Segment Routing

Santanu Dasgupta, Sr. Consulting Engineer
Clarence Filsfils, Distinguished Engineer

Aug 28 , 2013

Operators' Desire from the Network
Service Provider's are Under Pressure in the Zettabyte Era

- Simplicity
  - Less numbers of protocols to operate & troubleshoot
  - Less numbers of protocol interactions to deal with
  - Deliver automated FRR for any topology
- Scale
  - Avoid thousands of labels in LDP database
  - Avoid thousands of MPLS Traffic Engineering LSP's in the network
  - Avoid thousands of tunnels to configure

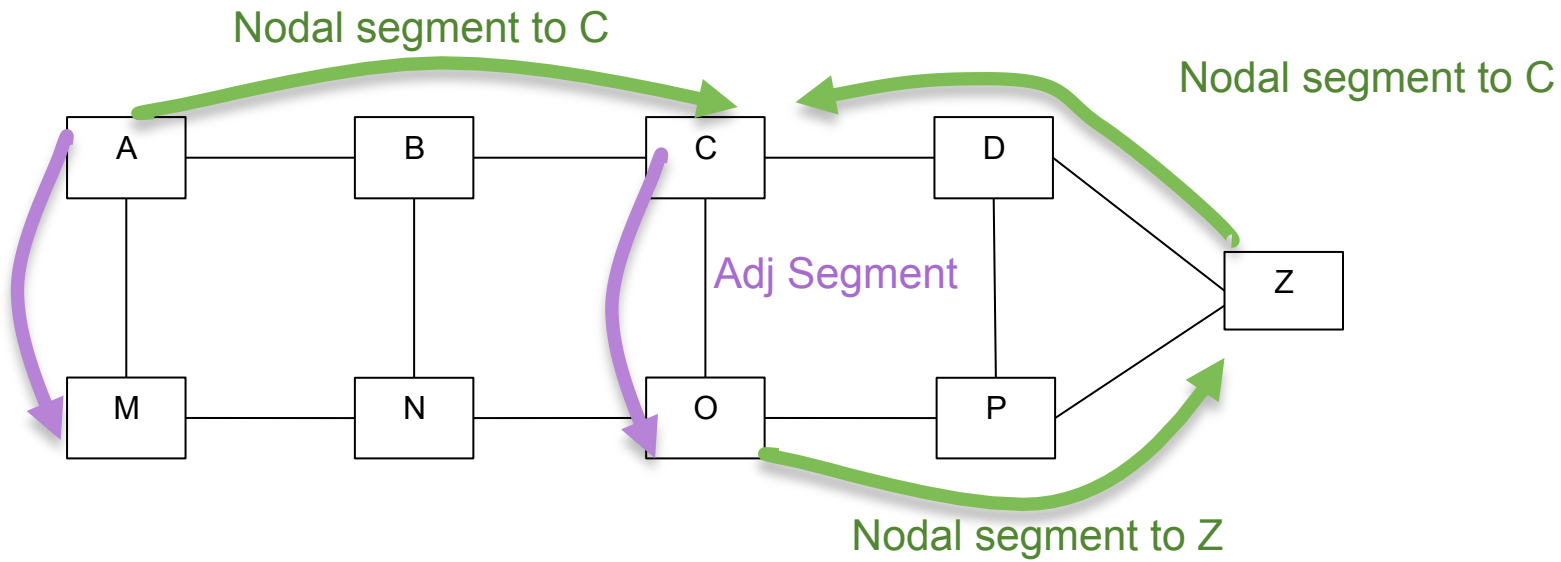
- Leverage all services supported over MPLS today (L3/L2 VPN, TE, IPv6)
  - Requires evolution and not revolution
- Bring the network closer to the applications
- IPv6 data plane a must, and should share parity with MPLS

Segment Routing – Introduction

- Forwarding state (segment) is established by IGP (ISIS or OSPF)
  - LDP and RSVP-TE are not required
  - Agnostic to forwarding dataplane: MPLS or IPv6
- MPLS Dataplane is leveraged without any modification
  - segment = label
  - push, swap and pop: all what we need
- Source Routing
  - Source computes the path and then encodes path as a label or stack of segments
- Architecturally designed to be Integrated with application
- Provide optimum scalability, resiliency, simplicity and virtualization

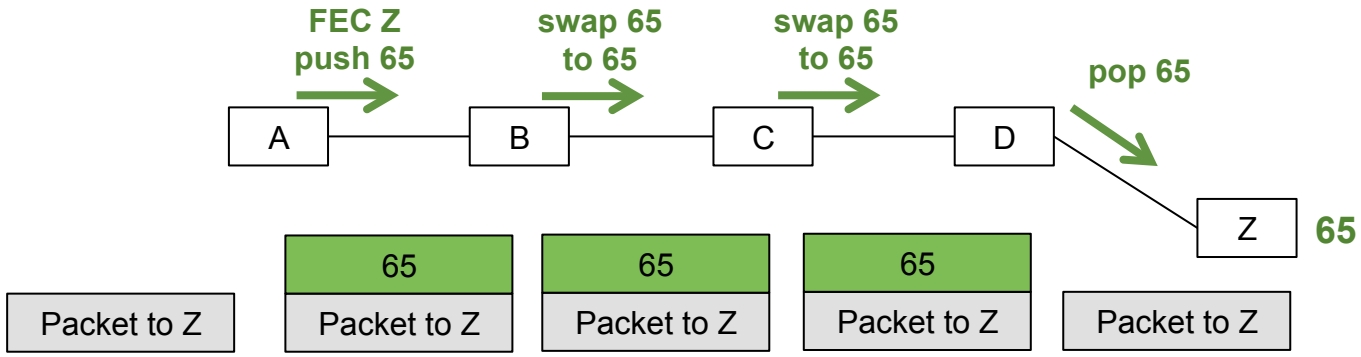
The State is No Longer in the Network, But In the Packet!

Segment Routing – Technology Basics
Simple Extension to IGP



- Simple extension to IS-IS or OSPF, automatically builds and maintains Segments
  - Nodal Segment – A Shortest path to the related node
  - Adjacency Segment – One hop through the related adjacency
- Excellent Scale: a node installs N+A FIB entries
  - N = nodal segments; A = adjacency segments

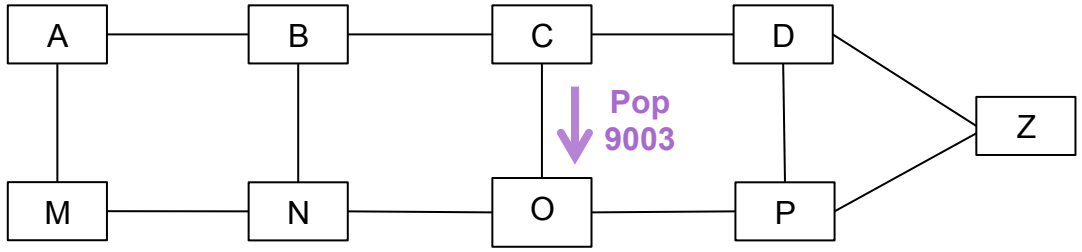
Nodal Segment



A packet injected anywhere with top label 65 will reach Z via shortest-path

- Node Z advertises its node segment (loopback 0)
  - For ISIS, its just a simple ISIS sub-TLV extension
- All remote nodes install the node segment to Z in the MPLS dataplane

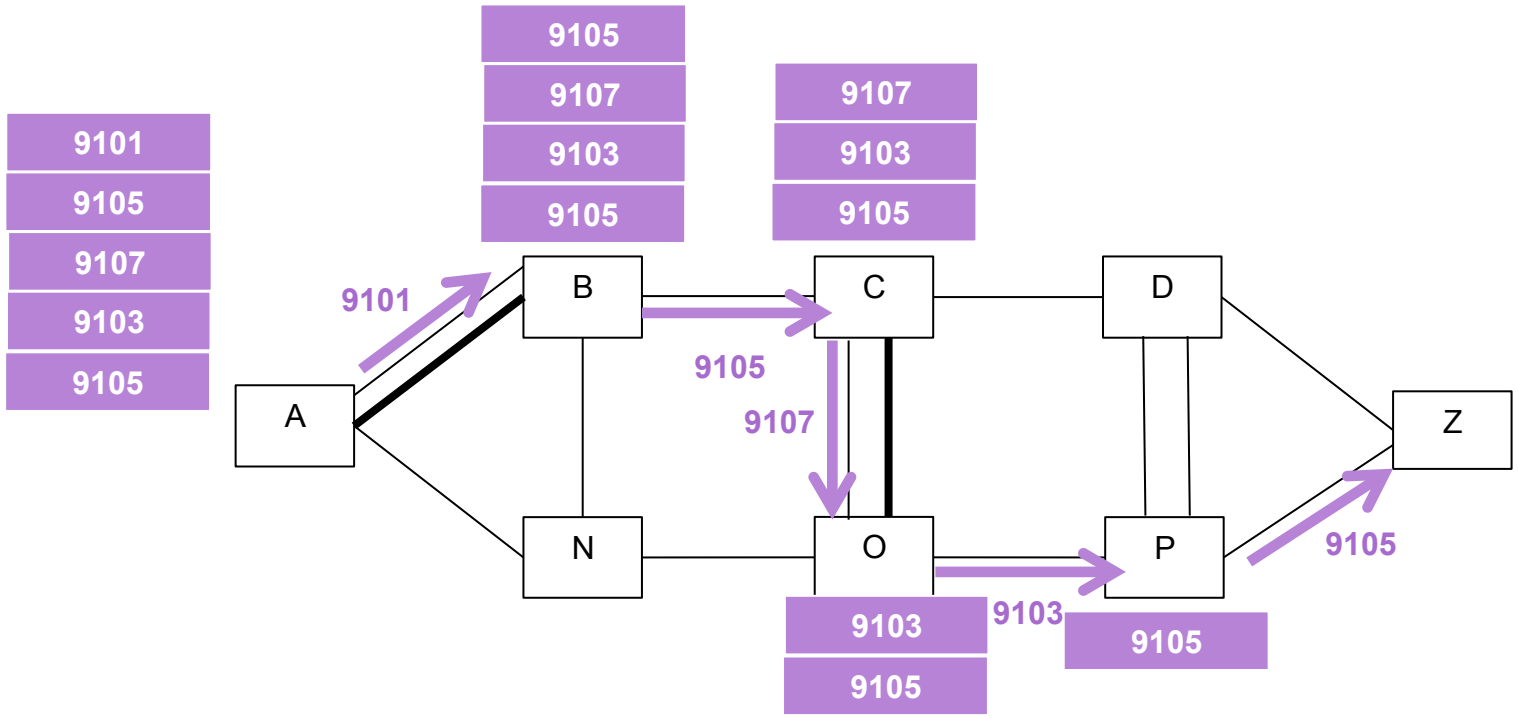
Adjacency Segment



A packet injected at node C with label 9003 is forced through datalink CO

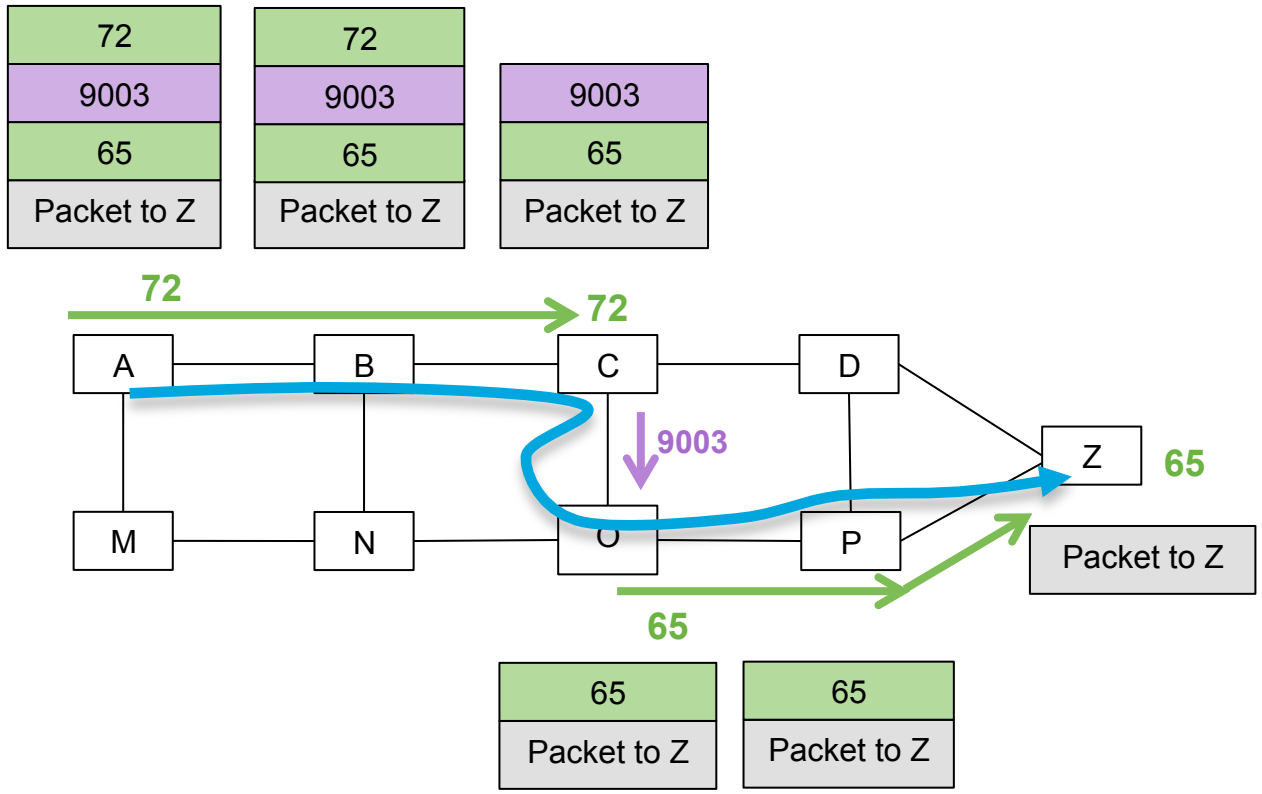
- Node C allocates a local label for CO link segment
- C advertises the adjacency label in IGP
  - e.g. for ISIS, it's a simple sub-TLV extension
- C is the only node to install the adjacency segment in MPLS dataplane (FIB)

Constructing A Path with Adjacency Segments



- Source routing along any explicit path
  - Stack of “adjacency segment” labels
- Segment Routing provides entire path control

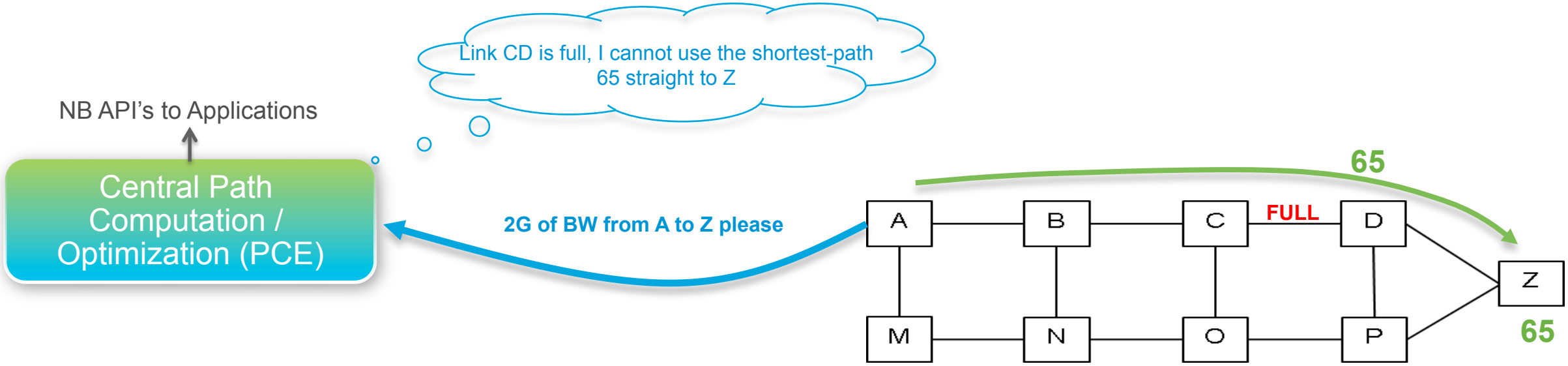
Constructing a Path Combining Node & Adj Segments



- Source Routing along with the explicit path, stack of nodal and adjacency segments
- Any explicit path can be expressed: e.g. ABCOPZ

Use Case: Segment Routing with Central Optimization (PCE)

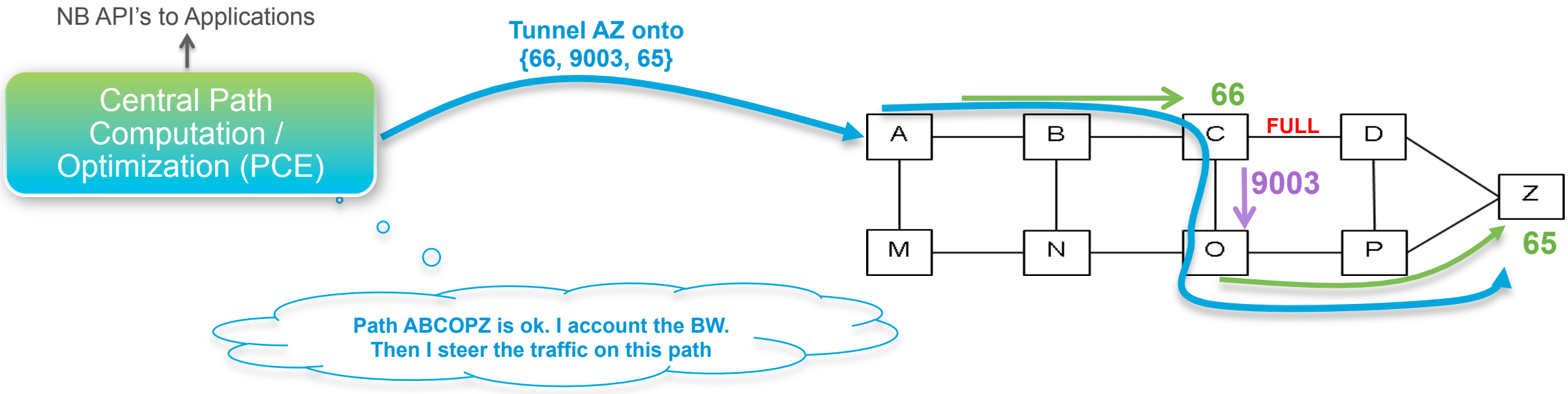
Traffic Engineering with Bandwidth Admission Control



- The network is simple, can respond to rapid changes and is programmable
  - perfect support for centralized optimization efficiency, if required

Use Case: Segment Routing with Central Optimization (PCE)

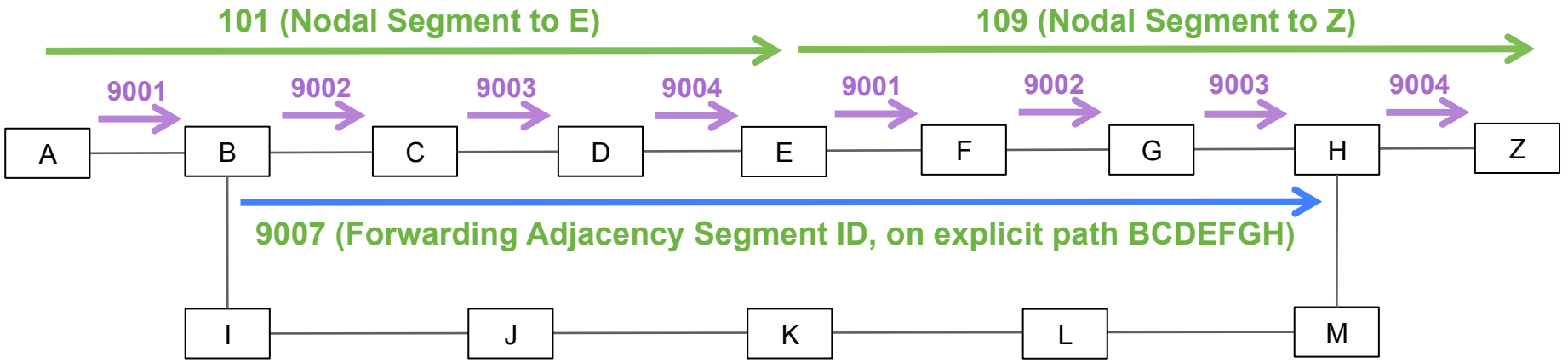
Traffic Engineering with Bandwidth Admission Control



- The network is simple, can respond to rapid changes and is programmable
- The Central Path Computation and Optimization system (PCE) may have Northbound API's through which applications can make requests (such as BW 2G from A to Z with max latency of "X" milliseconds)
- The router nodes in the network needs to have Programmatic interfaces such as PCEP or I2RS to facilitate southbound programming of the network by the PCE system to reflect changes

Use Case: TE Without Bandwidth Admission Control

Deterministic non-ECMP Path



- SR can express deterministic non-ECMP path as a list of adjacency segments

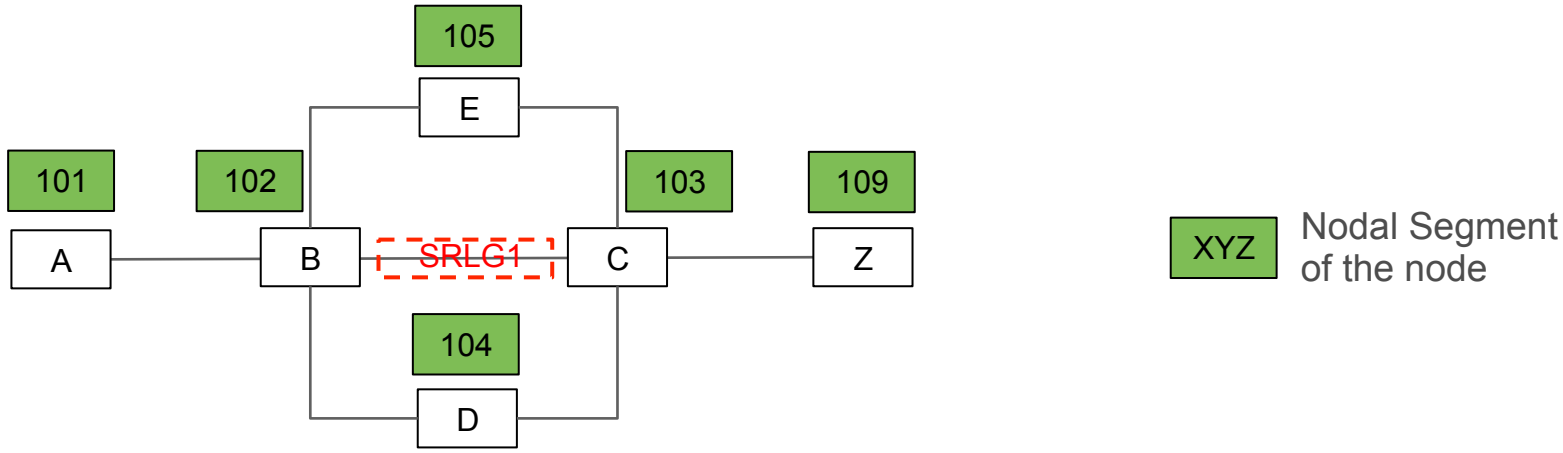
  A specific non-ECMP path i.e. ABCDEFGHZ can be expressed by by a label stack {9001, 9002, 9003, 9004, 9001, 9002, 9003, 9004}

- The label stack can be compressed by following –

  The use of nodal segment of E as 101 and Z as 109, the same path can be expressed as {101, 109}

  Use of Forwarding Adjacency between node B and H with explicit path BCDEFGH and Adjacency Segment ID of 9007, the same path can be expressed as {9001, 9007, 9004}

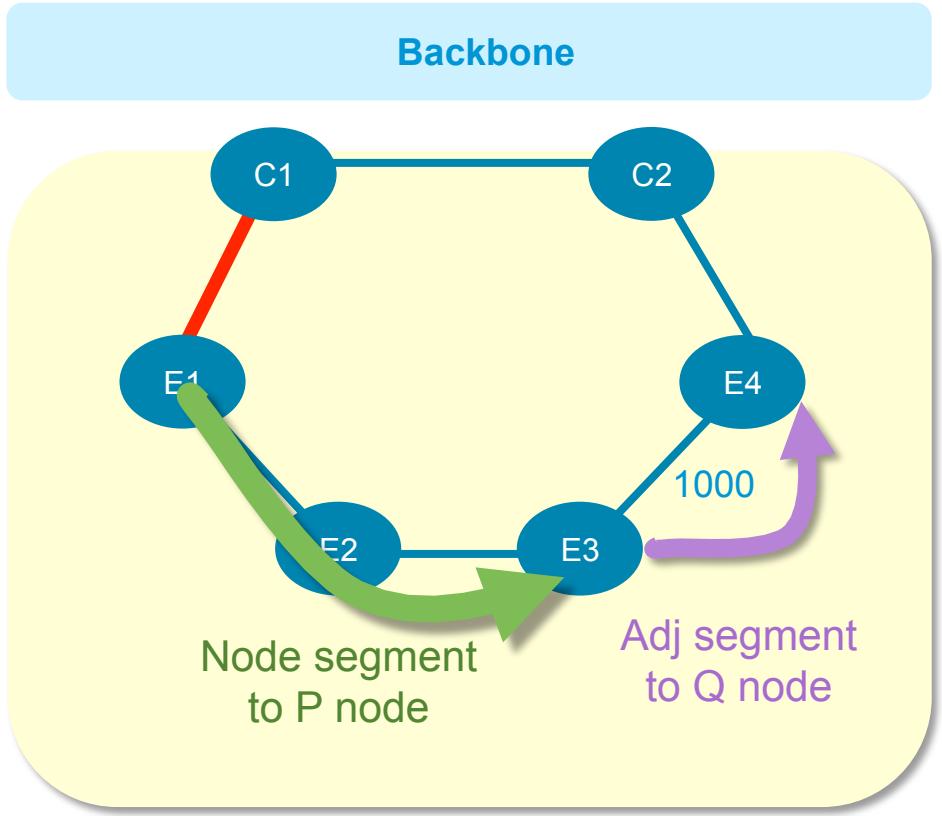
Use Case: TE Without Bandwidth Admission Control
Distributed CSPF Based TE



- A SR head-end router can map the result of its distributed CSPF computation into an SR segment list
- The operator configures a policy on A → Z destined traffic must avoid SRLG1. SRLG1 is link BC
- The SRLG get flooded in the link state IGP. A may implement the policy like the following way –
  - Prunes the links affected by the SRLG1, computes an SPF on the rest topology and picks one SPF paths, say ABDCZ
  - Translates the path as a list of segments – so ABDCZ can be expressed as two nodal segments {104, 109}
  - It monitors the status of the LSDB and upon any change impacting the policy, it either re-computes a path meeting the policy or update its translation as a list of segments

Automated & Guaranteed FRR in Any Topology

- Leverages the IP FRR framework
- IP-based FRR is guaranteed in any topology
  - draft-bryant-ipfrr-tunnels-03.txt
- Directed LFA (DLFA) is guaranteed when metrics are symmetric
- No extra computation (RLFA)
- Simple repair stack
  - node segment to P node
  - adjacency segment from P to Q

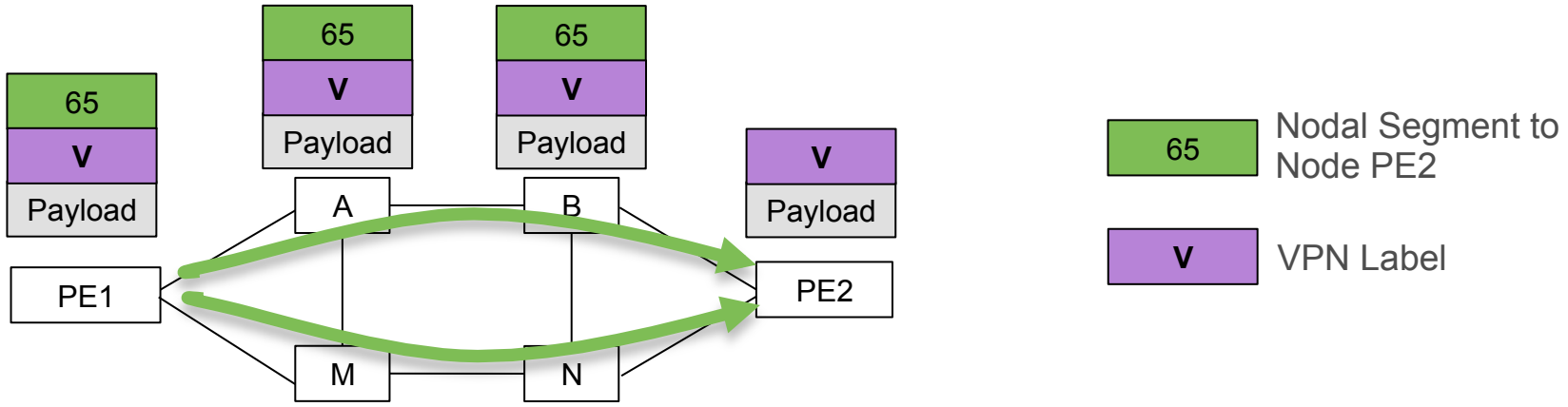


Classic IP/MPLS vs. Segment Routing

Control and Data Plane Comparison in context of IPv4

	Classic IP/MPLS	Segment Routing
Control Plane (Infrastructure)	IGP (IS-IS / OSPF) LDP RSVP-TE PCE + other knobs such as IGP-LDP synchronization, LDPoRSVP etc.	IGP (IS-IS / OSPF) with simple extensions added PCE
Control Plane (Services)	MP-BGP (L3VPN) T-LDP (L2VPN)	MP-BGP (L3VPN) T-LDP (L2VPN)
Data Plane	MPLS Data Plane	MPLS Data Plane

Use Case: Simple & Efficient Transport of MPLS services

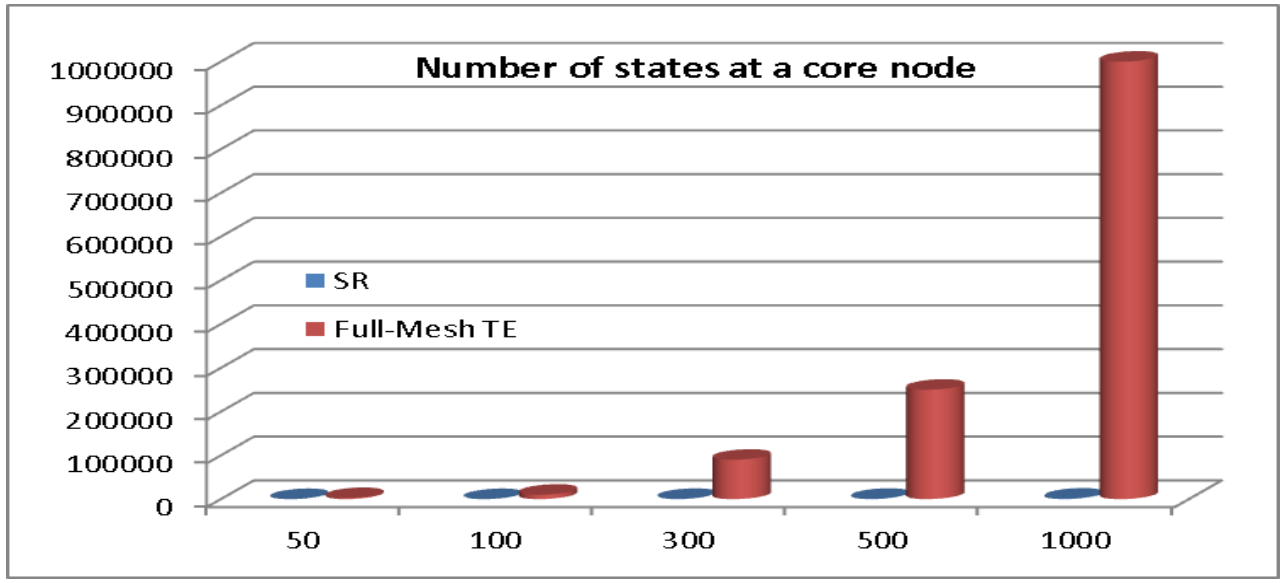


VPN services ride on the node segment to PE2

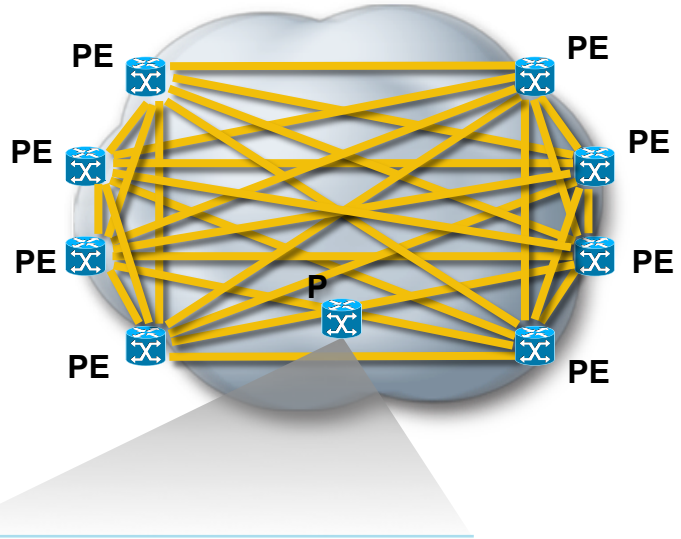
- Transport of MPLS Services – L3VPN, L2VPN
- Efficient packet networks leverage ECMP-aware shortest-path!

  node segment!
- Simplicity - Less protocol(s) to operate, no complex protocol interaction such as LDP – ISIS synchronization to troubleshoot

Use Case: Simple and Scalable Traffic Engineering



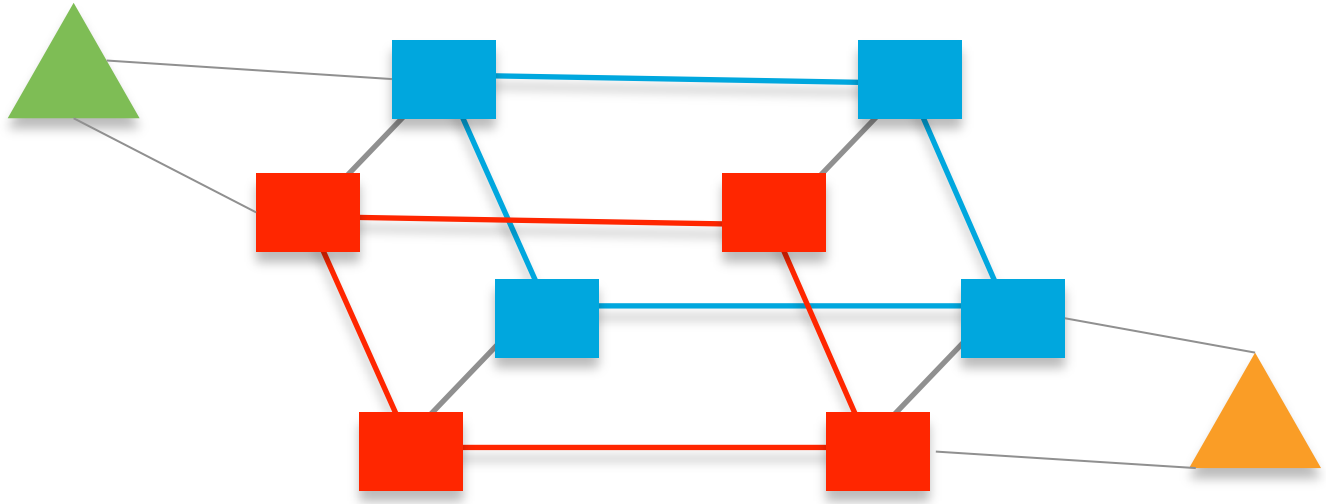
- SR router scales much more than with RSVP-TE
  - The state is not in the router but in the packet
  - Node + Adj vs. Node^2
- No requirement of RSVP-TE protocol
  - And knobs such as LDPoRSVP etc.



	In Label	Out Label	Out Interface
Node Segment Ids	L1	L1	Intf1
	L2	L2	Intf1
	…	…	…
	L8	L8	Intf4
Adjacency Segment Ids	L9	Pop	Intf2
	L10	Pop	Intf2
	…	…	…
	Ln	Pop	Intf5

FIB remains constant

Dual Plane Core



- Each pop has two core routers
  - a blue one and a red one
  - typically in different building/locations
- The blue routers are interconnected and form the blue plane
  - the red routers are interconnected and form the red plane
- The grey links between blue and red routers have bad metric
  - once a packet is within a plane, it reaches its destination without leaving the plane (except if the plane is partitioned)

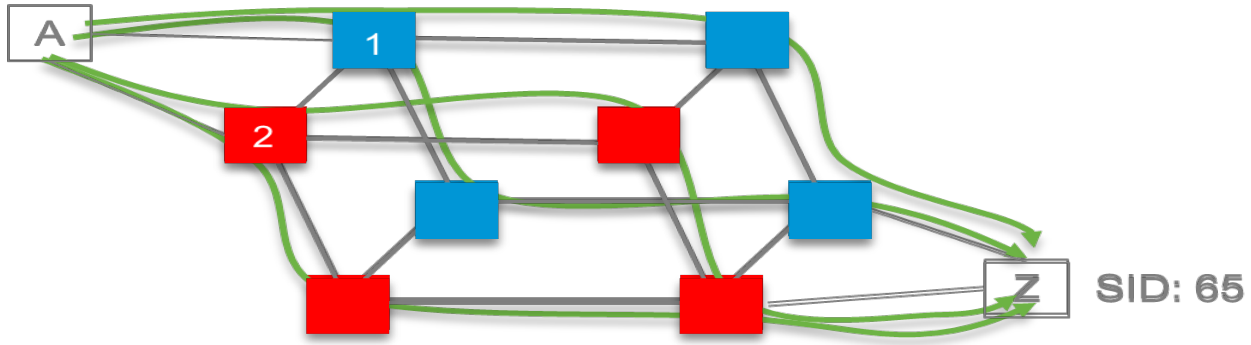
Use Case: Simple Disjointness in Dual Plane Core

TE Without Bandwidth Admission Control – Anycast Node Segment

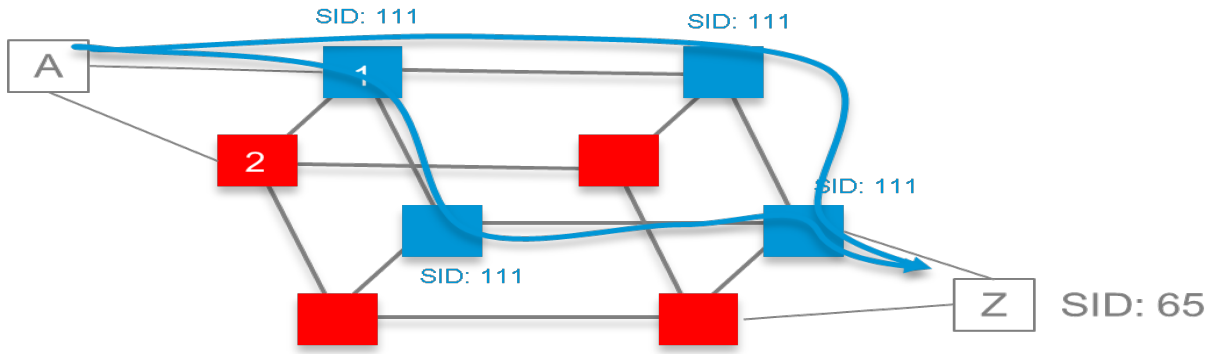
SR avoids state in the core
SR avoids enumerating RSVP-TE tunnels for each ECMP paths

- A sends traffic with [65]

  Classic ECMP “a la IP”



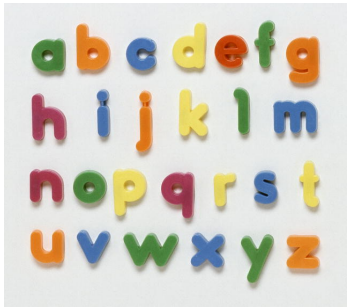
- A sends traffic with [111, 65]
  - All the blue routers advertise the same anycast loopback (1.1.1.1/32) with the same anycast nodal segment 11
  - Packets get attracted in blue plane and then use classic ECMP

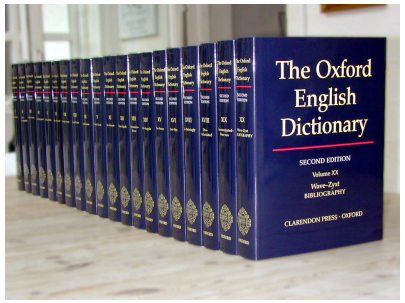


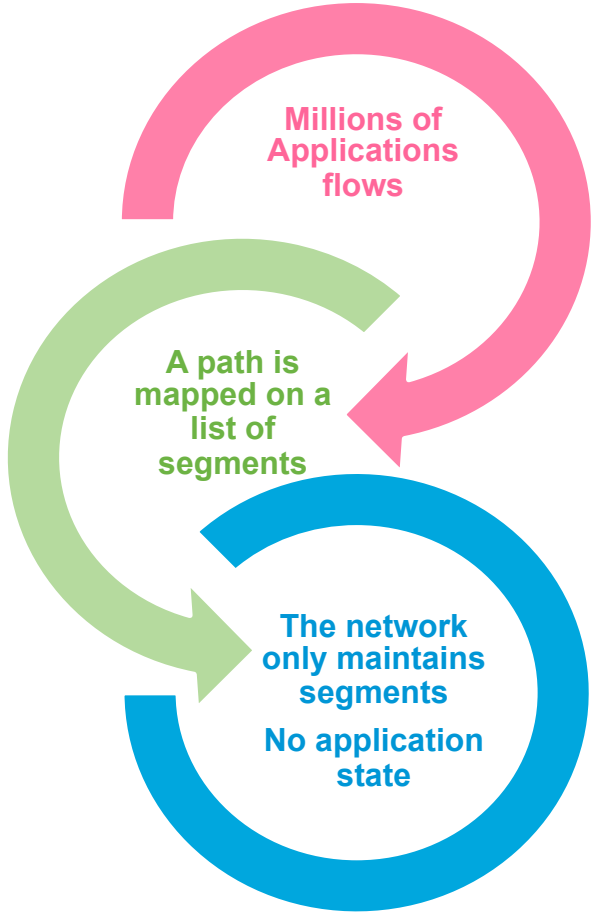
ECMP-awareness!

Scalability and Virtualization

- Each engineered application flow is mapped on a path
  - A lager operator may require millions of such paths
- A path is expressed as an ordered list of segments
- The network maintains segments
  - Typically around thousands of segments
  - Completely independent of application size/frequency & flow scale
- Excellent scaling and virtualization
  - The application state is no longer within the router but within the packet







Ongoing Standardization Effort at IETF

Multiple Vendors and Operators are Collaborating

IS-IS for IP Internets
Internet-Draft
Intended status: Standards Track
Expires: September 21, 2013

S. Previdi, Ed.
C. Filsfils, Ed.
A. Bashandy
Cisco Systems, Inc.
M. Horneffer
Deutsche Telekom
B. Decraene
S. Litkowski
Orange
I. Milojevic
Telekom Srbija
R. Shakir
British Telecom
S. Ytti
TDC Oy
W. Henderickx
Alcatel-Lucent
J. Tantsura
Ericsson
March 20, 2013

Segment Routing with IS-IS Routing Protocol
draft-previdi-filsfils-isis-segment-routing-02

Abstract

Segment Routing (SR) enables any node to select any path (explicit or derived from IGPs SPT computations) for each of its traffic classes. The path does not depend on a hop-by-hop signaling technique (neither LDP nor RSVP). It only depends on a set of "segments" that are advertised by the IS-IS routing protocol. These segments act as topological sub-paths that can be combined together to form the desired path.

Segment Routing – In Summary

- Simple to deploy and operate
  - Leverage MPLS services & hardware
  - straightforward ISIS/OSPF extension, no need of LDP & RSVP-TE
- Provide optimum scalability, resiliency, simplicity and virtualization
- Integration with application through central optimization/PCE system
- IETF standardization effort going on – you are welcome to join & contribute!
- Early EFT Code available for demo
- Stay tuned!

The State is No Longer in the Network, But in the Packet!