
  - Represents a tunnel (e.g., RSVP-signaled LSP)

The SR TE Approach

- SR segments define different types of path
  - Some traverse one router hop
  - Some traverse multiple router hops
- SR header is inserted into each packet
  - Lists each segment that a packet traverses
  - But not necessarily each node
- Network contains enough information to route a packet through a multi-hop segment
  - This information is advertised by the IGP
    - Or installed some other way
      - Central controller with a southbound protocol
        - » Such as BGP-LU

Three Encapsulation Environments

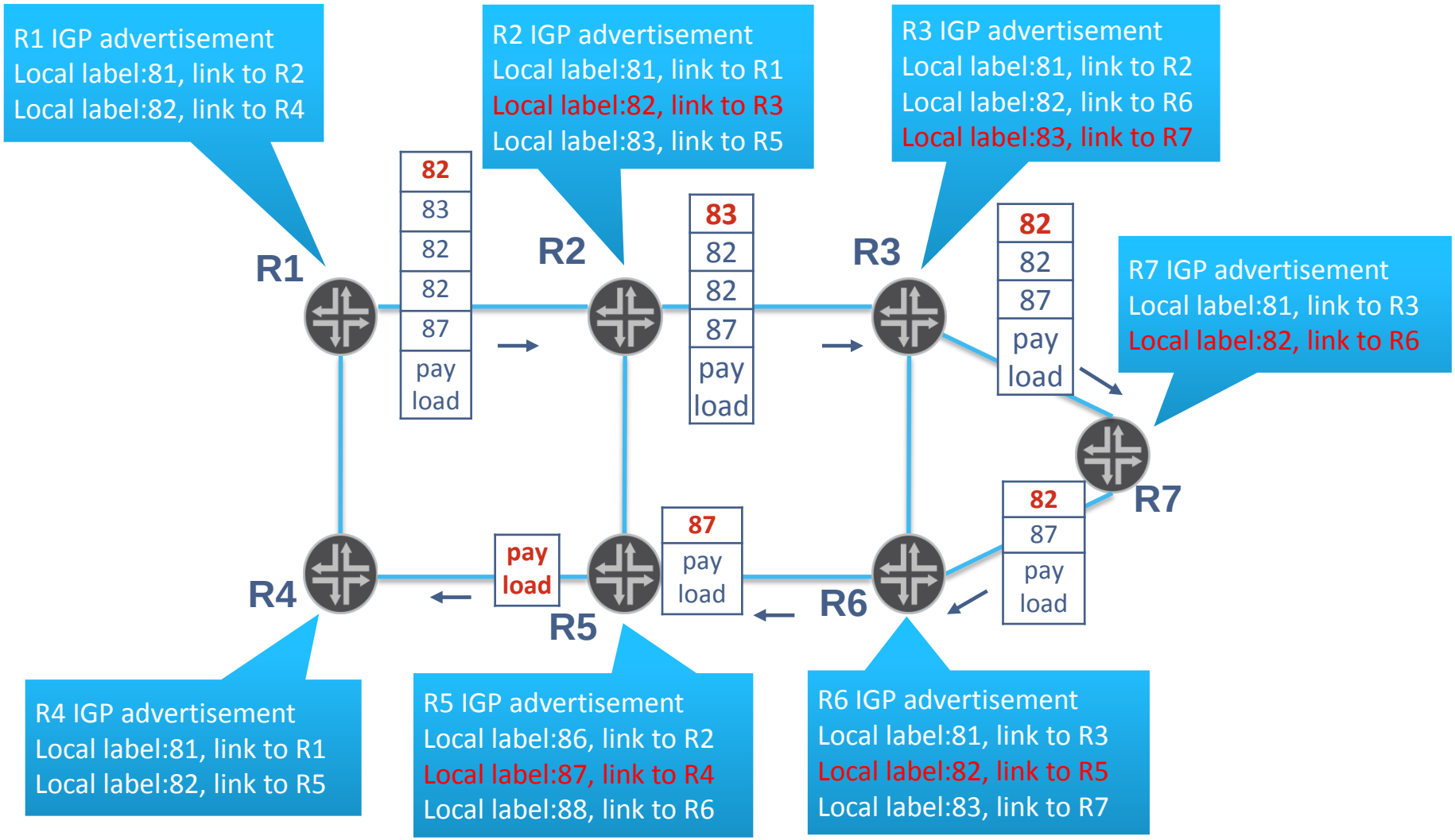
- MPLS
  - SR header is an MPLS label stack
  - Each label in the stack represents a segment
- IPv6
  - SR Header is an IPv6 header with a Segment Routing Extension Header (SRH)
  - SRH contains a list of IPv6 addresses
  - Each IPv6 address represents a segment
- MPLS-over-UDP
  - MPLS SR label stack encapsulated in UDP-over-IP
  - Routed through IPv4 or IPv6

MPLS FORWARDING

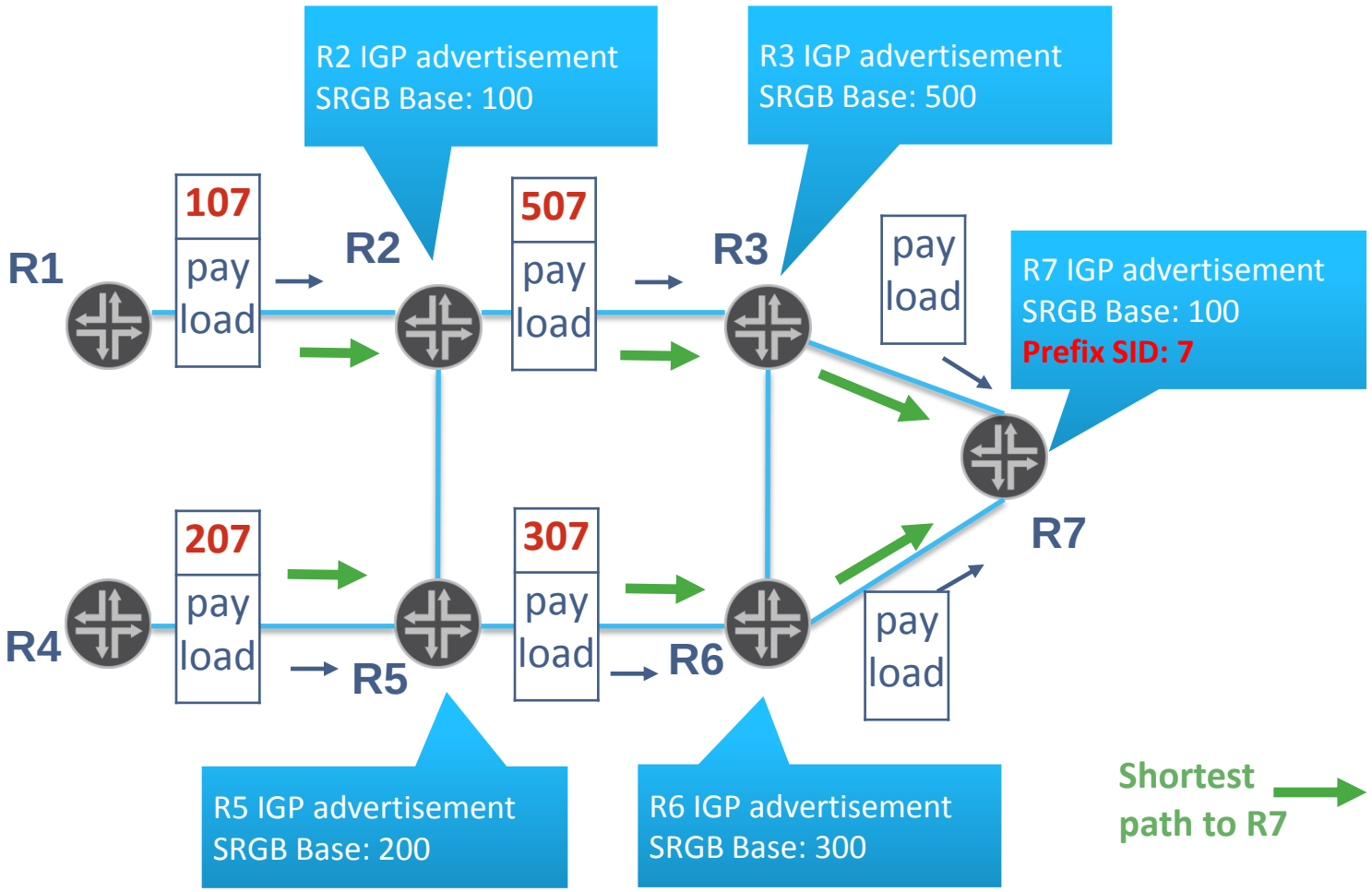
Local SIDs / Global SIDs

- SIDs are not labels
  - But they are encoded (carried) in labels
- Some SIDs have node-local significance
  - Nodes allocate local SIDs and to local labels
  - No need for domain-wide co-ordination
- Some SIDs have domain-wide significance
  - SIDs are allocated in a manner similar to that used for private IP (RFC 1918) addresses
  - Domain-wide coordination required (using the IGP)
    - Each node reserves a block of labels
      - The SR Global Block (SRGB)
    - Global label equals SRGB base value plus SID

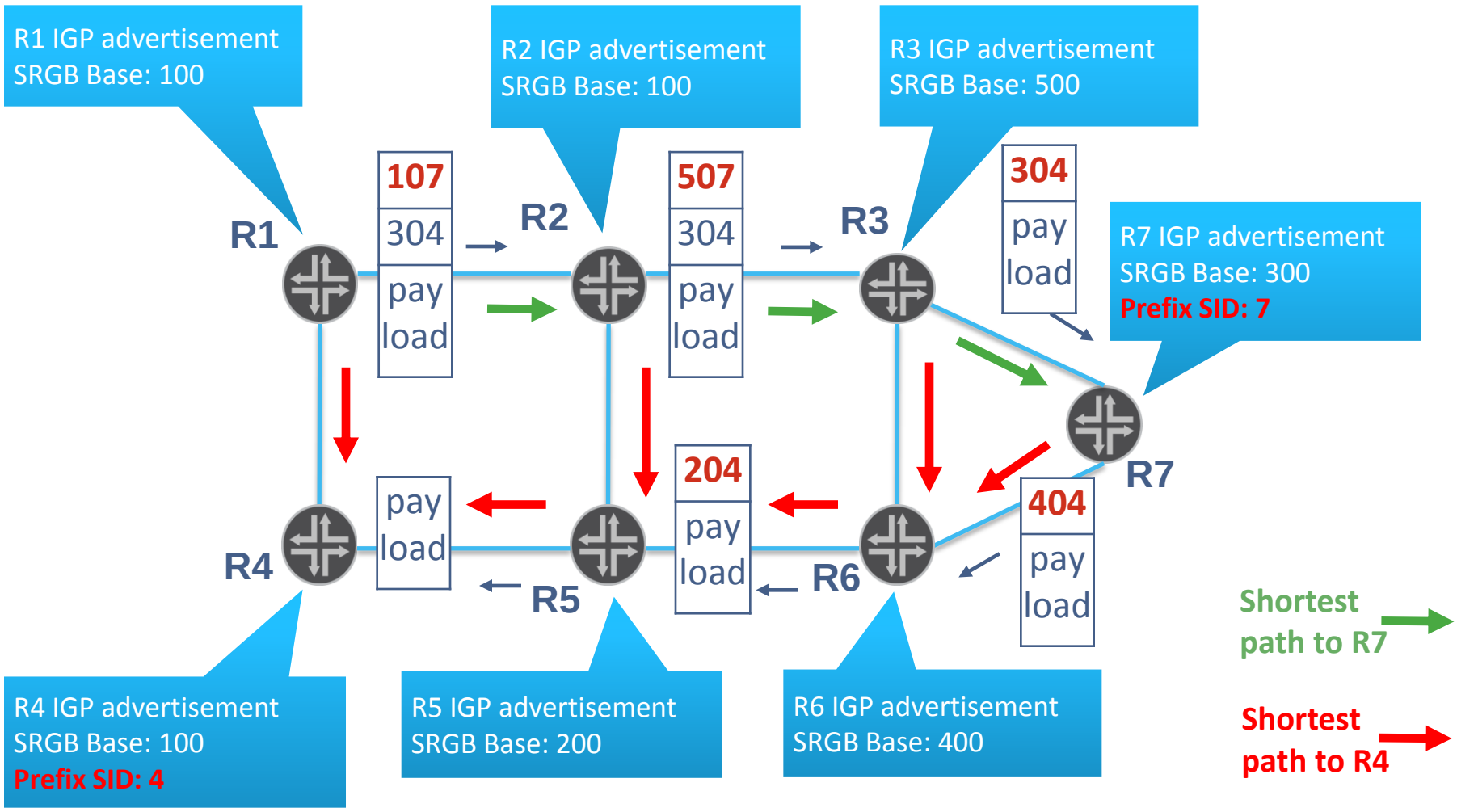
R1 to R4 : Adjacency Segments



Any Node to R7: Using Prefix Segment



R1 to R4 via R7 : Prefix Segment

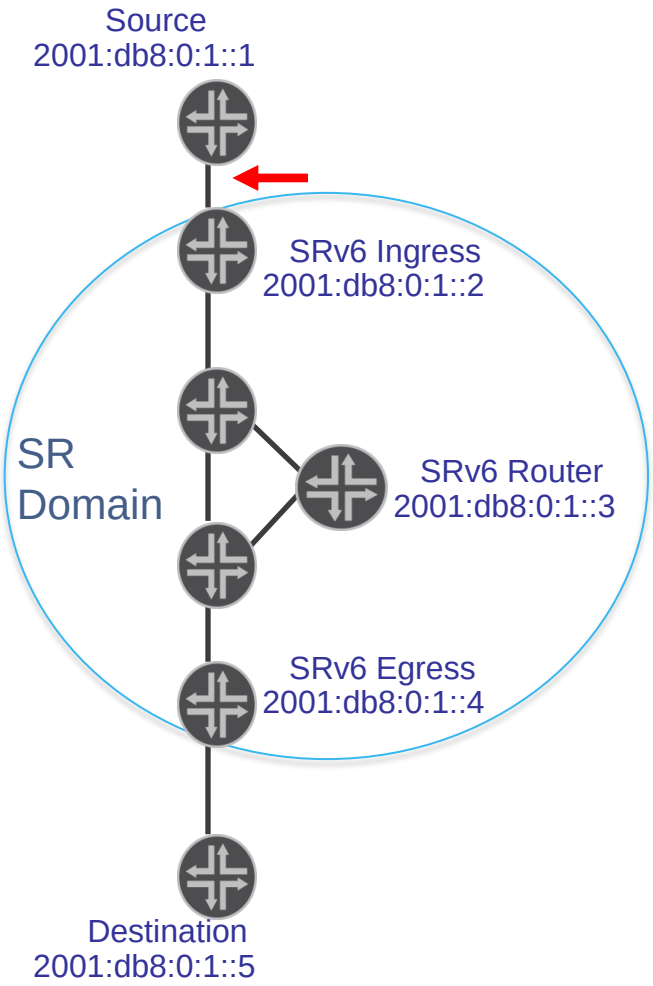


IPV6 FORWARDING

Modes

- Encapsulating mode
  - SR ingress router encapsulates payload packet in an IPv6 header
  - Source node includes a routing extension header between the IPv6 header and payload
  - (This is what the specs say)
- Simplified mode
  - SR ingress inserts a routing extension header between the payload IPv6 header and payload data
  - (This is what the prototypes implement)
- In both cases:
  - The routing extension header carries the stack of SIDs

Segment Routing Header (SRH) : (1 of 6)

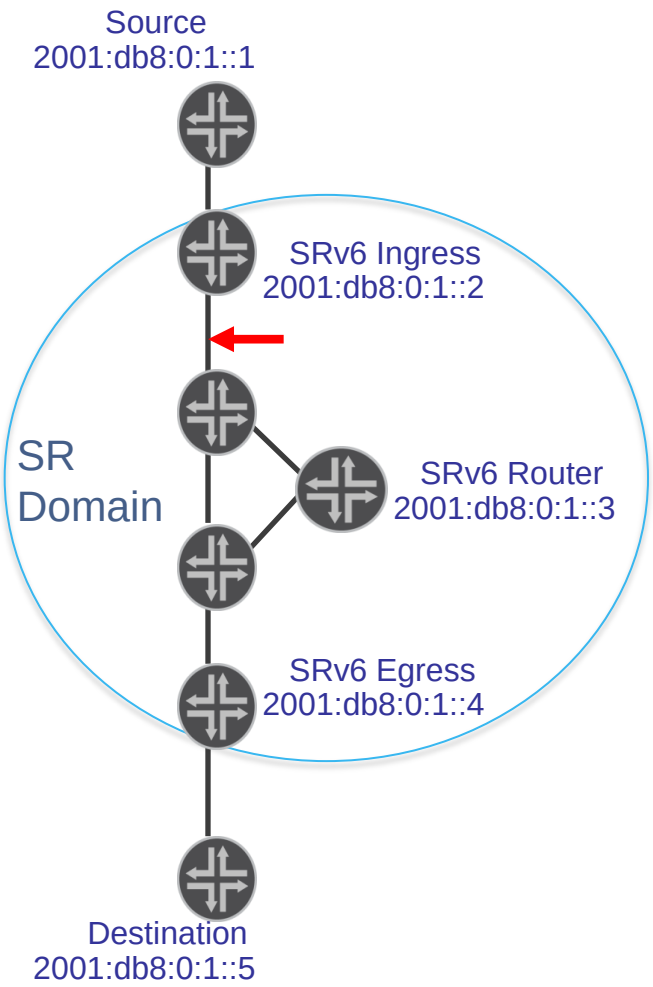


Ver	DSCP	Flow Label	
Payload Length 40		Next HDR TCP	Hop Limit 255
Source Address 2001:db8:0:1::1			
Destination Address 2001:db8:0:1::5			
TCP Header and Payload			

IPv6 HEADER

TCP Header/Payload

Segment Routing Header (SRH) : (2 of 6)



Ver	DSCP	Flow Label	
Payload Length 136		Next HDR SRH	Hop Limit 255
Source Address 2001:db8:0:1::2			
Destination Address 2001:db8:0:1::3			

IPv6 HEADER

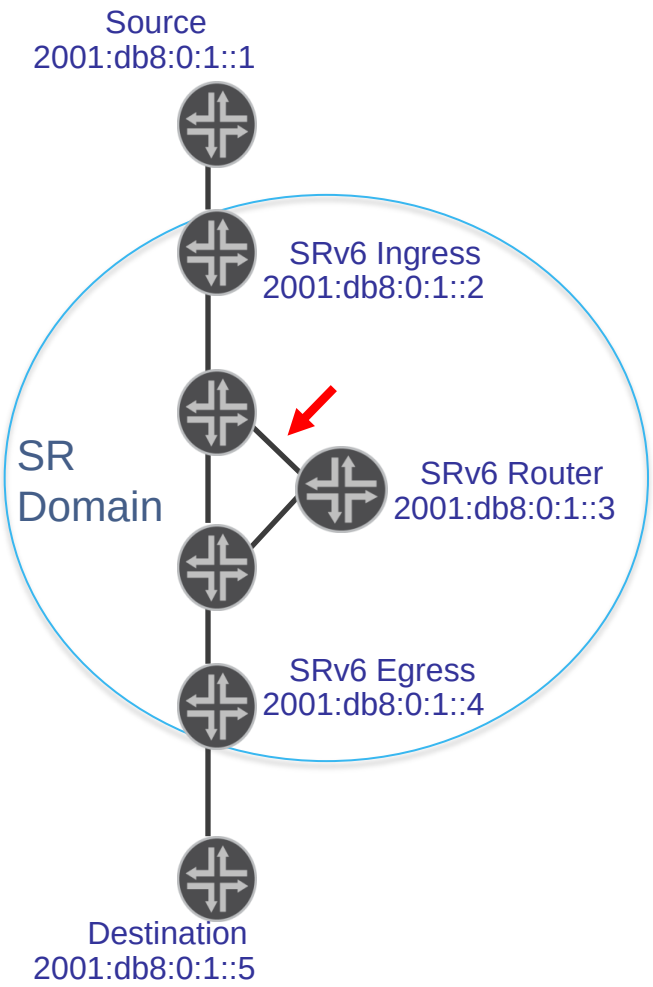
Next HDR TCP	Length 56	HDR Type 4	Seg Left 1
Last Entry 2	Flags		Tag
Segment 0 2001:db8:0:1::4			
Segment 1 2001:db8:0:1::3			
Segment 2 2001:db8:0:1::2			

Segment Routing Extension Header

Ver	DSCP	Flow Label	
Payload Length 40		Next HDR TCP	Hop Limit 255
Source Address 2001:db8:0:1::1			
Destination Address 2001:db8:0:1::5			
TCP Headed and Payload			

Payload (TCP/IPv6)

Segment Routing Header (SRH) : (3 of 6)



Ver	DSCP	Flow Label	
Payload Length 136		Next HDR SRH	Hop Limit 254
Source Address 2001:db8:0:1::2			
Destination Address 2001:db8:0:1::3			

IPv6 HEADER

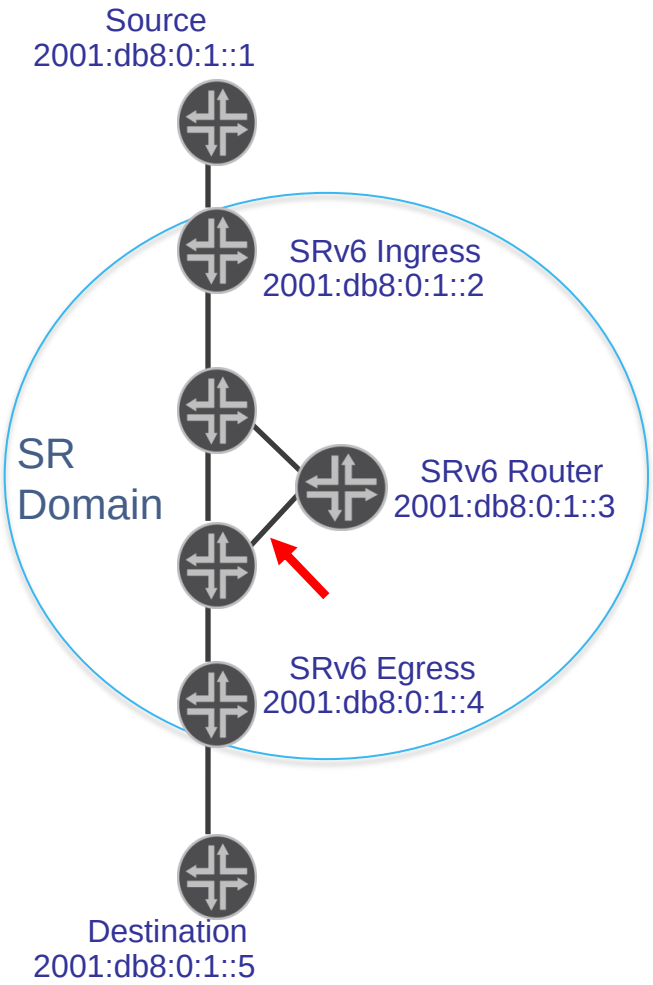
Next HDR TCP	Length 56	HDR Type 4	Seg Left 1
Last Entry 2	Flags		Tag
Segment 0 2001:db8:0:1::4			
Segment 1 2001:db8:0:1::3			
Segment 2 2001:db8:0:1::2			

Segment Routing Extension Header

Ver	DSCP	Flow Label	
Payload Length 40		Next HDR TCP	Hop Limit 255
Source Address 2001:db8:0:1::1			
Destination Address 2001:db8:0:1::5			
TCP Headed and Payload			

Payload (TCP/IPv6)

Segment Routing Header (SRH) : (4 of 6)



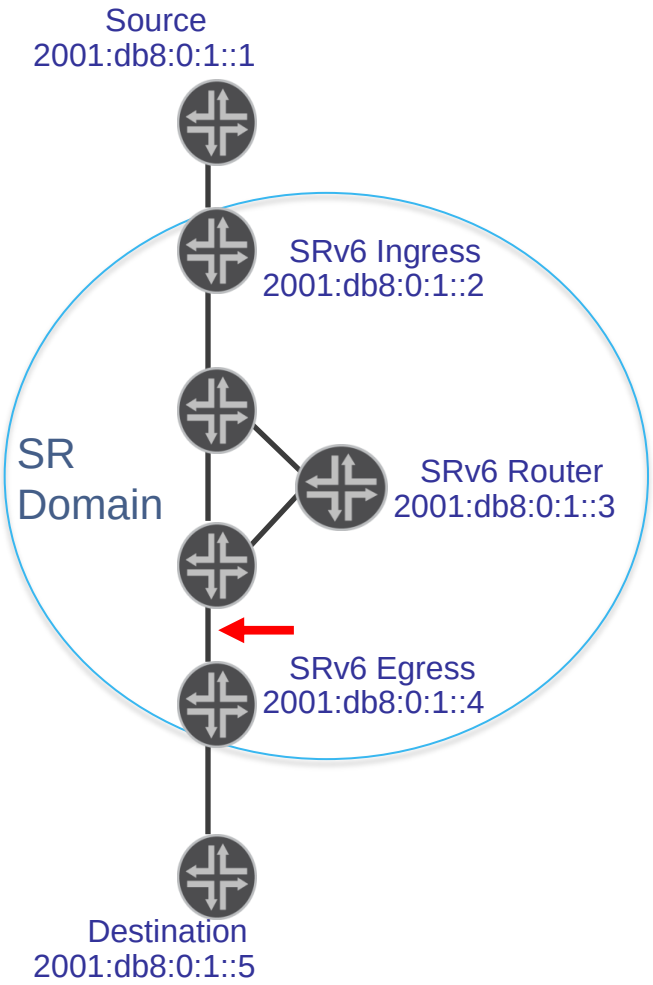
Ver	DSCP	Flow Label	
Payload Length 136		Next HDR SRH	Hop Limit 253
Source Address 2001:db8:0:1::2			
Destination Address 2001:db8:0:1::4			
Next HDR TCP	Length 56	HDR Type 4	Seg Left 0
Last Entry 2	Flags		Tag
Segment 0 2001:db8:0:1::4			
Segment 1 2001:db8:0:1::3			
Segment 2 2001:db8:0:1::2			
Ver	DSCP	Flow Label	
Payload Length 40		Next HDR TCP	Hop Limit 255
Source Address 2001:db8:0:1::1			
Destination Address 2001:db8:0:1::5			
TCP Headed and Payload			

IPv6 HEADER

Segment Routing Extension Header

Payload (TCP/IPv6)

Segment Routing Header (SRH) : (5 of 6)



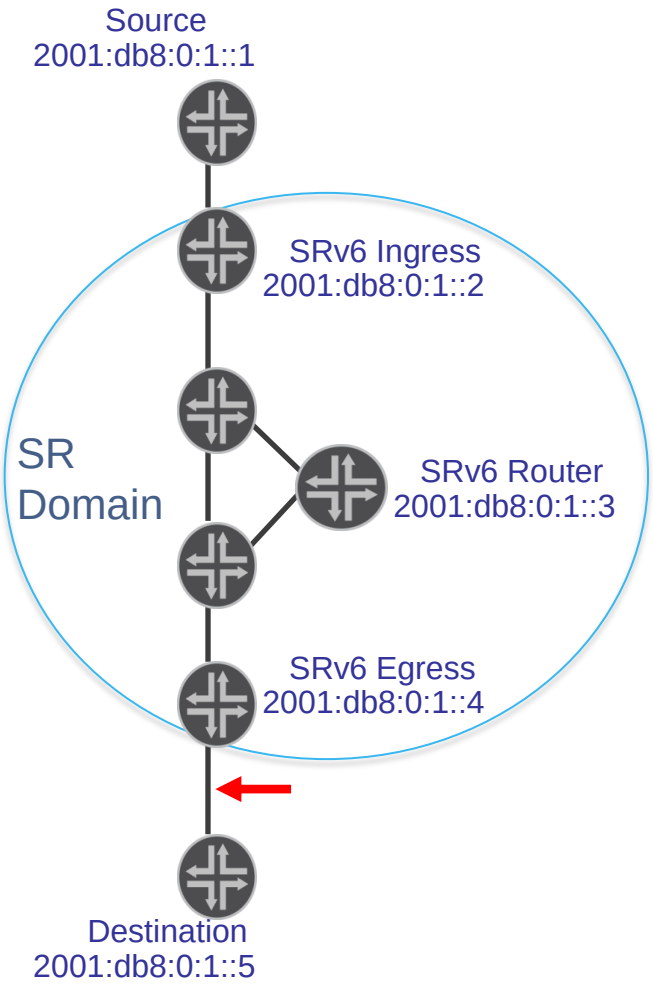
Ver	DSCP	Flow Label	
Payload Length 136		Next HDR SRH	Hop Limit 252
Source Address 2001:db8:0:1::2			
Destination Address 2001:db8:0:1::4			
Next HDR TCP	Length 56	HDR Type 4	Seg Left 0
Last Entry 2	Flags		Tag
Segment 0 2001:db8:0:1::4			
Segment 1 2001:db8:0:1::3			
Segment 2 2001:db8:0:1::2			
Ver	DSCP	Flow Label	
Payload Length 40		Next HDR TCP	Hop Limit 255
Source Address 2001:db8:0:1::1			
Destination Address 2001:db8:0:1::5			
TCP Headed and Payload			

IPv6 HEADER

Segment Routing Extension Header

Payload (TCP/IPv6)

Segment Routing Header (SRH) : (6 of 6)



Ver	DSCP	Flow Label	
Payload Length 40		Next HDR TCP	Hop Limit 254
Source Address 2001:db8:0:1::1			
Destination Address 2001:db8:0:1::5			
TCP Header and Payload			

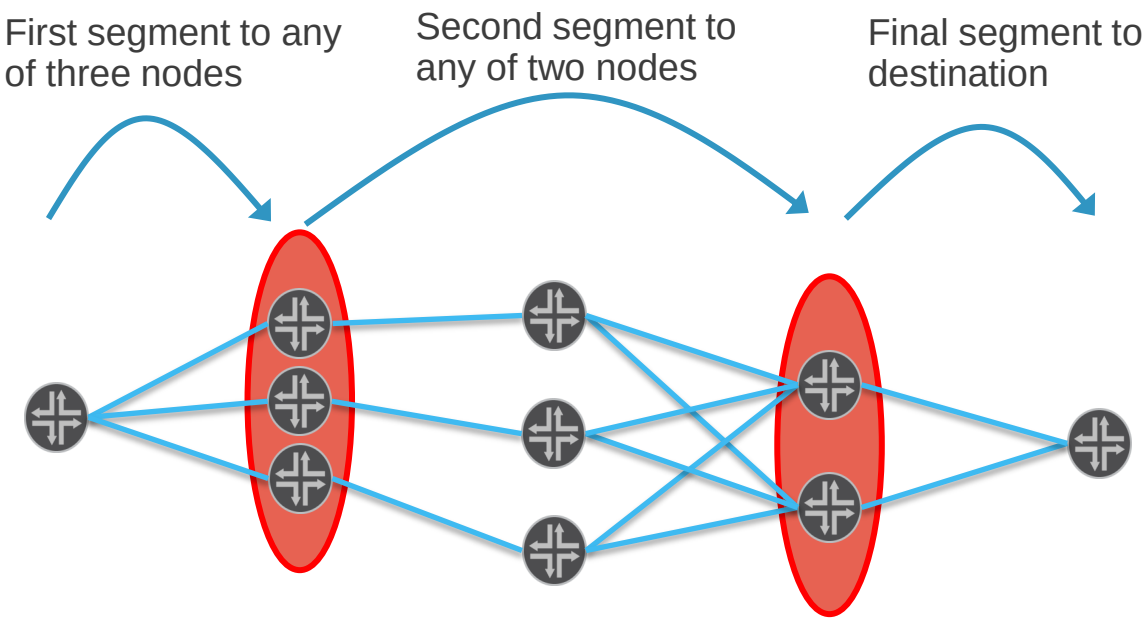
IPv6 HEADER

TCP Header/Payload

ADVANCED TYPES OF SID

Multiple Points of Presence

- An ***Anycast SID*** identifies a set of nodes via a non-unique prefix
- Choice is made as an IGP shortest path first to the nearest member of the prefix set
- May use ECMP
- Helps survive failures and allows load balancing
- Set of nodes are usually geographically close



Identifying SR Paths or Tunnels

- ***Binding SID***s are bound to (i.e., identify) other SR paths or tunnels
- This allows one SR path to include another SR path or a tunnel by reference
- If Binding SID identifies another SR path then the SR forwarding operation is:
  - Step beyond the Binding SID (decrement “Segments Left” or pop label)
  - Insert additional labels for the identified SR path
- If Binding SID identifies a tunnel then the forwarding operation is:
  - Step beyond the Binding SID (decrement “Segments Left” or pop label)
  - Encapsulate the packet and send it down the tunnel
- Useful for scaling the SID stack at the packet ingress
- Useful for traversing legacy networks

Benefits and Drawbacks to SRv6

- Segment routing is a very powerful concept
  - Many use cases and many advantages
- SRv6 header might be “quite large”
  - 16 bytes per SID
  - This causes MTU issues
  - Some silicon may face challenges and we want SR to be widely available
- Standardization issues around IPv6 header options mean that a real specification is still some way away
  - We ***need*** standards to ensure interoperability
- Two SR approaches will be expensive
  - Both have to be developed and tested even if you only buy one of them

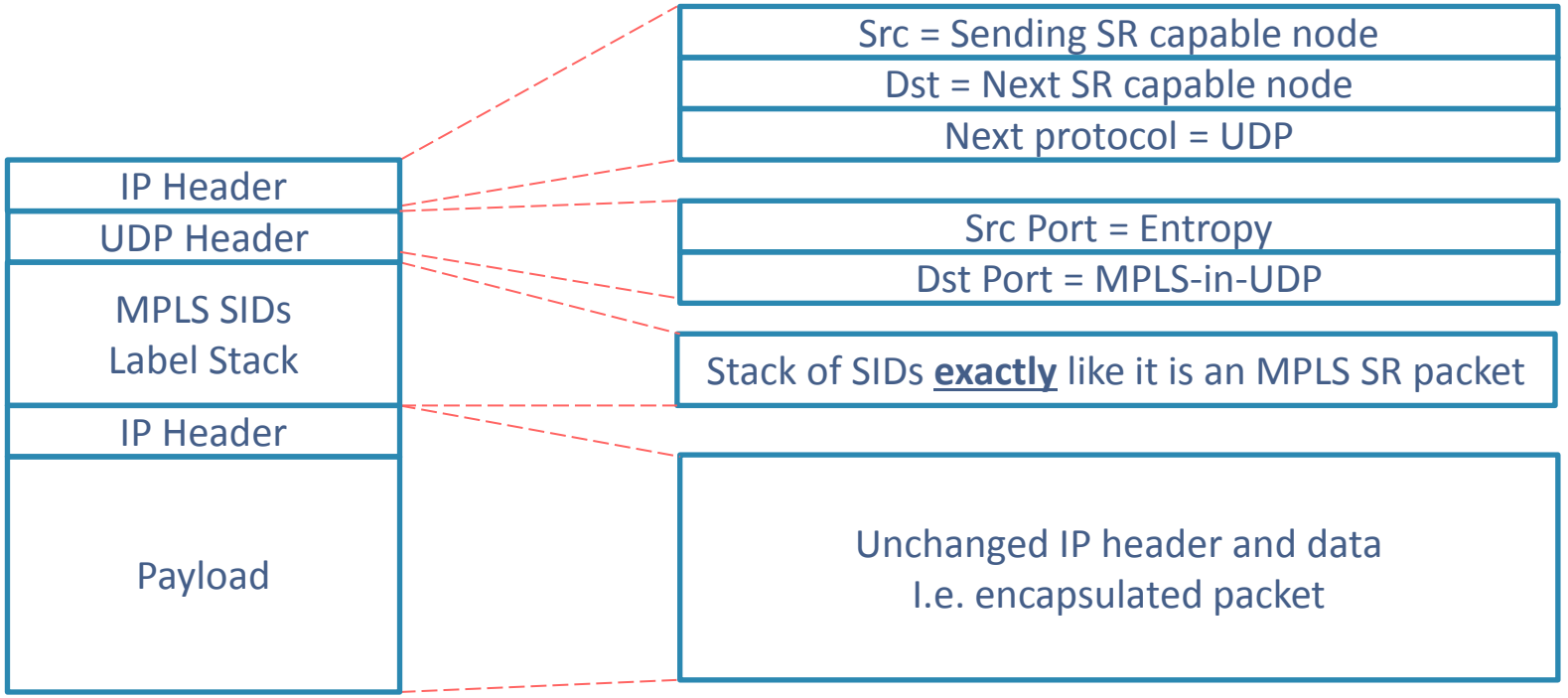
MPLS-SR-OVER-UDP

Main Objectives

- Get all of the Segment Routing function
- Tunnel MPLS-SR over an IPv6 network
  - E.g., connect two MPLS-SR data centers
- Slot into a native IPv6 network
  - Don’t need to use MPLS forwarding (some people don’t like it!)
  - Phased introduction with non-SR routers
- Simplify SRv6
  - Address the scaling concerns (reduce header size)
  - Avoid standardization controversy
  - Use existing technologies and mechanisms
  - Avoid expense/complexity of two SR solutions

MPLS-SR-over-UDP

- We already know how to carry MPLS over UDP (RFC 7510)
  - Very useful for “bridging” MPLS islands (such as data centers)
- New IETF work:
  - draft-bryant-mpls-unified-ip-sr
  - Carry MPLS-SR in UDP
  - Very simple way to get all the SR function in an IPv6 network
  - Get SR in IPv4 “for free”



MPLS-SR-in-UDP Processing

- IGP and control plane just like MPLS-SR
- Source processing is just like MPLS-SR
  - But encapsulate in UDP and IP to first router identified by first SID
- Legacy transit nodes
  - It is just an IP packet, so simply forward it
- SR-capable transit nodes
  - Process MPLS-SID stack as normal
  - Encapsulate in UDP and IP and send to router identified by next SID
- Final hop just strips outer header and forwards payload packet

SR CONTROL PLANE

SR ❤ IGPs

- LSDB provides information required for CSPF computation
- LSDB provides information required to create SR FIB entries
- ISIS and OSPF have been enhanced to flood SR information throughout the IGP domain
- SR requires an IGP and little else!

Path Computation

- SR ingress imposes label stack for the path
- Someone has to work out the path to use
  - It's the normal TE problem
- Performed on SR ingress router
- Or on central controller
  - PCE-based
- Or imposed by operator as config

FIB Creation (MPLS)

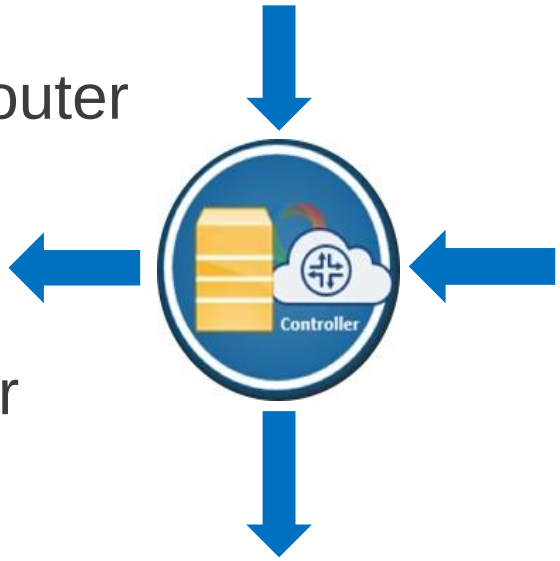
- On each node, for each global SID
  - Create a FIB entry that swaps the label (if required) and forwards through the IGP shortest path
- On each node, for each local SID
  - Create a FIB entry that pops a label and forwards through the correct link
- LSDB provides information for FIB creation
- (SRv6 forwarding is just IPv6 forwarding)

Central Controller

- Benefits
  - Central control has global view of reserved bandwidth
    - Not available at any other point in the network
  - Facilitates analytics driven policy
    - Controller receives telemetry
    - Based on telemetry, controller alters policy
- Risks
  - Concentrated point of failure / congestion
  - Potential performance bottle neck
  - Risks mitigated by redundant controllers
    - May require some form of synchronization

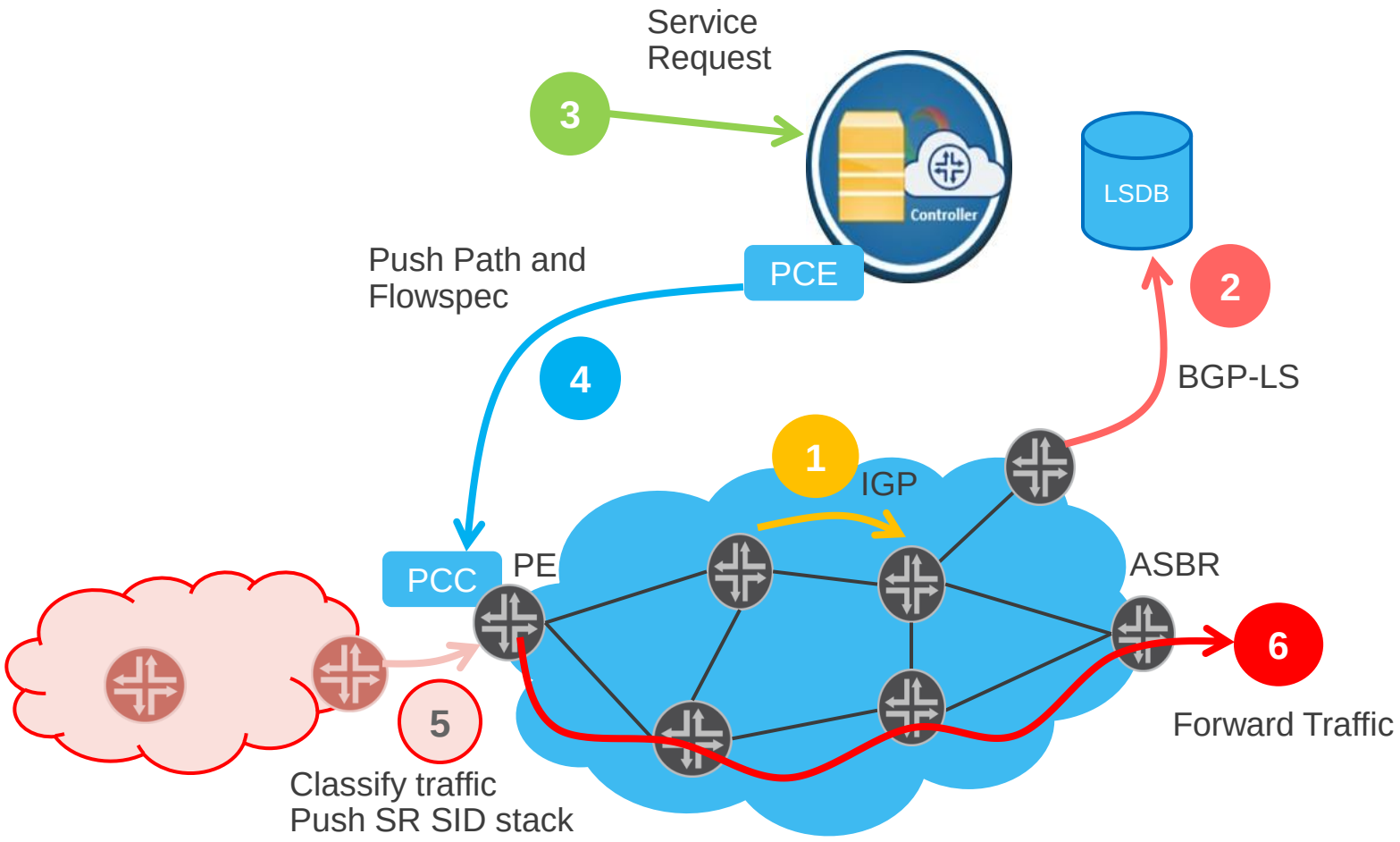
Controller Protocol Options

- Controller acquires LSDB
  - Controller participates (passively) in IGP
  - BGP-LS exports LSDB to controller
- Controller sends segment list to ingress router
  - PCEP
  - BGP
  - Programmable RPD
- Controller imposes policy at ingress router
  - What traffic to place on a SR path
  - Flowspec additions to PCEP or BGP



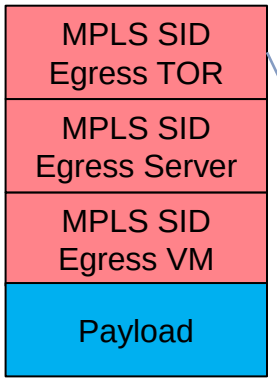
SOME USE CASES FOR SR

PCE With Segment Routing



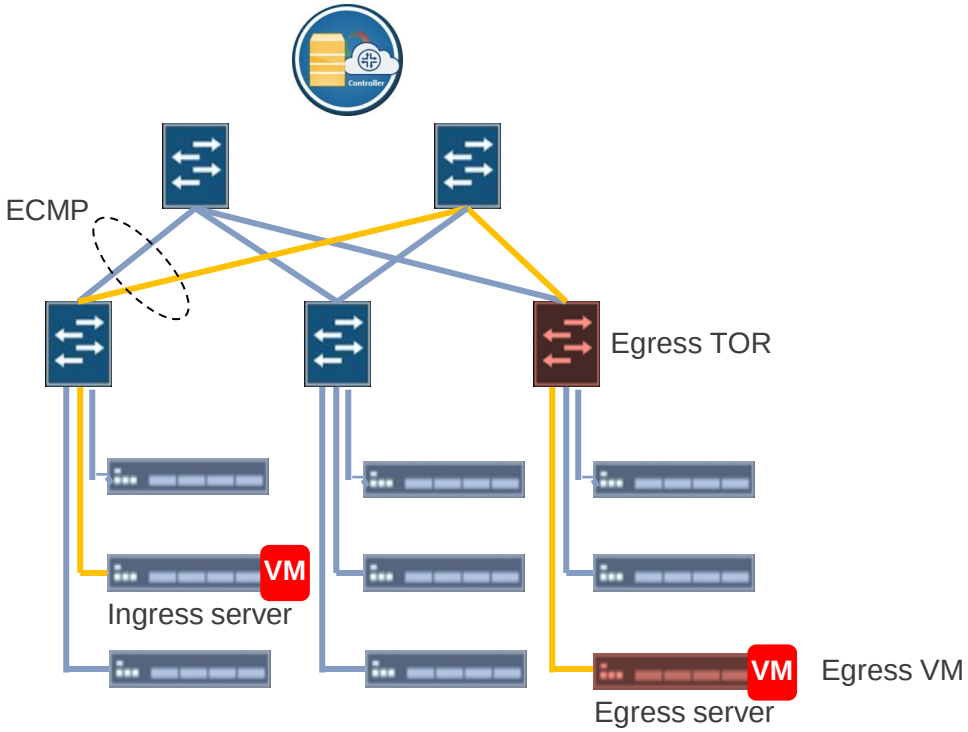
Segment Routing in the Data Center

- BGP-LU used to coordinate SIDs/labels in the network
- Controller builds paths
- Controller pushes paths
- Controller programs egress server



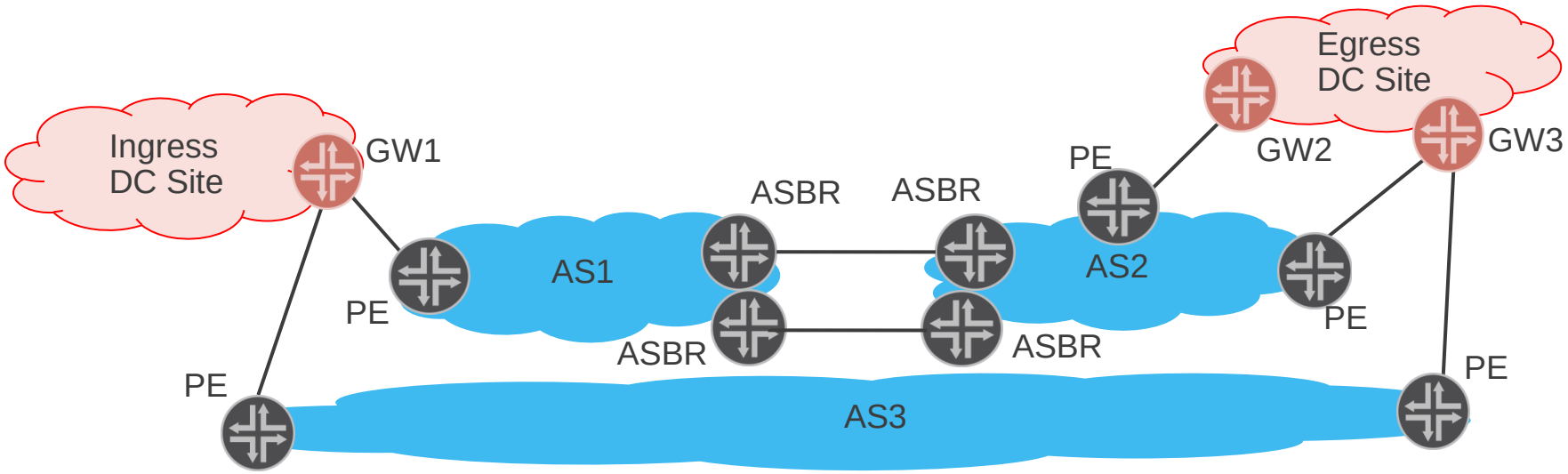
Transmitted packet

"Loose hop" load balances over ECMP through spine switches

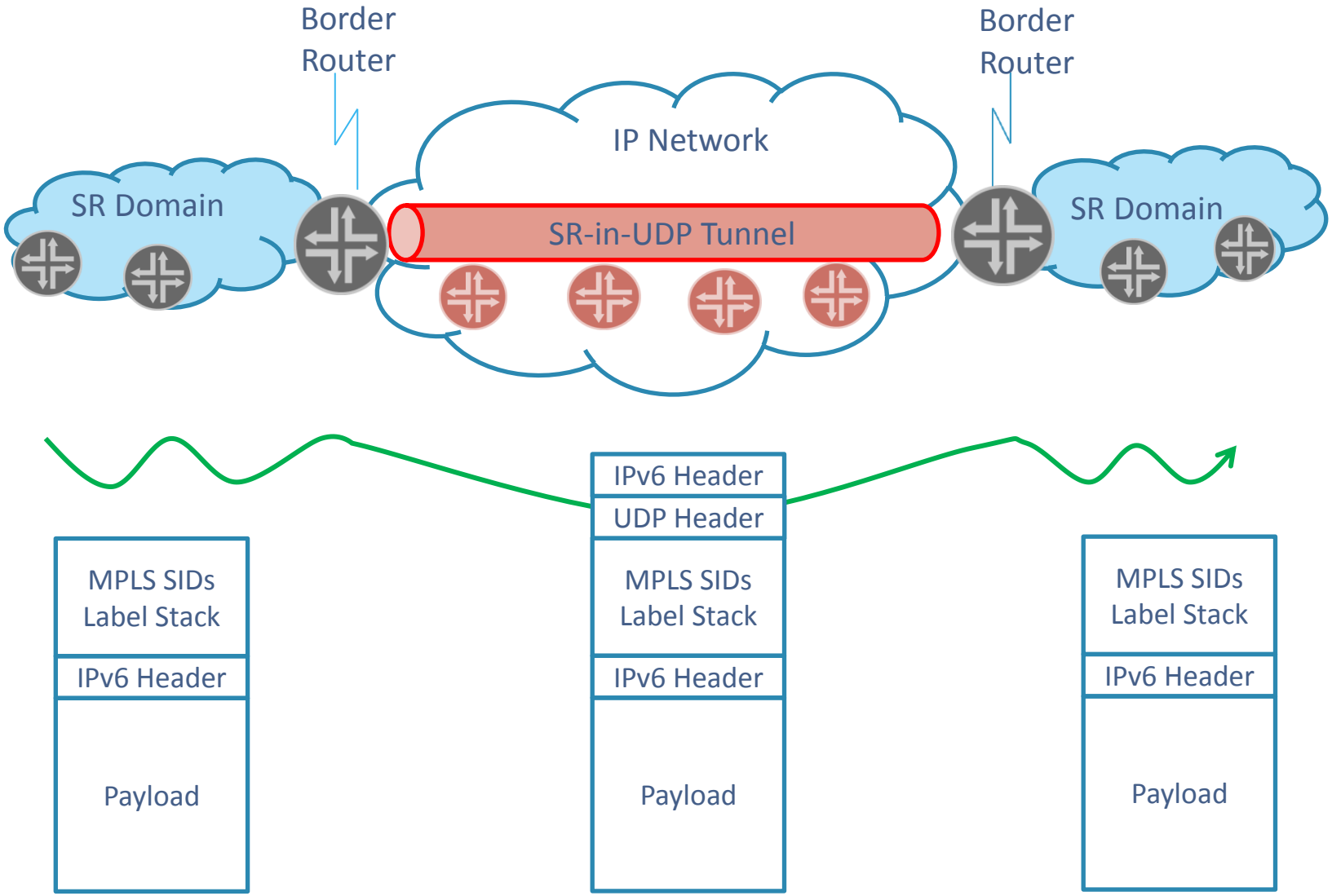


Routing Between Data Centers

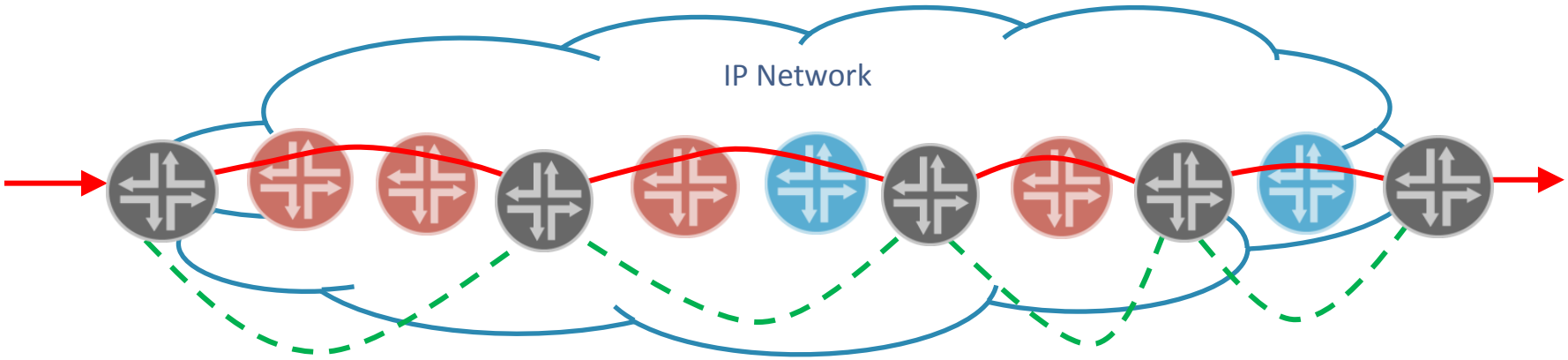
- draft-drake-bess-datacenter-gateway
- draft-farrel-spring-sr-domain-interconnect
- Gateways already advertise reachability to prefixes in the DC sites
- Gateways already advertise Tunnel Encapsulation attributes
- Two new features
  - All gateways advertise on behalf of all other gateways
    - Can now resolve dual homing paths
  - New “SR tunnel” type : Binding SID
    - Can now build end-to-end SR paths



Use Case : Tunnelling SR Across a Non-SR Core



Use Case : SR in a Mixed Mode IP Network



Legacy IP Router

SR-Capable Router Not In SID Stack

SR-Capable Router In SID Stack

Native IP Forwarding

SR Forwarding Path

CONCLUSION

Conclusion

- SR moves state from the network to the packet
  - Simplifies protocols
- Some problems remain to be addressed
  - OAM, Fast Reroute
- Optimisations on early proposals are possible
- Operational experience is required

Questions?

afarrel@juniper.net