Scaling considerations in MPLS networks

Ina Minei

ina@juniper.net

Disclaimer

- **The views presented are of the author and do not necessarily represent Juniper Networks.**
- **This presentation is vendor-independent.**
- **Basic familiarity with MPLS is assumed.**

Topics

- **Scaling – what matters**
- **How much state?**
- **What is the price for configuring/managing the network?**

Scaling

- The magic word
- Tradeoff between the cost of extra state and the benefit brought by the extra state.
- Pitfalls to avoid when doing scaling analysis

The cost of extra state – why care?

- Finite physical resources both in the control plane (e.g. control plane memory, CPU) and in the data plane (e.g. forwarding resources).
- Finite logical resources per/LSR – number of RSVP sessions, number of VRFs supported.
- Network-wide state - affects manageability, load on the protocols.
- Operational/management complexity– configuration, troubleshooting, monitoring.

The cost of extra state – what state?

- Number of LSPs
- Forwarding-plane state
- Control-plane state
- The overhead of maintaining control-plane state
- Operational / management complexity– configuration, troubleshooting, monitoring.

The benefit of extra state

- New services – for example, FRR
- Easier configuration – for example, need not change the default router configuration
- More flexibility - for example, allows finer granularity of the reservations
- More information – for example, for accounting and billing purposes
- Simpler operation – easier to understand solution (for example option A vs. option C in inter-AS VPNs)

Scaling – Tradeoffs

- **The goal is to find the optimum tradeoff between cost and benefit, *for a particular deployment*.**
- **For the same application, the answer may not be the same for two providers.**

Common pitfalls when doing scalability analysis

I. **Not taking a system-wide view**
II. **Comparing incompatible things**
III. **Comparing things in the incorrect context**

Avoiding common pitfalls I - Taking a system-wide view

1. Look at both control-plane and data-plane state.
2. Look at what happens both in steady state and failure/re-optimization scenarios.
3. Look at the entire network, not just at one particular LSR (examine the impact of topology).
4. Look not just at the amount of state created, but also at the cost of maintaining it.
5. Look at all aspects of a solution, not just state created (for example, configuration and troubleshooting complexity must also be taken into account).

Avoiding common pitfalls II - Comparing compatible things

- There is no value in comparing solutions to different problems – for example, L3VPN and L2VPN
- Any comparison should be done while keeping all other factors equal.

Avoiding common pitfalls III - Comparing things in the correct context

- **It is true that a large number of LSPs is not desirable, but not when the goal is the optimization of the link utilization.**

Goals

- Understand how the choices of different signaling protocols, features and configuration options affects the amount of state created, and what are the tradeoffs involved.
- See a few of the common mistakes when doing scalability analysis.
- See a few of the techniques available for improving scaling in MPLS deployments.

Topics

- Scaling – what are the tradeoffs?
- How much state?
- What is the price for configuring/managing the network?

What state we care about

- Number of LSPs
- Forwarding-plane state
- Control-plane state
- The overhead of maintaining control-plane state

Number of LSPs – why care?

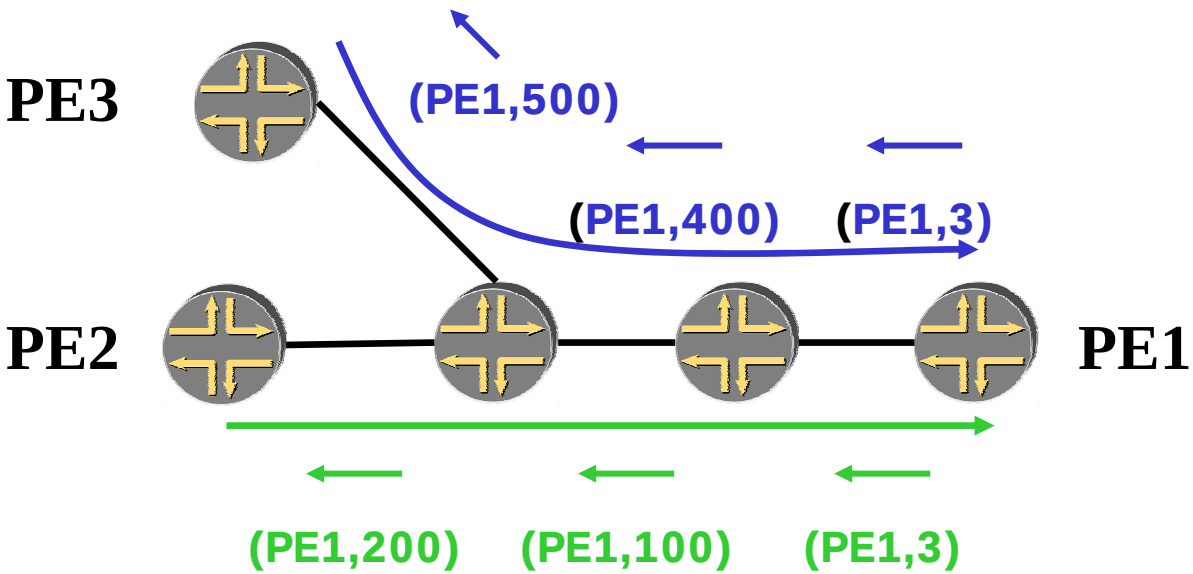
- Protocol state, forwarding state – in the entire network, and per/LSR
- The overhead of maintaining control-plane state – in the entire network, and per/LSR
- Management – will be discussed in the last part of the presentation.

What affects the number of LSPs?

- Choice of signaling protocol (RSVP/LDP)
- Protocol specific issues
  - LDP – independent/ordered control
  - RSVP – reservation granularity, make-before-break, fast-reroute.
- Impact of the topology

Choice of signaling protocol (1) RSVP

- RSVP sets up unidirectional point-to-point (P2P) LSPs.
- A P2P LSP has one head end (ingress) and one tail end (egress).
- LSP creation is initiated by the ingress.

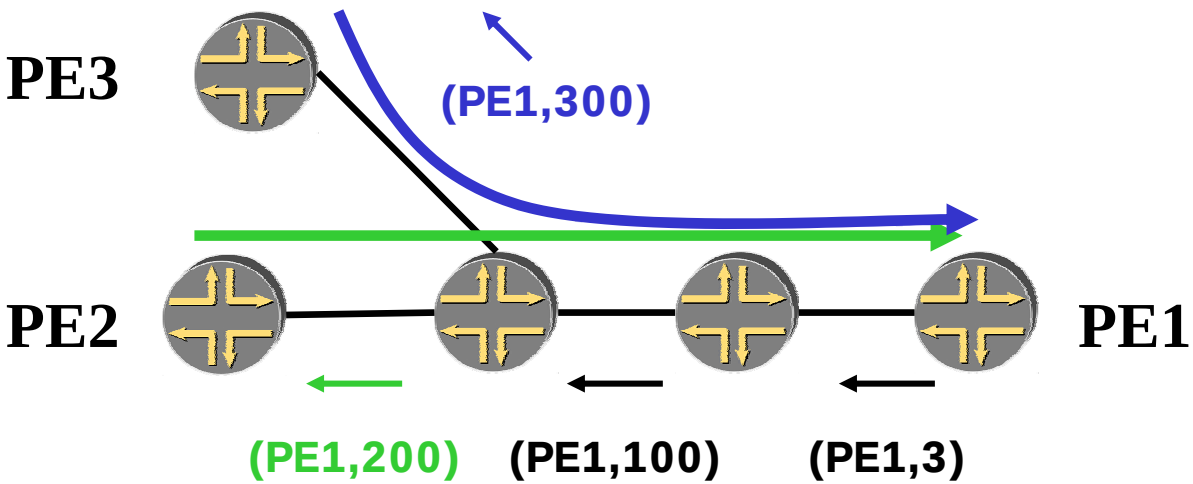


Choice of signaling protocol (2) RSVP

- **Implications:**
  - **The head-end must know who the tail-ends are.**
  - **The number of RSVP sessions at the head end grows proportionally to the number of tail ends.**
  - **Each LSP is considered one session and creates its own forwarding state in the network. (Two LSPs from two different head ends to the same tail end will create twice the amount of state).**

Choice of signaling protocol (3) LDP

- LDP sets up unidirectional multi-point-to-point (MP2P) LSPs.
- An MP2P LSP has one tail end, but multiple head-ends (like an inverted tree).
- LSP setup is initiated by the egress.



Choice of signaling protocol (4) LDP

- **Implications:**
  - **The head-end does not have to maintain explicit knowledge about who the tail-ends are. The tail-ends do not have to know who the head-ends are.**
  - **The number of LDP sessions at the head-end is proportional to the number of directly connected neighbors and has no relation to the number of LSPs in the network.**
  - **LSPs from two different head-ends to the same tail end may share state in the network.**

Choice of signaling protocol (5)

- **Morals:**
  - **The choice of signaling protocol (RSVP/LDP) affects the number of distinct LSPs traversing transit nodes in the network and the amount of forwarding state created.**

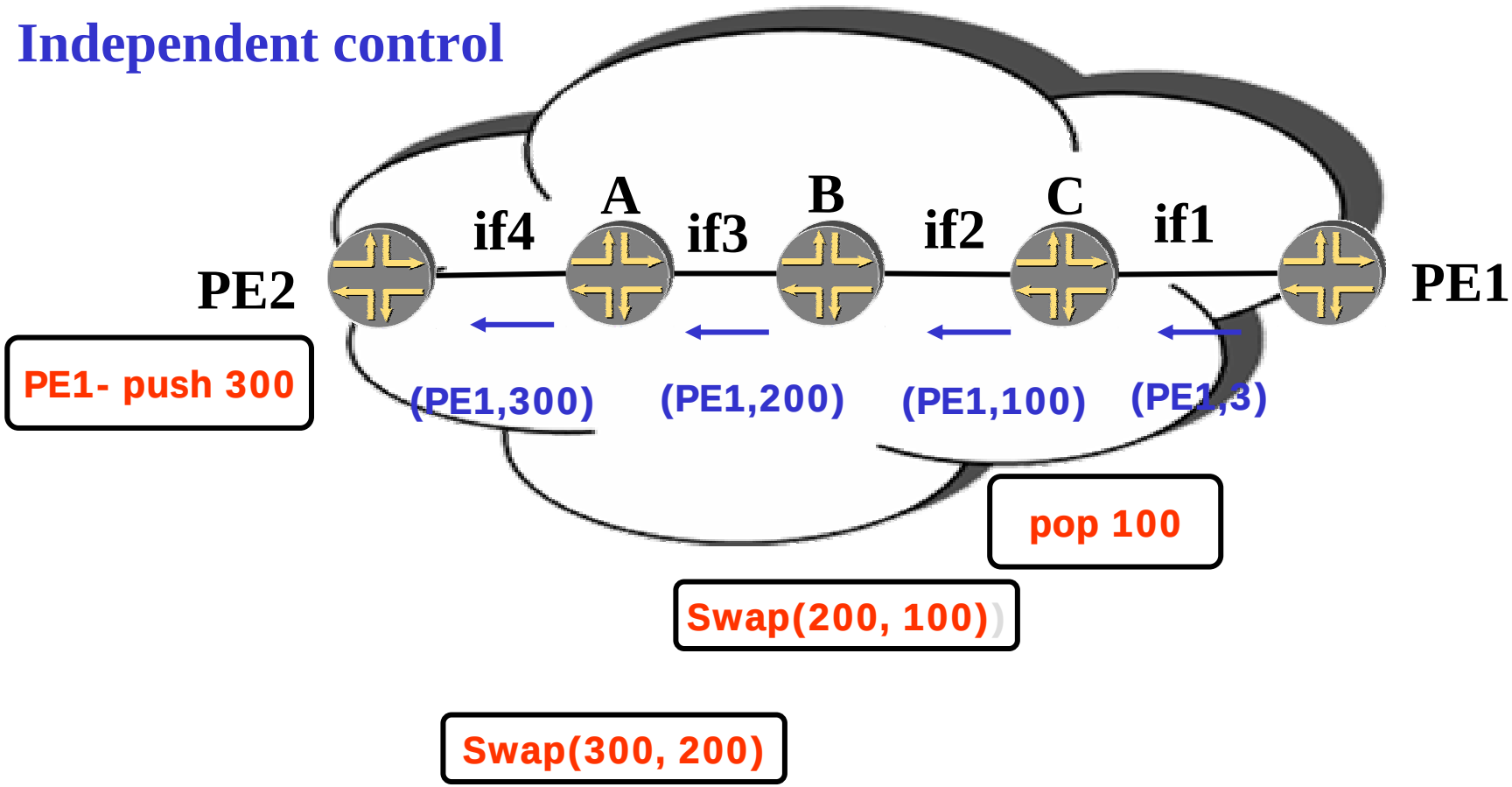
What affects the number of LSPs?

- Choice of signaling protocol (RSVP/LDP)
- Protocol specific issues
  - LDP – independent/ordered control
  - RSVP – reservation granularity, make-before-break.
- Impact of the topology

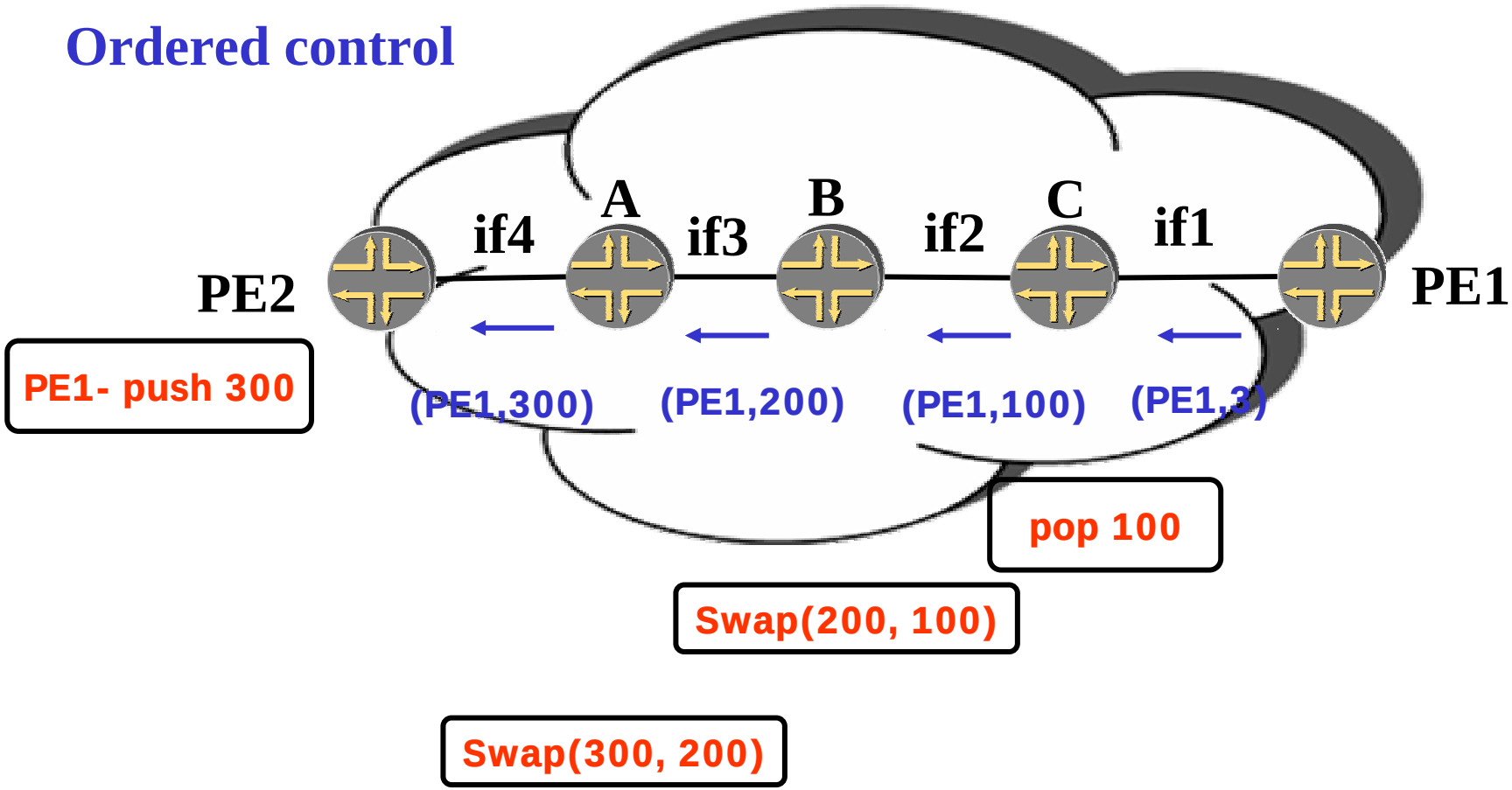
Number of LSPs - Protocol-specific decisions LDP independent/ordered control

- **Independent vs. ordered control**
  - **Independent control – each LSR advertises a label for a FEC independently of any other LSR. Forwarding state is installed for the label received over the IGP path for that FEC.**
  - **Ordered control – the egress LSR advertises a label for the FEC. An LSR that receives a label for a FEC over the best IGP path for the FEC advertises it to its neighbors and installs forwarding state.**

Number of LSPs – LDP independent control



Number of LSPs – LDP ordered control



Number of LSPs - Protocol-specific decisions LDP independent/ordered control

- For an LSP to establish, all routers in the path should advertise a label for a FEC. What are the FECs that should be advertised by default?
  - Ordered control – the loopback address
  - Independent control – all routes in the IGP
- The difference can be between a couple a hundred of FECs carried in LDP (ordered control) and a couple of thousand of them (independent control).
- The FECs advertised can be controlled through configuration -> amount of state can be equal.

Number of LSPs - Protocol-specific decisions LDP independent/ordered control

- **Cost of the default advertisement policy for independent control:**
  - **Maintain and advertise labels for FECs that are not interesting – protocol overhead, memory consumption, more difficult troubleshooting, extra state in the MIBs.**
  - **Create forwarding state for FECs that are not interesting (e.g. interface addresses).**
- **Benefit of the default advertisement policy for independent control:**
  - **Simple configuration**

Number of LSPs - Protocol-specific decisions LDP independent/ordered control

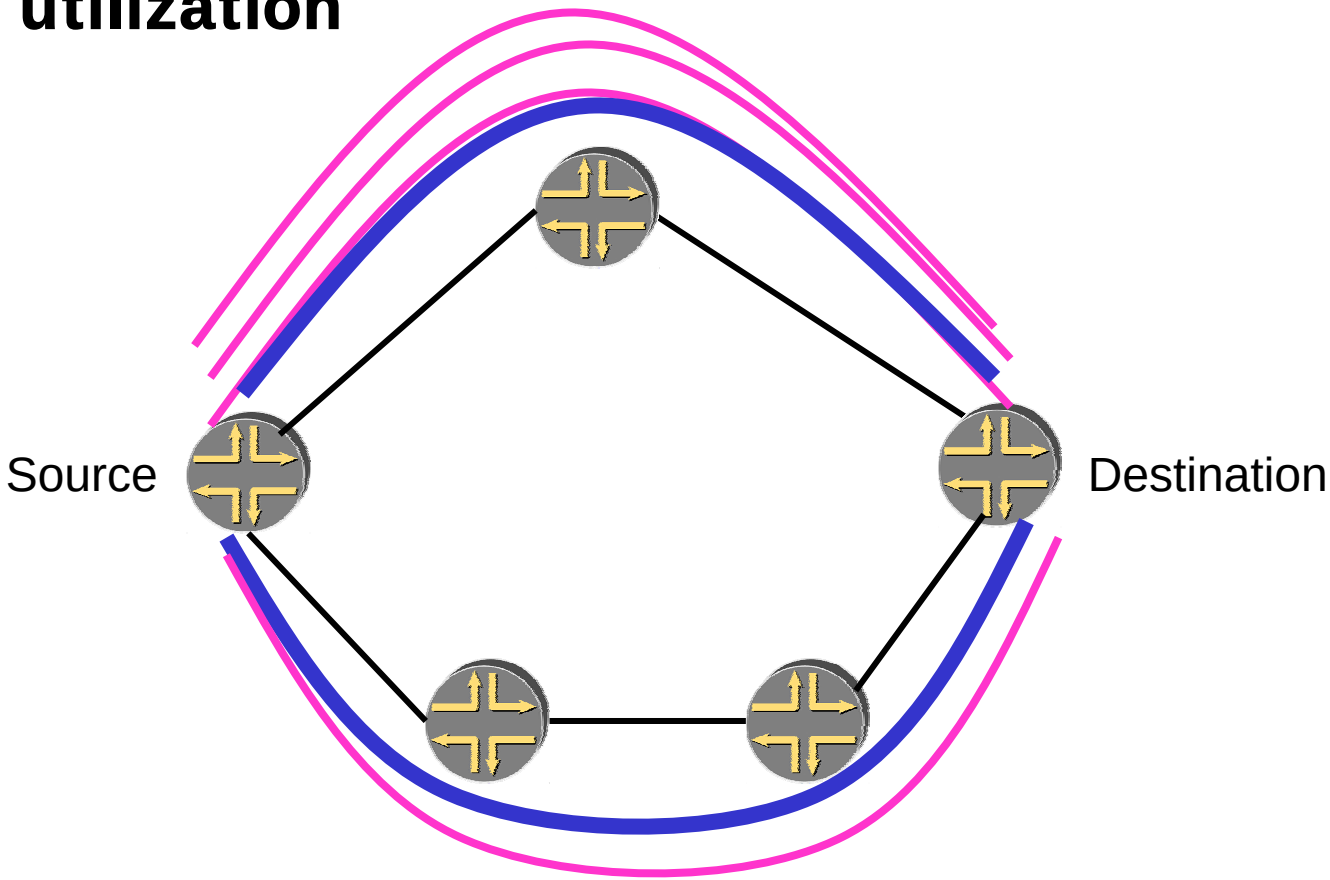
- **Moral:**
  - **Choices within the same protocol can impact the number of LSPs created.**
- **Scaling analysis pitfall: Not taking a system-wide view (ignoring other aspects of the solution)**
  - **Although the amount of state is the same, the operational expense of the configuration required to produce this state is not equal.**

What affects the number of LSPs?

- Choice of signaling protocol (RSVP/LDP)
- Protocol specific issues
  - LDP – independent/ordered control
  - RSVP – reservation granularity, make-before-break.
- Impact of the topology

Number of LSPs - Protocol-specific decisions – RSVP reservation granularity

- **Reservation granularity – how big of an LSP to set up? A single big LSP or several small ones?**
  - **Effect of reservation granularity on the link utilization**



Number of LSPs - Protocol-specific decisions – RSVP reservation granularity

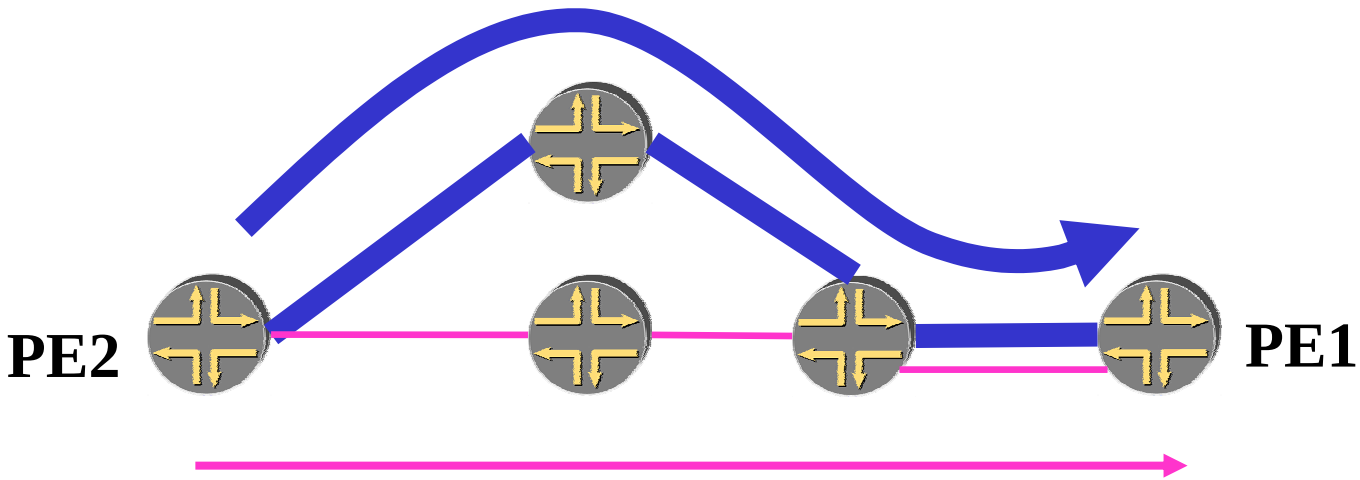
- If the goal is to maximize link utilization, smaller LSPs are better. Similar to how the density of objects in a bin increases as the size of the objects decreases.
- Moral:
  - Smaller reservations create the need for more LSPs.

Number of LSPs - Protocol-specific decisions – RSVP reservation granularity

- **Cost – more LSPs**
- **Benefit - better utilization of the link capacity.**
- **Scaling analysis pitfall: Comparing things in the incorrect context.**
  - **The context is different, because the goal is different. The number of LSPs required at the conceptual level may be different than the number of LSPs created in the network.**

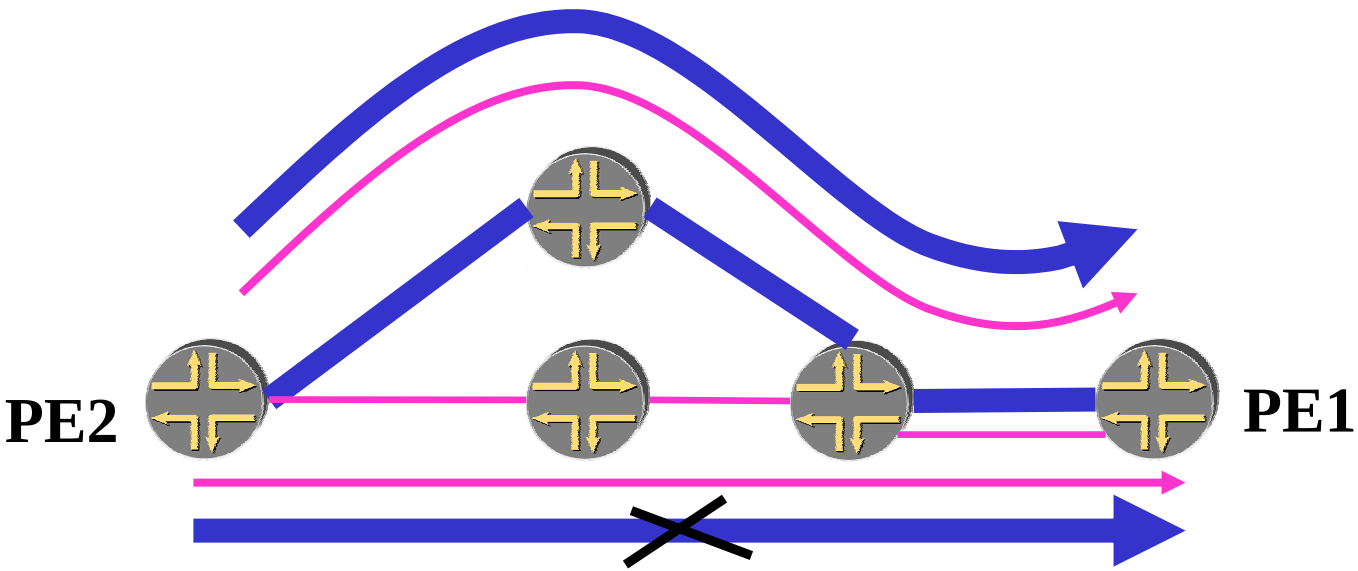
Number of LSPs - Protocol-specific decisions – RSVP Reservation granularity – how to size LSPs

- The minimum link capacity is the gating factor on the size of the LSPs crossing the link.
- In a network with links of different capacities, do alll LSPs have to conform to the smallest-link rule?



Number of LSPs - Protocol-specific decisions – RSVP Reservation granularity – how to size LSPs

- **Benefit – fewer LSPs;**
- **Cost – less available paths in the network for big LSPs. Solution using priority manipulation.**

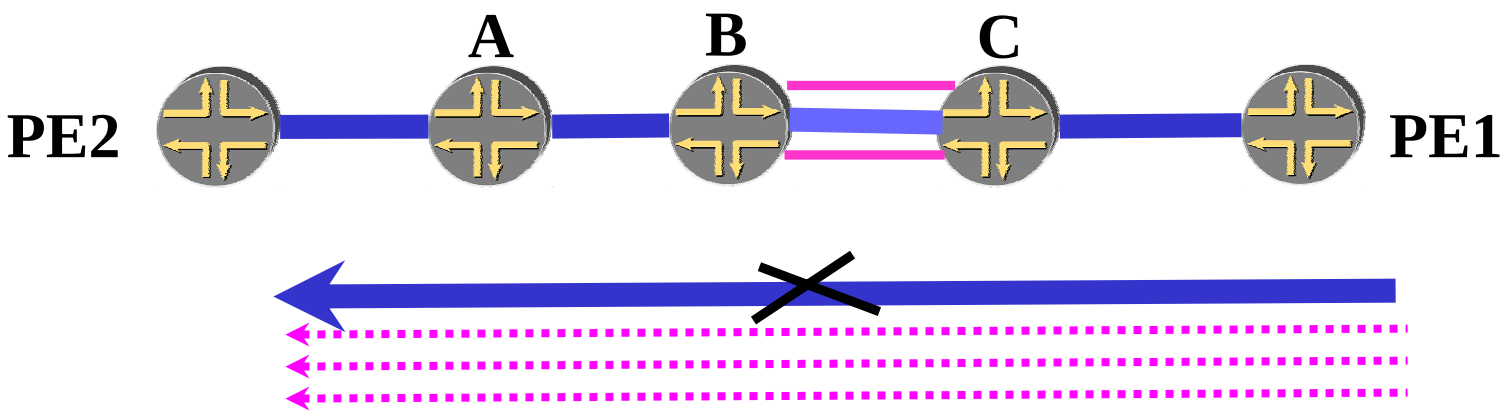


Number of LSPs - Protocol-specific decisions – RSVP Reservation granularity – how to size LSPs

- **Moral:**
  - **The LSP properties (reservation size, priority) is affected by link capacity. This may impose creation of more LSPs than what is required at the conceptual level.**

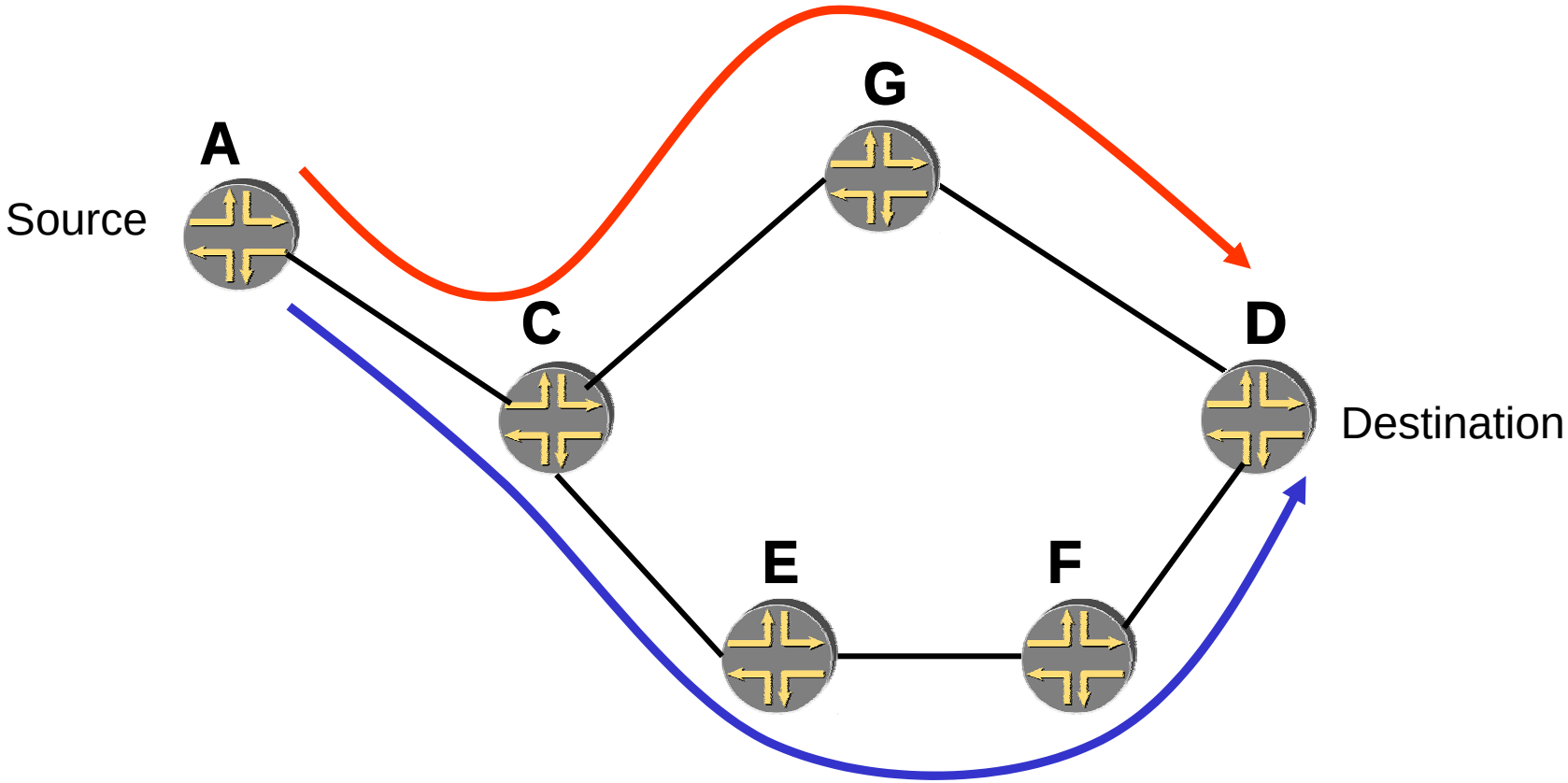
Number of LSPs - Protocol-specific decisions – RSVP - Reducing the number of LSPs

- The link capacity is the gating factor on any LSP traversing the link.
- Link-aggregation as a solution. What are the disadvantages?



Number of LSPs - Protocol-specific decisions – RSVP make-before-break

- **Make-before-break – set up a new path before tearing down the old path**



Number of LSPs - Protocol-specific decisions – RSVP make-before-break

- Cost – temporarily increases the number of LSPs in the network (extra forwarding state, extra protocol state)
- Benefit - useful for implementing auto-bandwidth, re-optimization. Used after fast reroute has kicked in, to move the LSP away from the protection path.

Number of LSPs - Protocol-specific decisions – RSVP make-before-break

Morals:

- **The number of LSPs and the forwarding state may grow temporarily (this must be taken into account when computing the average number of LSPs crossing an LSR).**
- **Must understand the behavior of features deployed (FRR, auto-bandwidth)**

Number of LSPs - Protocol-specific decisions – RSVP make-before-break

- **Scaling analysis pitfall: Not taking a system-wide view (ignoring what happens following re-optimization, failure)**
  - **The number of LSPs and the forwarding state may grow temporarily (this must be taken into account when computing the average number of LSPs crossing an LSR).**
  - **Must understand how particular features impact the number of LSPs created.**

What affects the number of LSPs?

- Choice of signaling protocol (RSVP/LDP)
- Protocol specific issues
  - LDP – independent/ordered control
  - RSVP – reservation ganularity, make-before-break.
- Impact of the topology

Number of LSP - Impact of topology (1)

- What is interesting:
  - total number of LSPs network-wide
  - number of LSPs traversing any particular node
- Common pitfall – computing the average number of LSPs traversing a node.
- Instead, must compute the number of LSPs that will traverse the most loaded node in the network.

Number of LSP - Impact of topology (2)

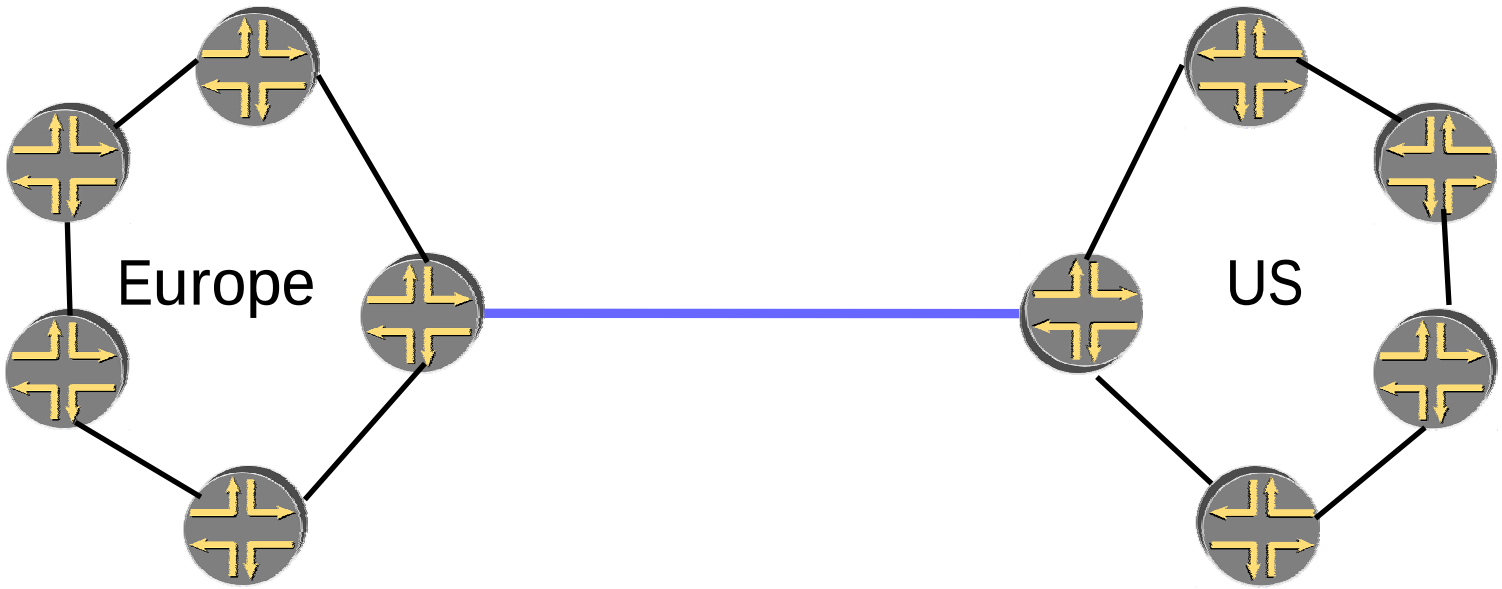
- 10 PoPs, each with a single WAN router (for simplicity).
- Full mesh topology
- Total number of LSPs = 10 * 9 = 90
- Maximum number of LSPs that can traverse a node = 90 – (LSPs for which the LSR is head-end or tail-end for) = 90-9-9 = 72
- Average number of LSPs per node = 72/10 = 8 LSPs

Number of LSP - Impact of topology (3)

- 5 PoPs in Europe, 5 PoPs in the US.
- The trans-continental links are from the DC PoP to the London PoP. (assume a single link for simplicity)
- The DC and London routers represent “choke points” in the network, because all inter-continental LSPs must pass through them.

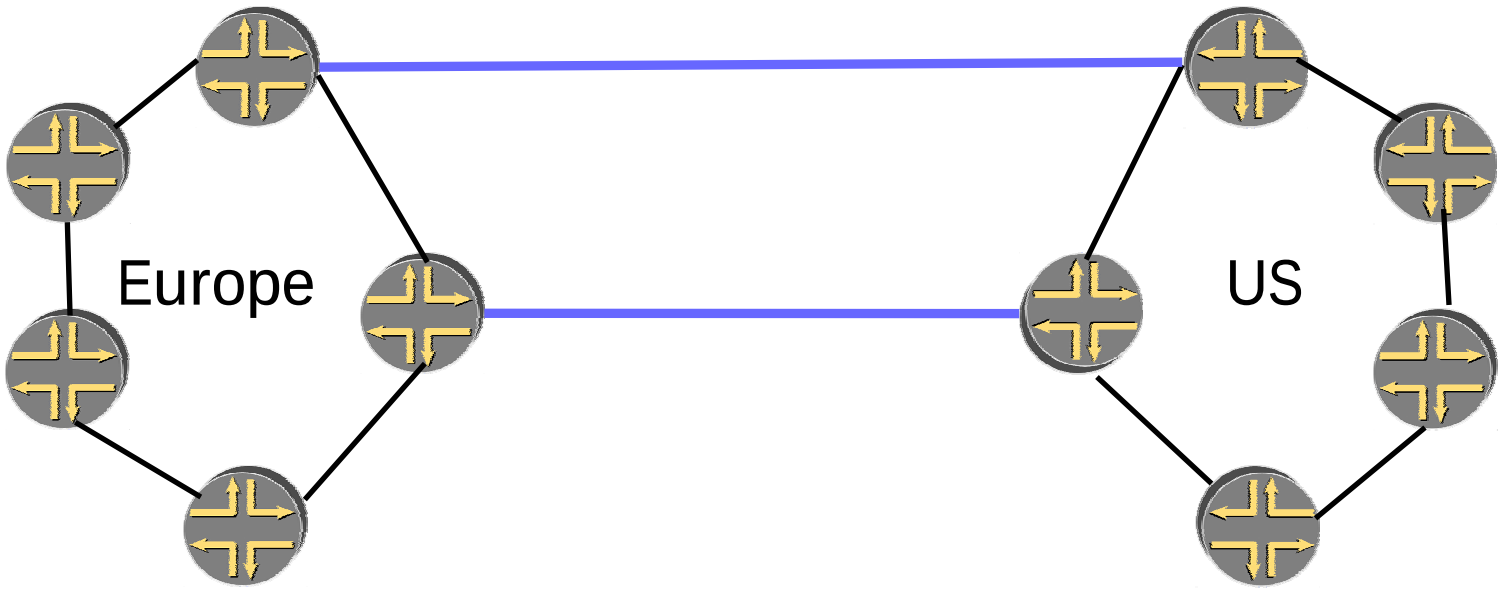
Number of LSP - Impact of topology (4)

- Number of LSPs transiting the choke points = 4 * 5 = 20 (in each direction). Total = 40 LSPs.
- Is this the final number?



Number of LSP - Impact of topology (5)

- What happens with two inter-continental links? 3 * 5 = 15 in each direction. 30 LSP total, can assume 15 on each router.
- What if one of the links fails?



Number of LSP - Impact of topology (6)

- **Scaling analysis pitfall:** **Not taking a system-wide view (ignoring the impact of topology).**
- **Morals:**
  - **Topology affects the paths of the LSPs, and the amount of state maintained by each LSR.**
  - **It is required to take a global view of the network when evaluating a particular deployment.**
  - **Must take into account the state that is created both in the steady state and following a failure or a re-optimization.**

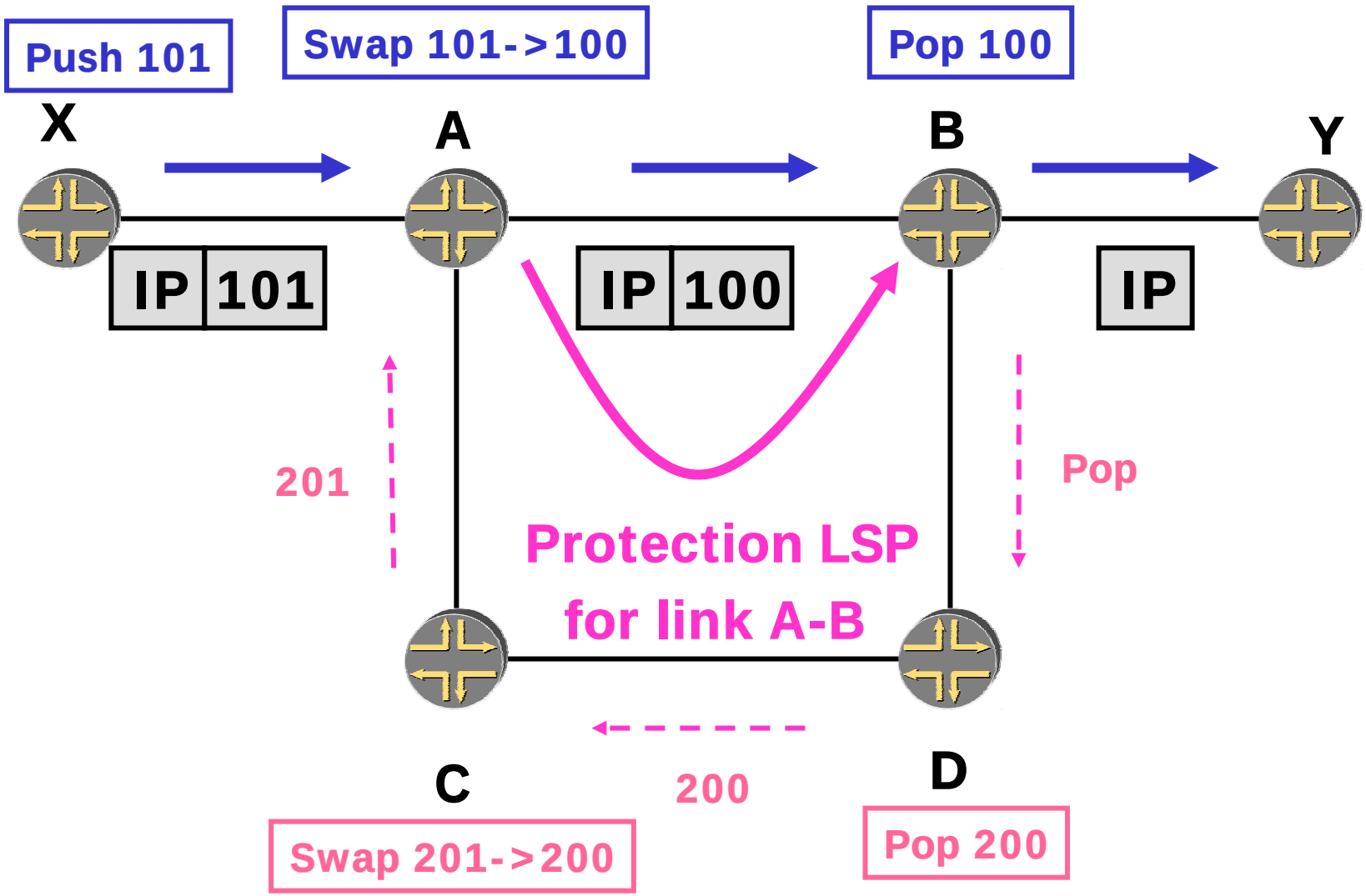
What state we care about

- Number of LSPs
- Forwarding-plane state
- Control-plane state
- The overhead of maintaining control-plane state

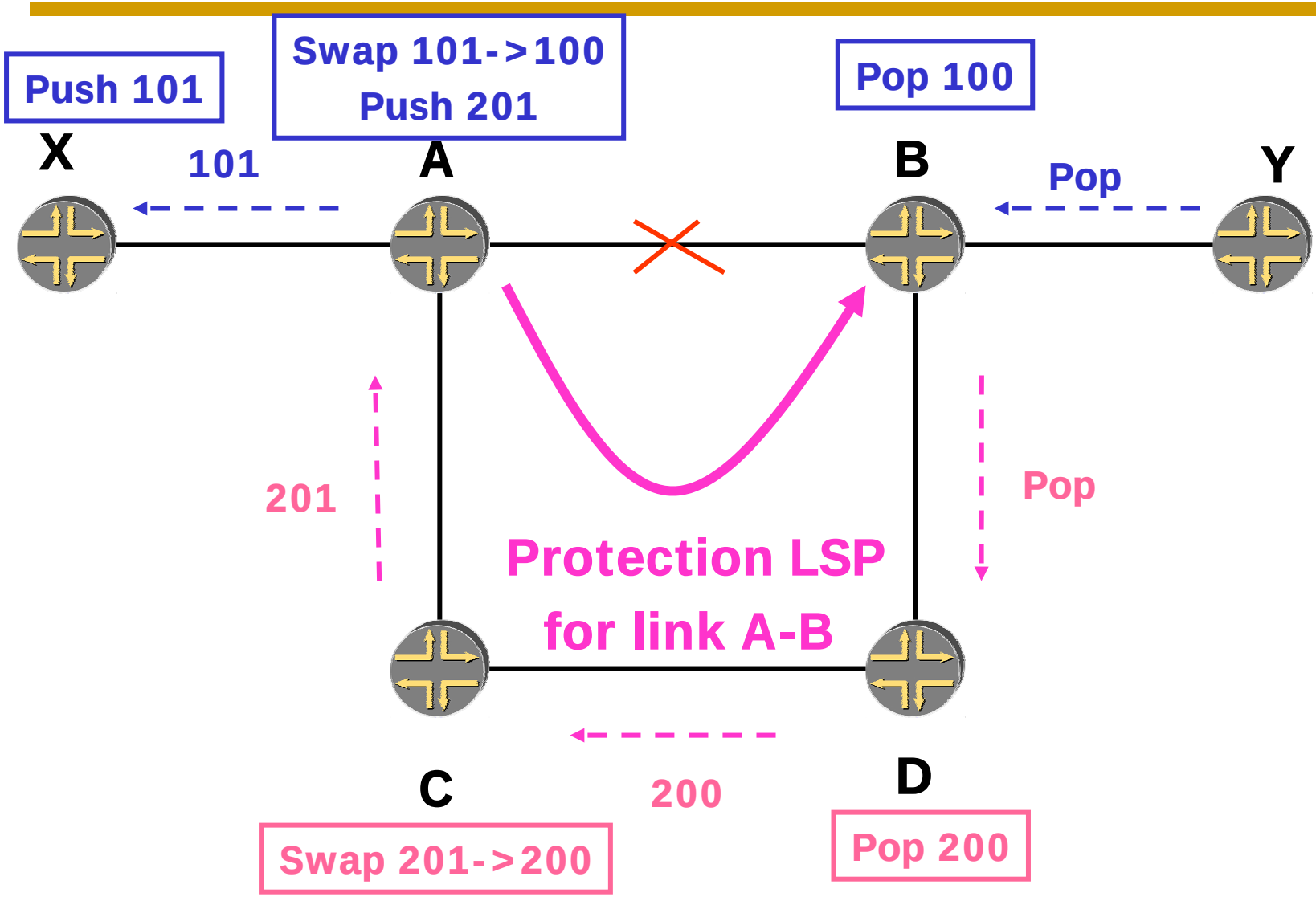
Forwarding-plane state - FRR

- **Local protection using fast reroute**
  - **Construct a “protection” LSP around a point of failure (the backup path). Nest the LSPs that traverse the point of failure onto the protection LSP.**
  - **The protected resource can be either a link or a node.**
  - **The backup path can be dedicated to a particular LSP (1:1 backup) or can be shared between several LSPs (facility backup)**

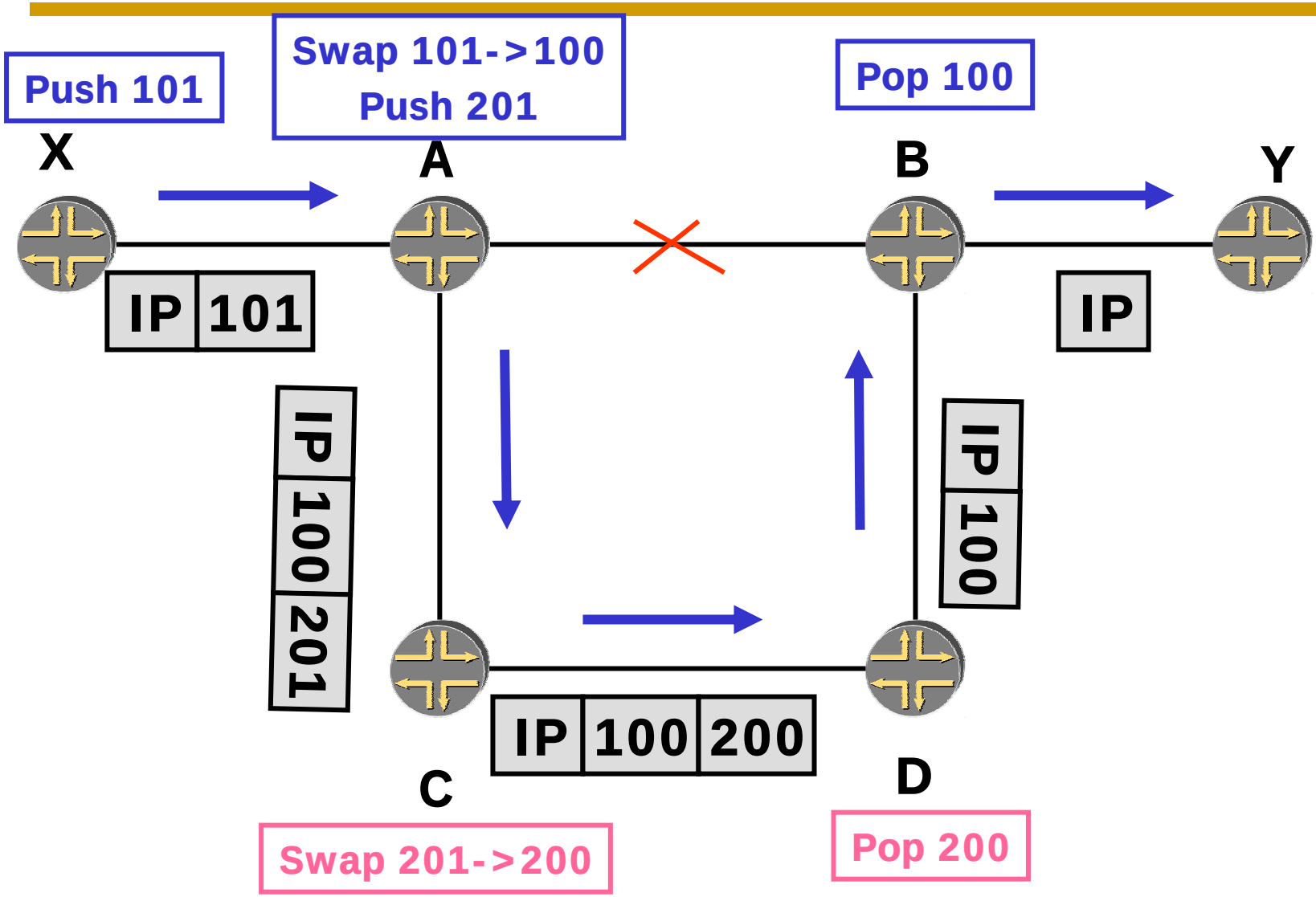
Fast reroute with MPLS (cont)



Fast reroute with MPLS (cont)



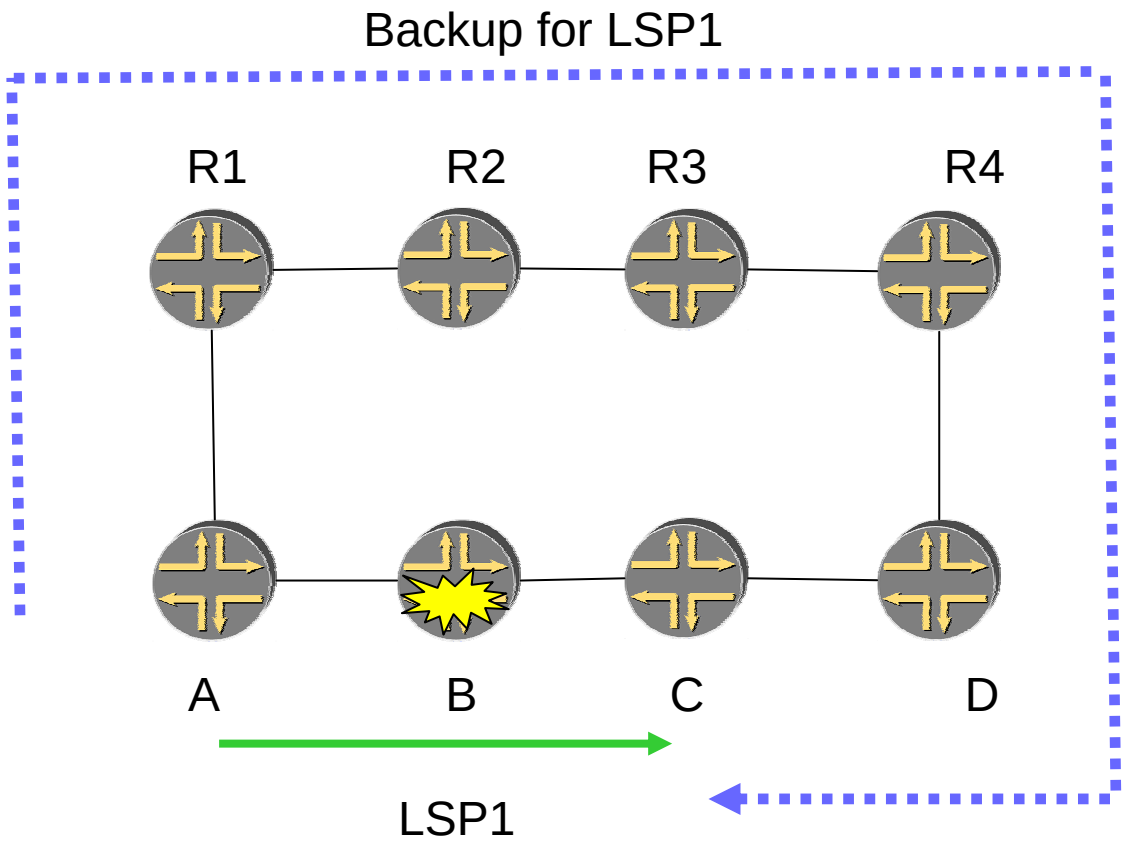
Fast reroute with MPLS (cont)



Forwarding-plane state – Local protection using fast-reroute

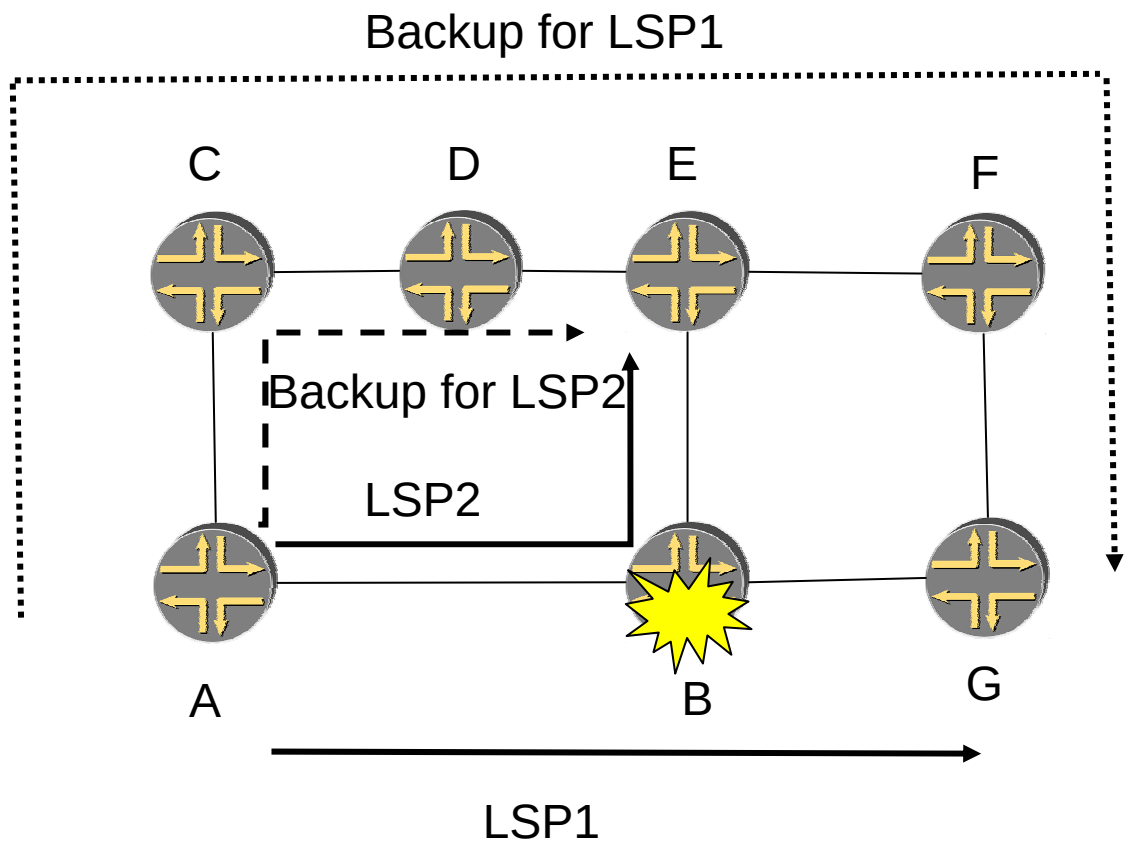
- The backup path creates extra state and consumes extra forwarding resources.
- Local protection using fast reroute. How much state is created for the backup?
  - Topology dependent
  - Protection type dependent:
    - one-to-one vs. facility
    - link vs. node
- Scaling analysis pitfall: Not taking a system-wide view.

Forwarding plane state – impact of topology



Forwarding plane state – impact of protection type

- **One-to-one vs. facility, node vs. link.**



What state we care about

- Number of LSPs
- Forwarding-plane state
- Control-plane state
- The overhead of maintaining control-plane state

Control plane state (1)

- **P2P LSPs vs MP2P LSPs (RSVP vs. LDP) comparison for full mesh connectivity:**
  - **Protocol state (session/adjacencies) that must be maintained – proportional to the number of interfaces for LDP, proportional to the number of LSPs crossing the router for RSVP (this is larger than the number of PEs).**
  - **Adding a new PE to the mesh will typically not increase the LDP protocol state, but will do so for RSVP.**

Control plane state (2)

- **Does this mean that LDP is more scalable than RSVP? No, because the two protocols do not provide the same functionality.**

- **Scaling analysis pitfall: Comparing incompatible things. Comparisons must be done keeping all other factors equal. If all that is required is the setup of an LSP, LDP will create less state.**

What state we care about

- Number of LSPs
- Forwarding-plane state
- Control-plane state
- The overhead of maintaining control-plane state

Overhead of control plane state maintenance

- **RSVP is soft-state, requires periodic refresh.**
- **RSVP refresh-reduction – classic example of how to reduce the overhead of control-plane state maintenance.**

Topics

- Scaling – what are the tradeoffs?
- How many LSPs?
- What is the price for configuring/managing the network?

Management aspects

- Configuration
- Troubleshooting
- Statistics collection
- Monitoring - liveness detection

Configuration – properties

- Configuration is always necessary
- Less configuration is better because:
  - Configuration is expensive to manage
  - Less configuration means less room for errors
- The amount of configuration depends on:
  - Particular design / implementation decisions in the software (e.g. vrf-import and vrf-export need not be configured separately if they are the same)
  - Protocol / application properties

Configuration – minimizing the amount of configuration (RSVP auto-mesh)

- **P2P LSPs with RSVP**
  - **The head end must know about all the tail ends.**
  - **LSPs must also be set up from the tail ends towards the head end.**
- **Typically achieved via configuration. What happens when a new PE is added to the network?**
- **The idea: use the IGPs to signal membership in an RSVP LSP mesh, instead of configuring the LSP endpoints.**

Configuration – minimizing the amount of configuration (RSVP auto-mesh)

- **Is configuration totally eliminated? No, because the group membership still needs to be configured.**

- **Moral: Ease of configuration is an important scaling property. For cases where the configuration complexity stems from the protocol itself, a solution can be found at the protocol level.**

Troubleshooting

- **The more state, the more difficult to troubleshoot.**
  - **More data to walk through**
  - **More state in the MIBs, more state for the show commands.**
- **Example: missing label in the LDP database.**

Statistics collection

- Used for billing and accounting, network planning, etc.
- The price of statistics maintenance: gathering, exporting and processing statistics incurs an overhead.
- Must maintain state only for items that are interesting for the particular the network design. For example: Are statistics necessary with the granularity of a VPN route? Which FECs need to be carried in LDP?

Monitoring

- **Liveness detection**
  - **Tradeoff between the polling frequency and number of LSPs monitored.**
  - **How many incoming packets to expect?**

- **Moral: must only poll for those LSPs that provide a useful function in the network.**

Summary (1)

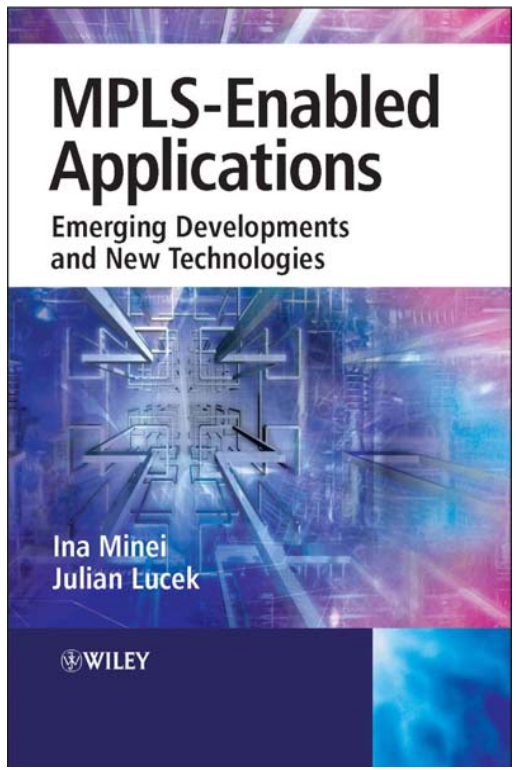
- The presentation only focused on the transport LSPs, the building block for other applications.
- The amount of state is not always obvious when taking a high-level view.
- The amount of state limits the equipment that can be used, and imposes operations/management restrictions.
- There is a tradeoff between the cost of the extra state and the benefit it brings. There is no such thing as a good solution, only a good solution for a particular deployment.
- There are ways to minimize the state.

Summary (2)

- State is not the only thing that must be taken into account when analyzing the properties of a solution.
- Must also look at:
  - The overhead of maintaining the state
  - Configuration complexity
  - Ease of troubleshooting

More info

MPLS-Enabled applications
http://www.juniper.net/training/jnbooks/



Thank you!

Please send comments to ina@juniper.net