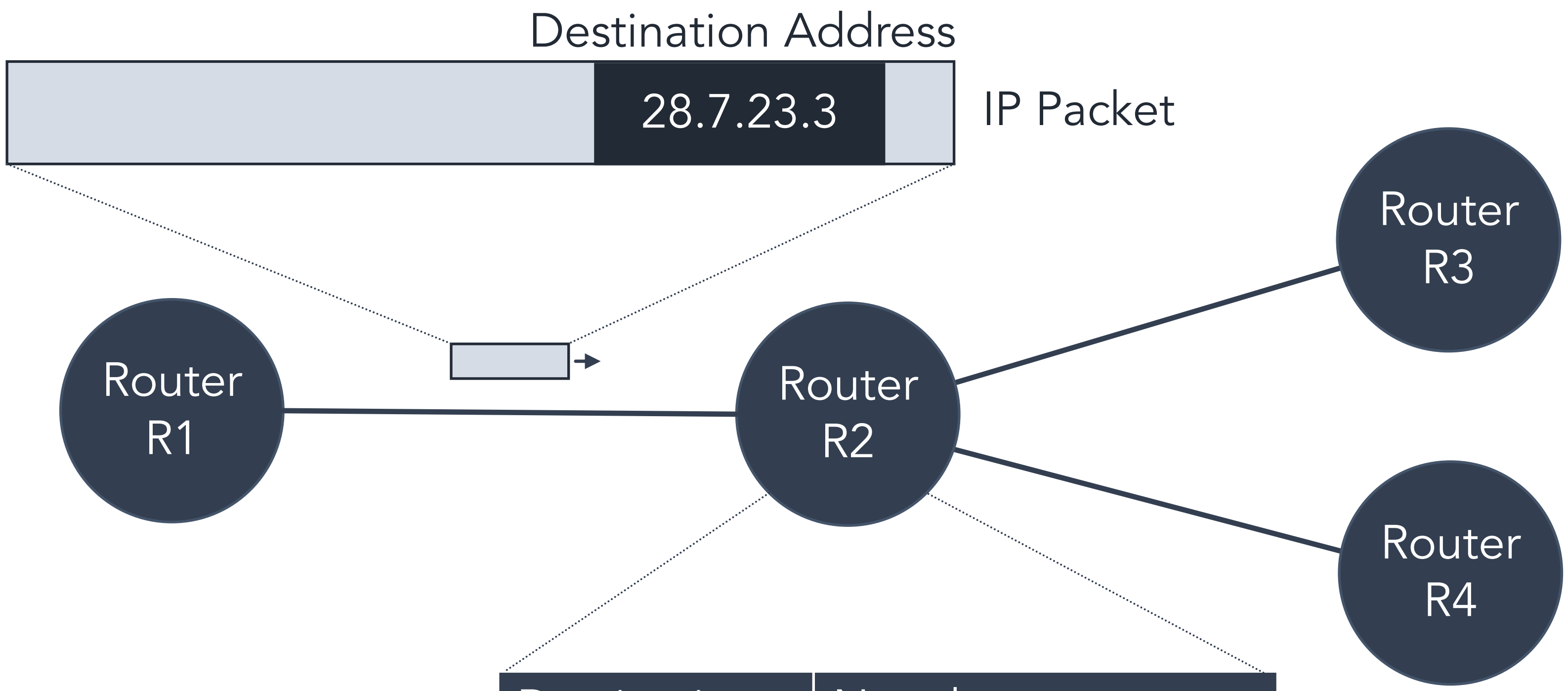
Traffic Engineering (TE)

QuTech, Delft

© Bruno Rijsman, brunorijsman@gmail.com, 11July 2019 v1

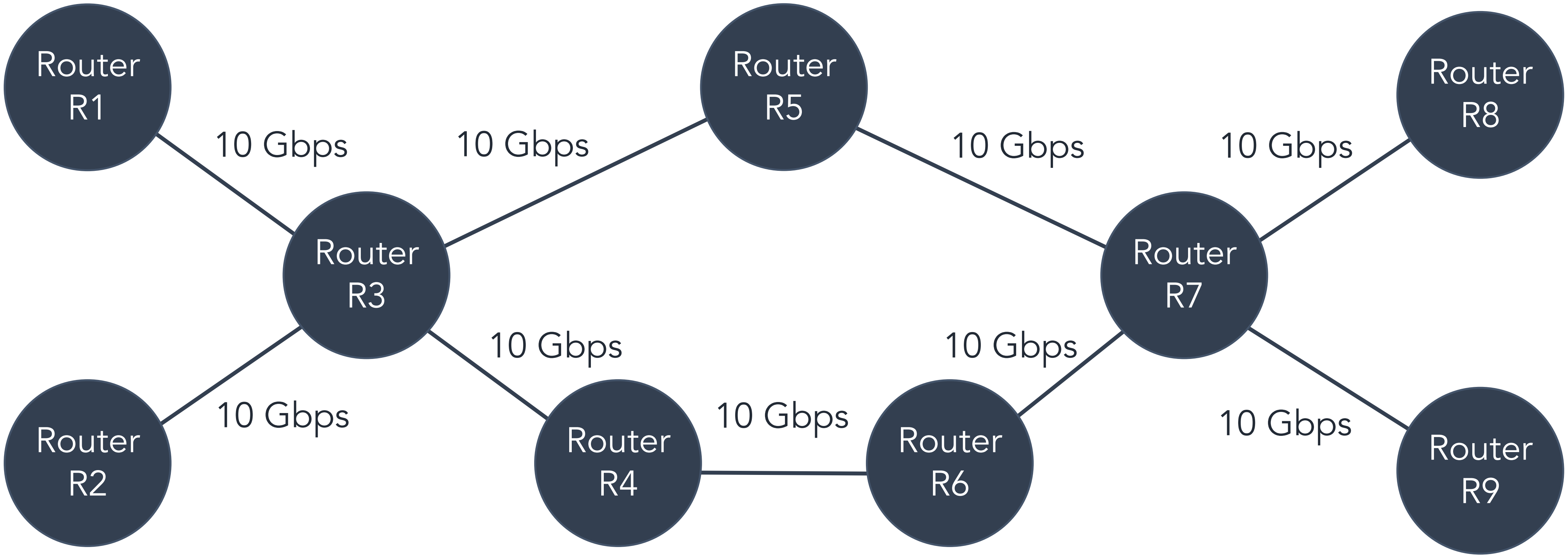
Stateless hop-by-hop IP forwarding



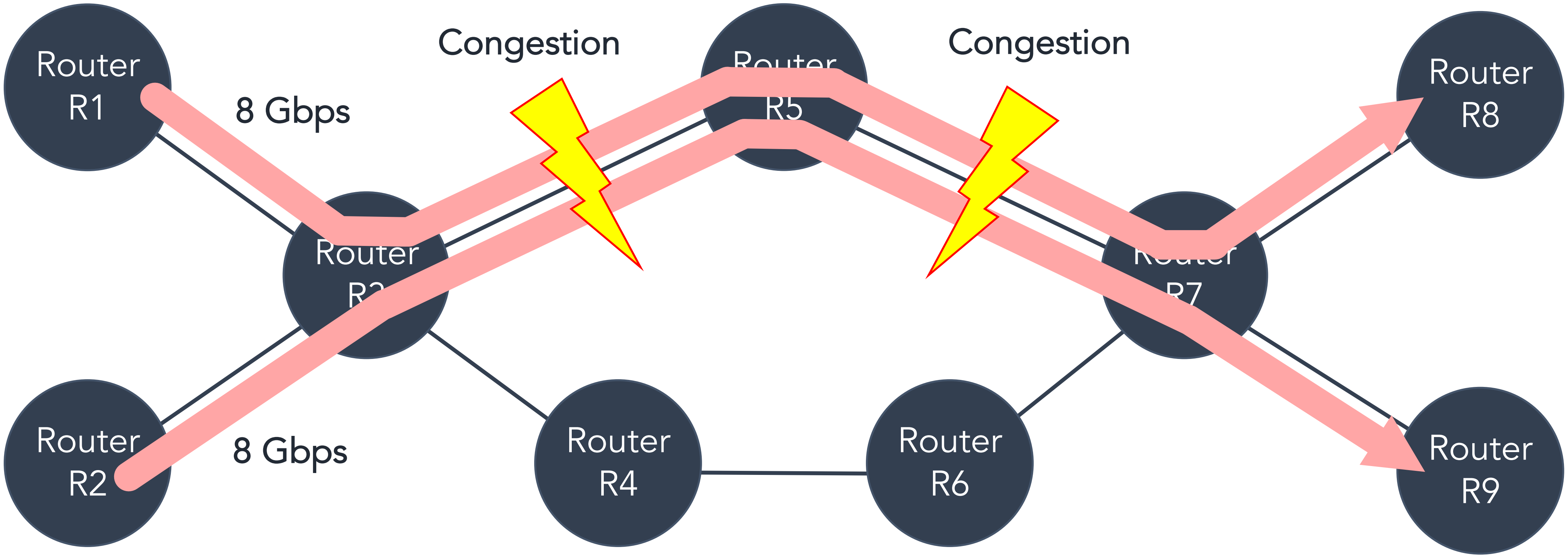
Destination	Next-hop
…	…
28.7.23.0/24	Router R3
…	…

Forwarding Table

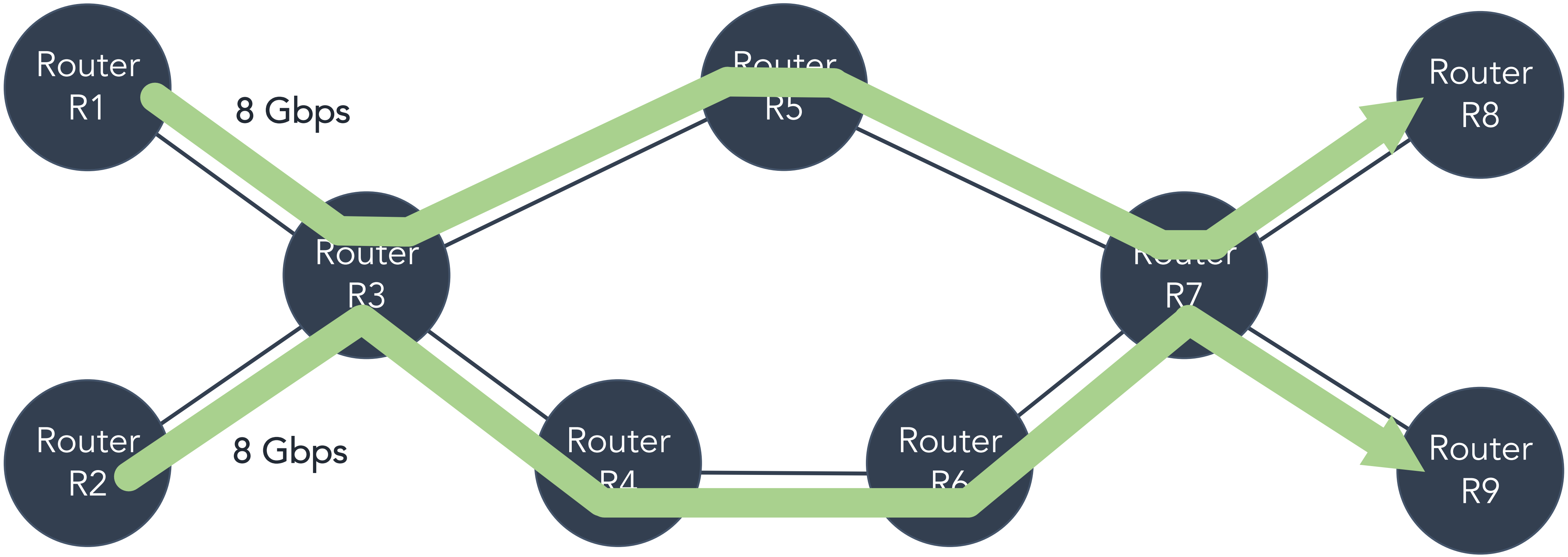
Example network



Congestion due to shortest-path forwarding



Traffic engineering to avoid congestion



Non-shortest path

When is TE used in classical networks?

TE only matters if there is (potential) congestion

- Traffic demand is high relative to available bandwidth
- Traffic demand does not automatically adapt (e.g. TCP back-off)
- Application is sensitive to latency, jitter, or drops (e.g. video or voice)
- Over-provisioning the network is not an option
- Cost of recovery in application layer is high (e.g. Wide Area Network)

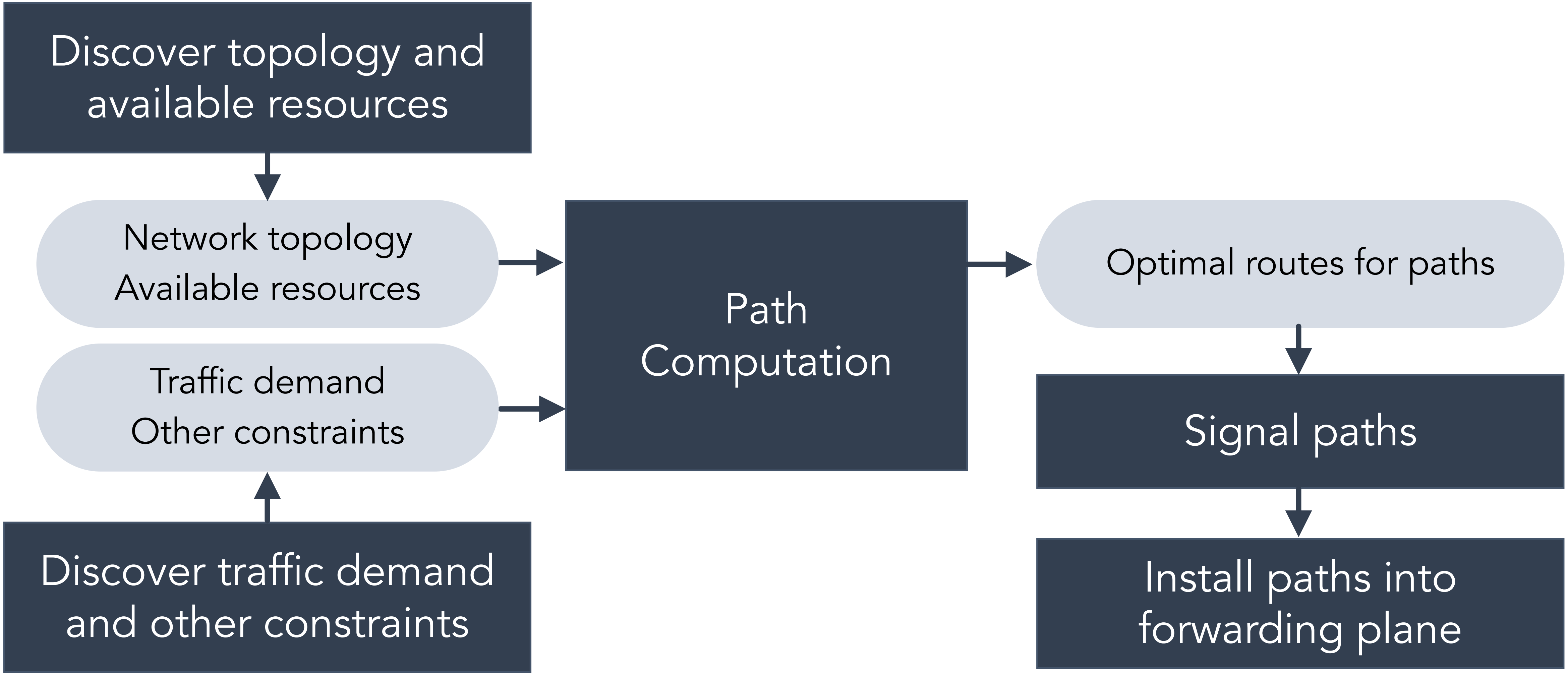
TE only works if traffic demand is known / predictable

- Traffic demand can be measured
- Traffic demand follows predictable pattern (e.g. diurnal pattern)
- Or, traffic demand can be scheduled (e.g. Google internal backbone)

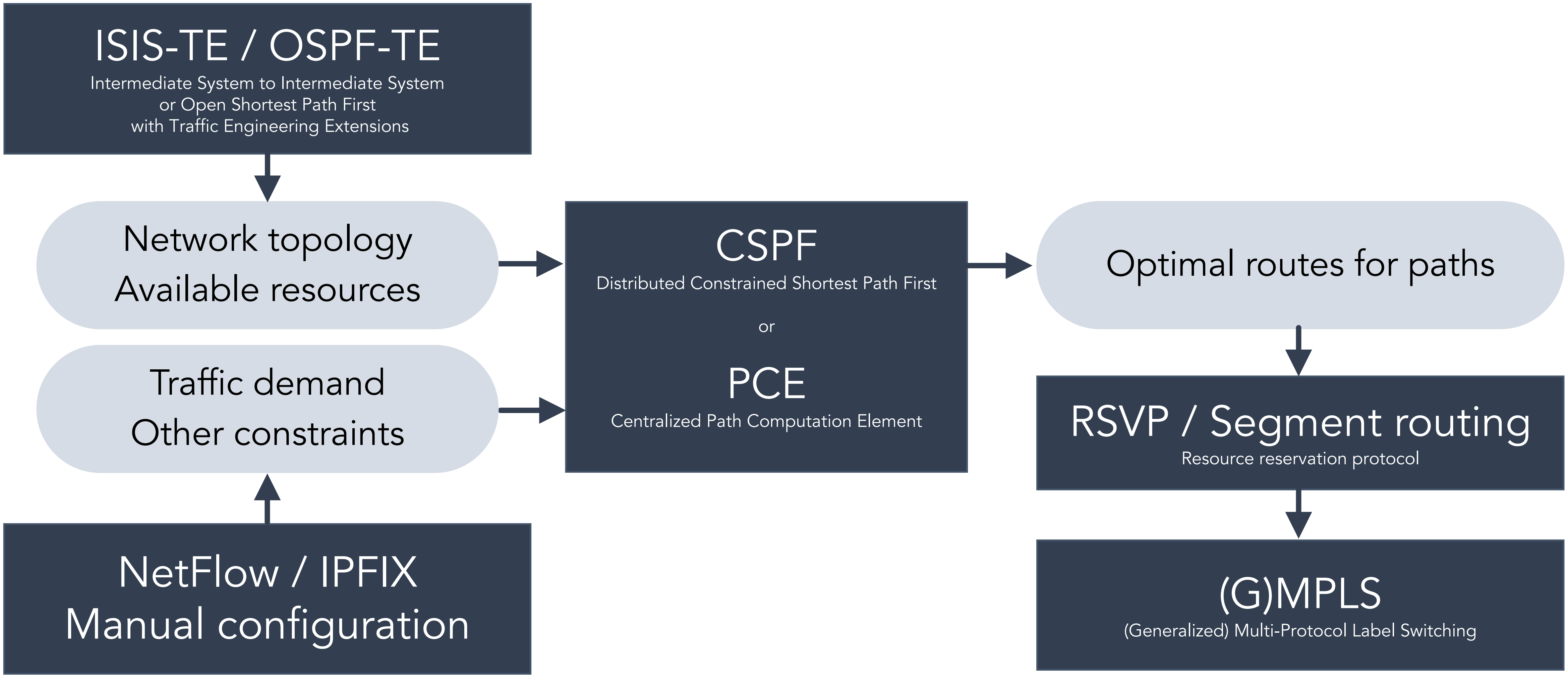
Lessons for quantum networks

- Bandwidth on quantum networks is extremely scarce
  - 10s to 100s of Bell pairs per second on point-to-point links
  - End-to-end Bell pairs may be even less (probabilistic swap, distillation, …)
- "Congestion" is not acceptable
  - Stored qubits have short decoherence times
  - Queueing delays will cause decoherence and loss of fidelity
- <u>Traffic engineering makes sense for quantum networks</u>

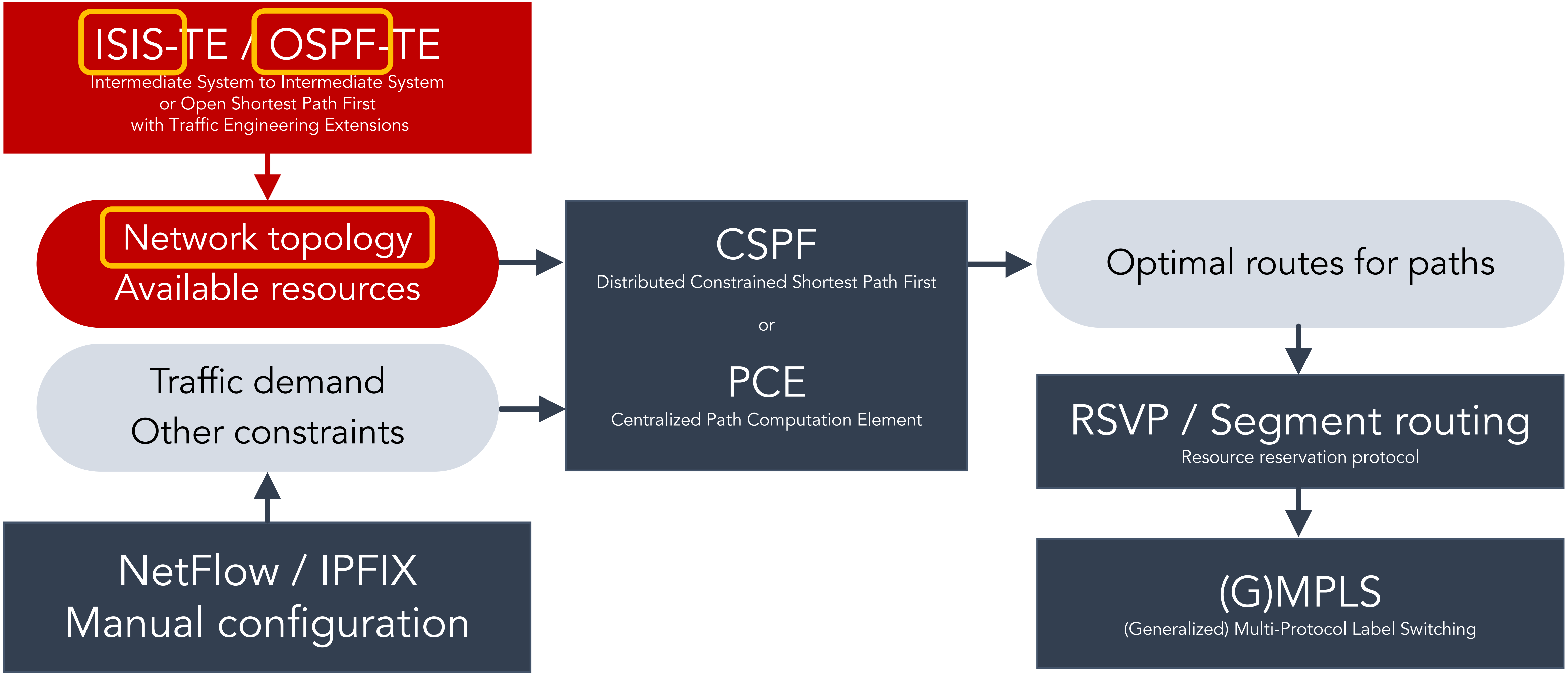
Traffic engineering functional blocks



Classical traffic engineering protocols



Topology discovery



Discover topology using link-state routing

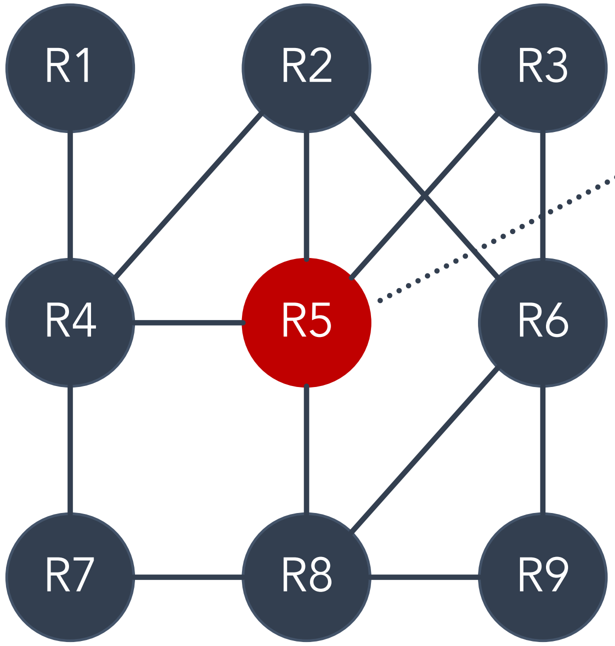
There are two widely used link-state routing protocols:

1. Open Shortest Path First (OSPF)
2. Intermediate System to Intermediate System (ISIS)

They are very similar. Each router:

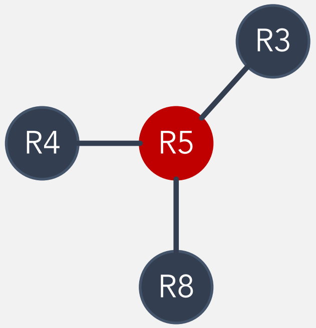
1. Discovers direct neighbors ("adjacencies") using hello packets
2. Summarizes all adjacencies in a link-state packet
3. Reliably floods its own link-state packet to all other routers
4. Stores the link-state packets from all other routers in a database
5. Constructs the topology of the network from the link-state database

Link state packet reports adjacencies

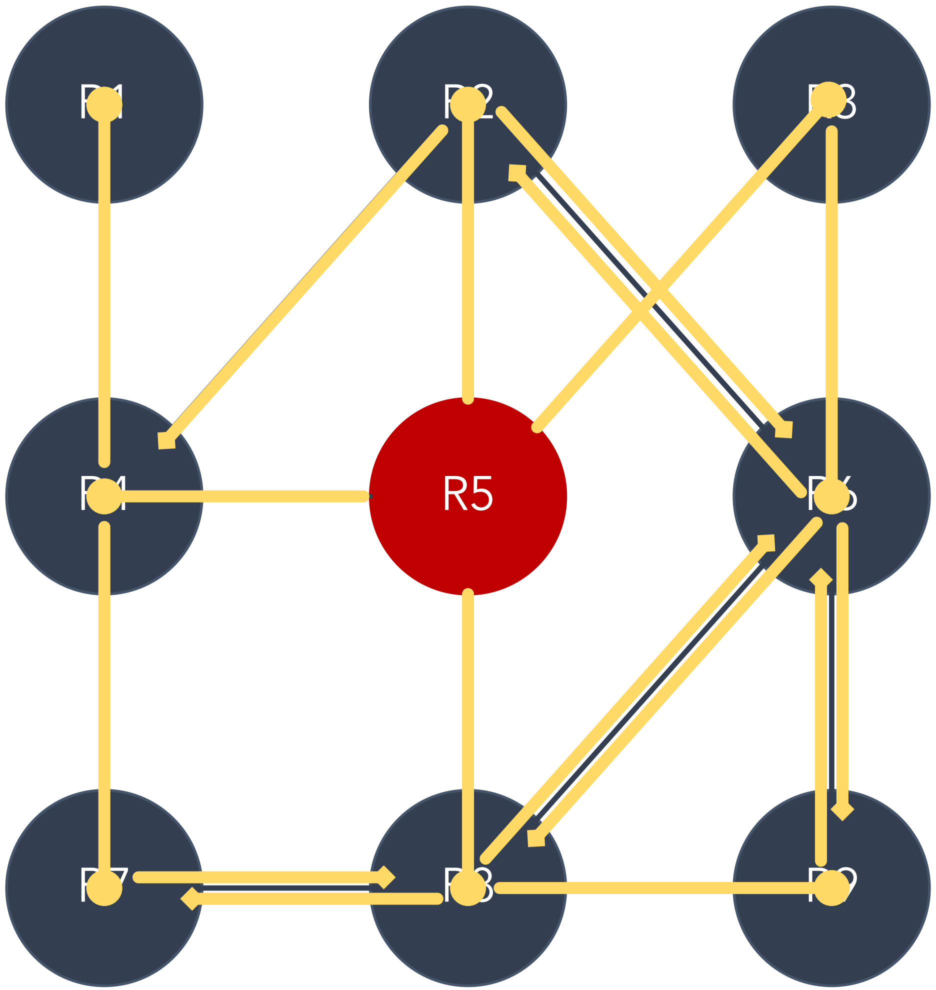


Link-state packet originated by R5

Adjacency	Metric
R3	10
R4	10
R8	10



Initial flooding of link-state packets (LSPs)

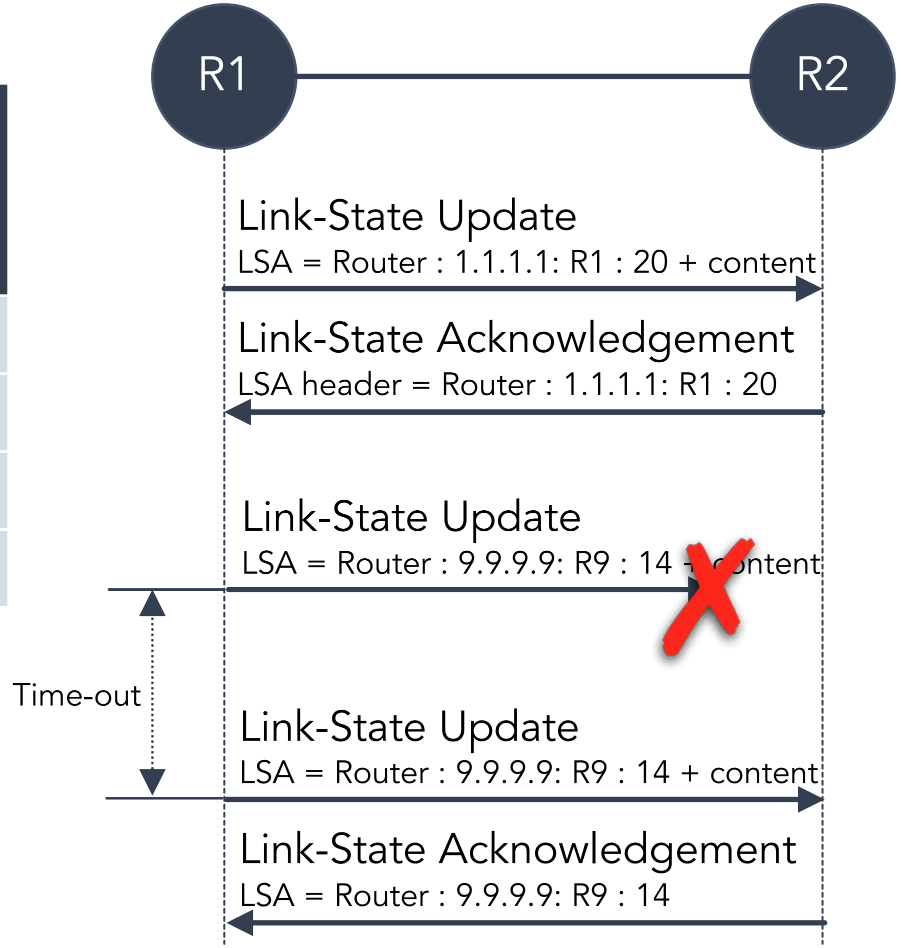


- LSP is identified by originator + type + identifier + version number
- Originator initiates flooding whenever LSP changes (e.g. adjacency up or down)
- Originator sends LSP to all adjacencies
- Receiver checks: do I have LSP version in database?
- If yes, ignore.
- If no:
  - Store LSP in database
  - Flood LSP to all adjacencies, except the one it was received from

Making the flooding reliable: acknowledgements

Link state database (LSDB) of R1

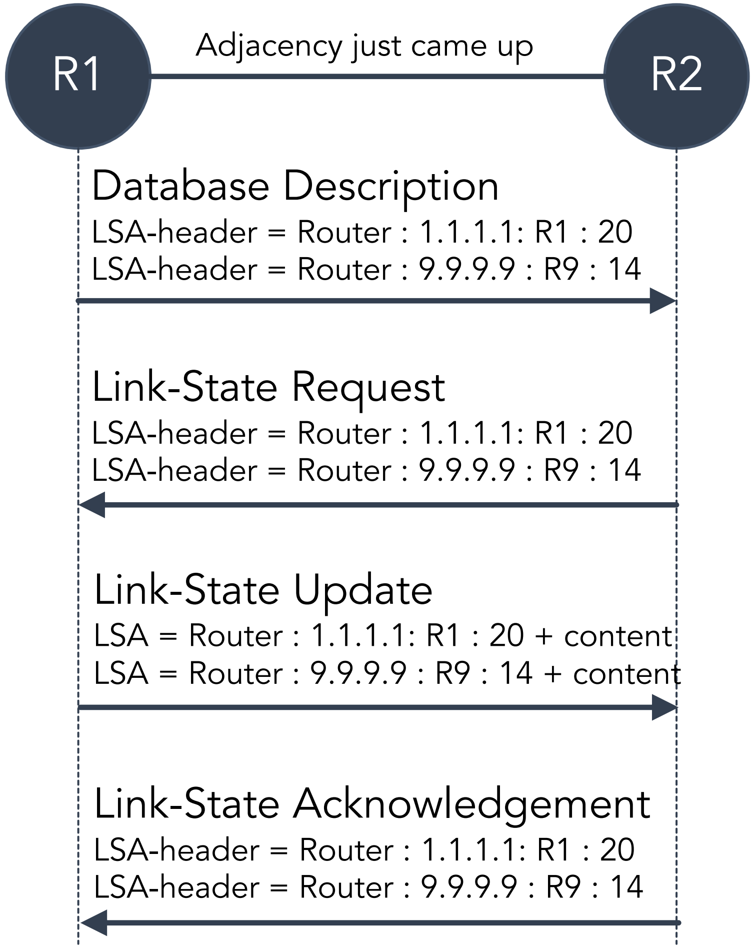
LSA Header				LSA Content
Type	ID	Advertising router	Sequence number	
…	…	…	…	…
Router-LSA	1.1.1.1	Router R1	20	…
Router-LSA	9.9.9.9	Router R9	14	…
…	…	…	…	…



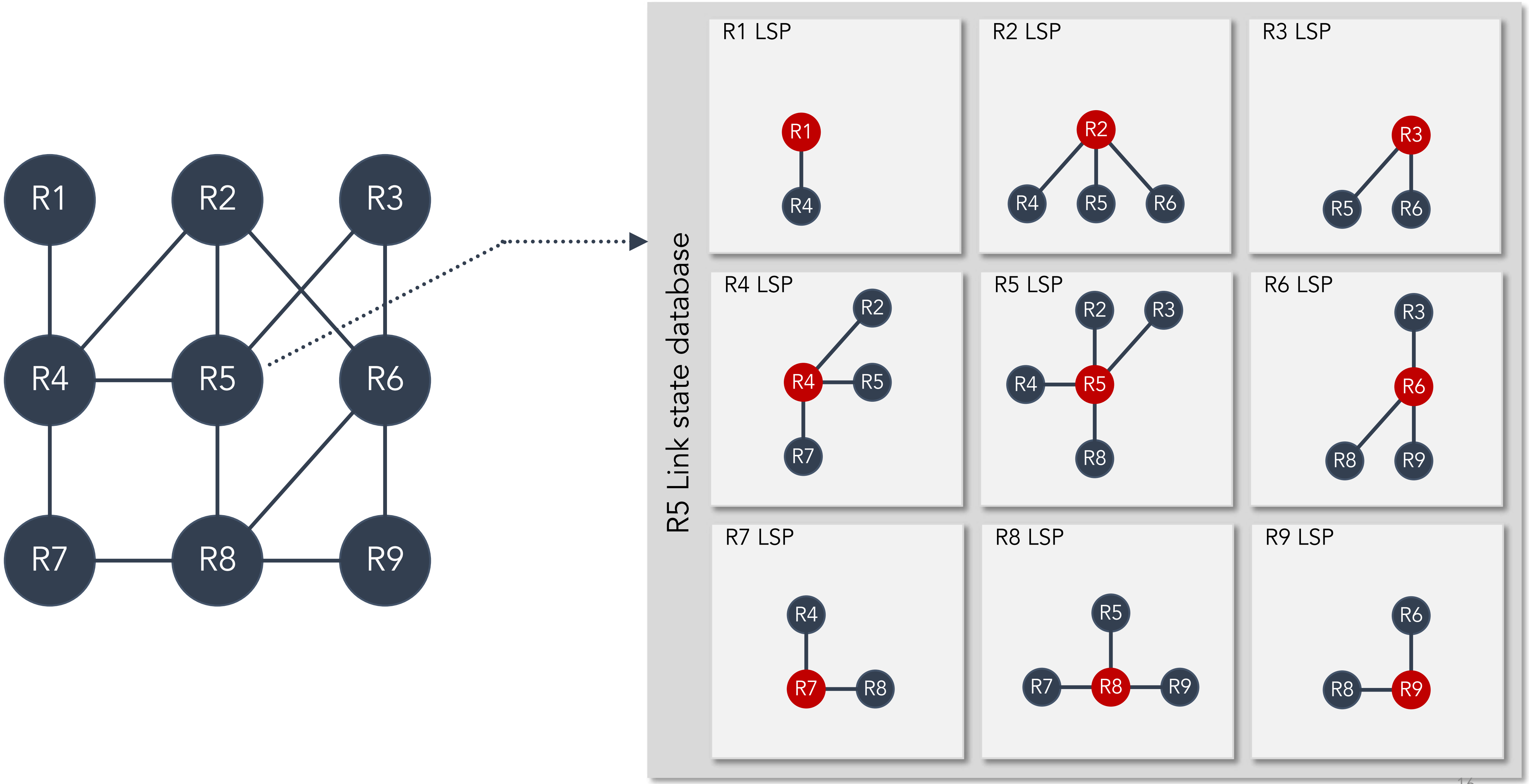
Initial sync: database description + request

Link state database (LSDB) of R1

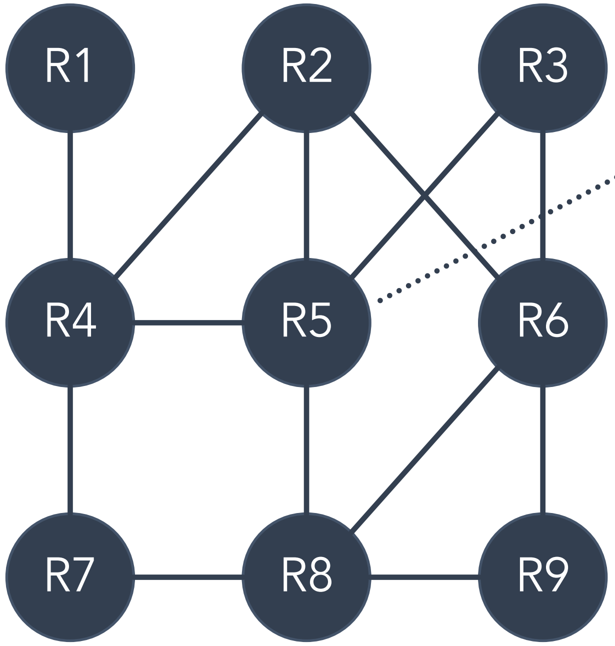
LSA Header				LSA Content
Type	ID	Advertising router	Sequence number	
…	…	…	…	…
Router-LSA	1.1.1.1	Router R1	20	…
Router-LSA	9.9.9.9	Router R9	14	…
…	…	…	…	…

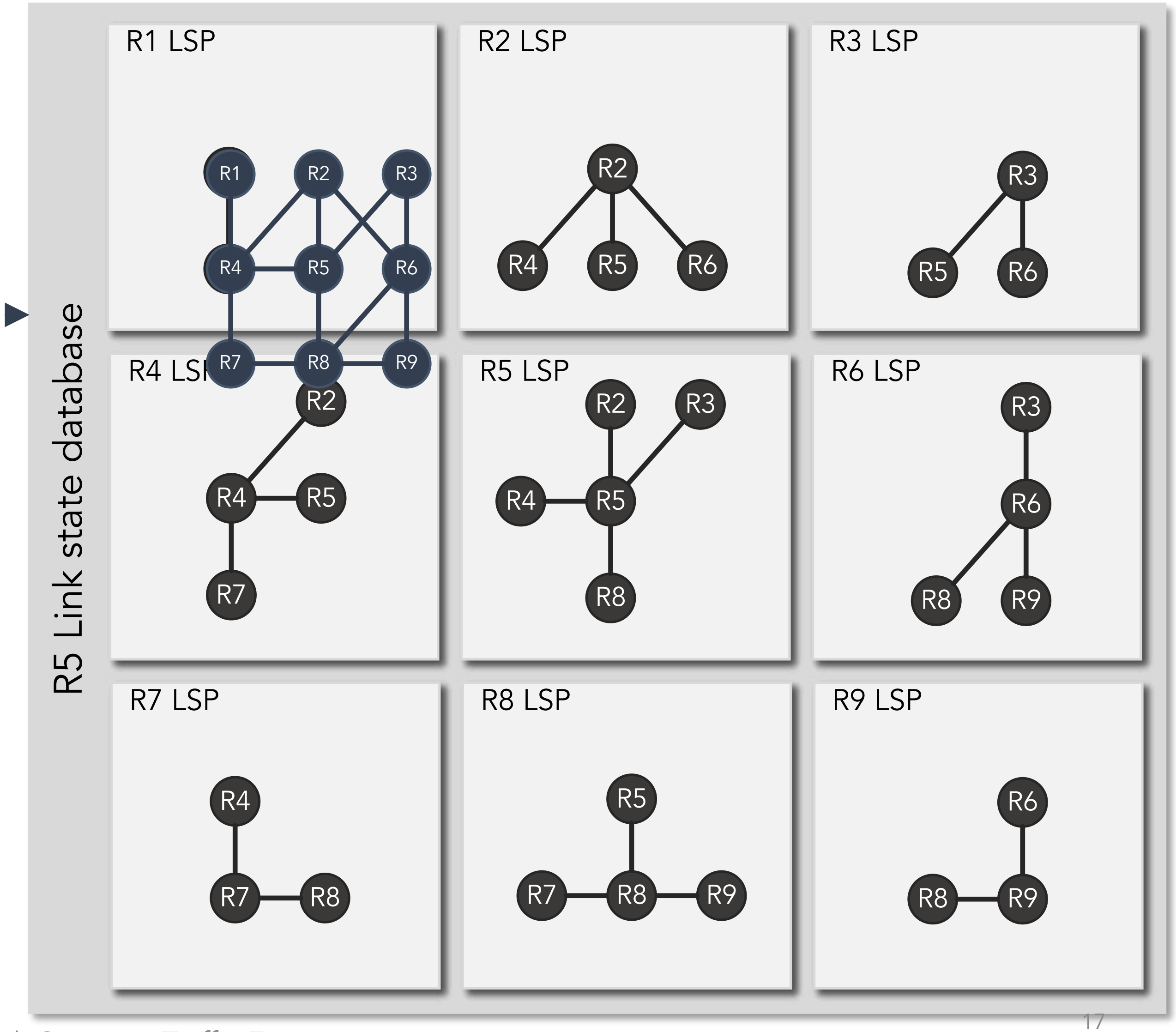


Each router has link-state database (LSDB)



Construct topology from link-state database

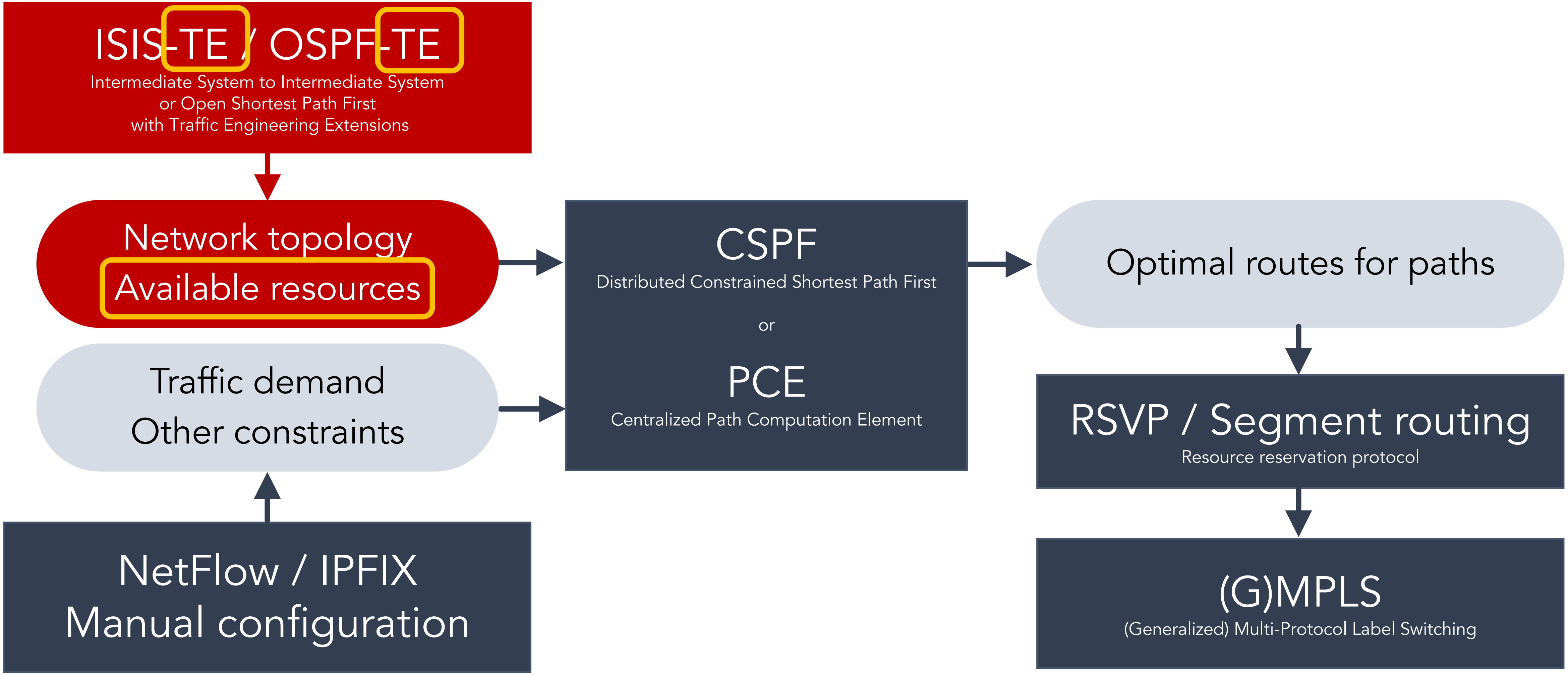




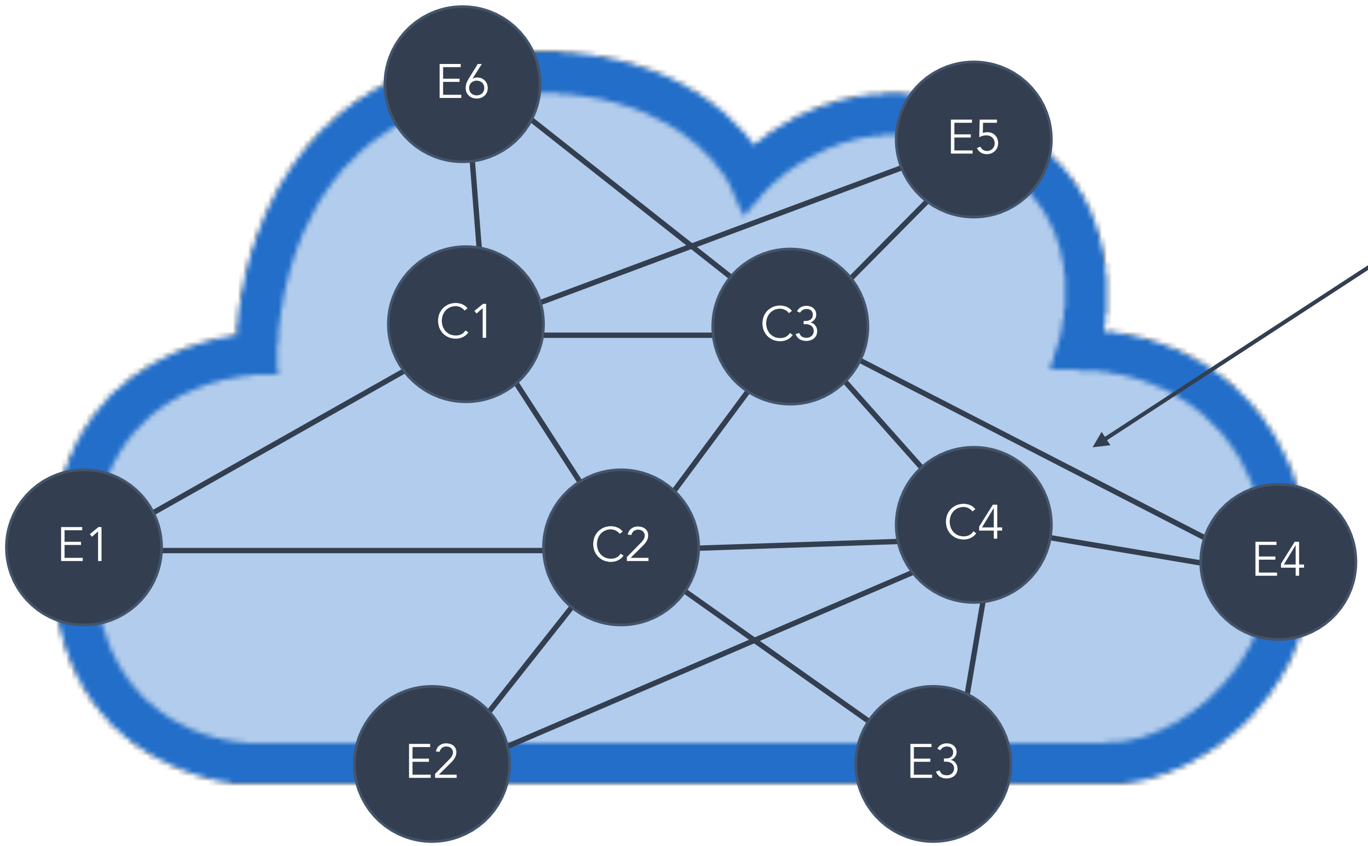
Lessons for quantum networks

- Quantum Networks contains both classical links and quantum links
- The classical topology and the quantum topology may be different
- These existing protocols already have multi-topology extensions
- <u>It makes sense to use / generalize OSPF or ISIS for topology discovery in Quantum Networks</u>

Resources discovery



Available bandwidth on every link

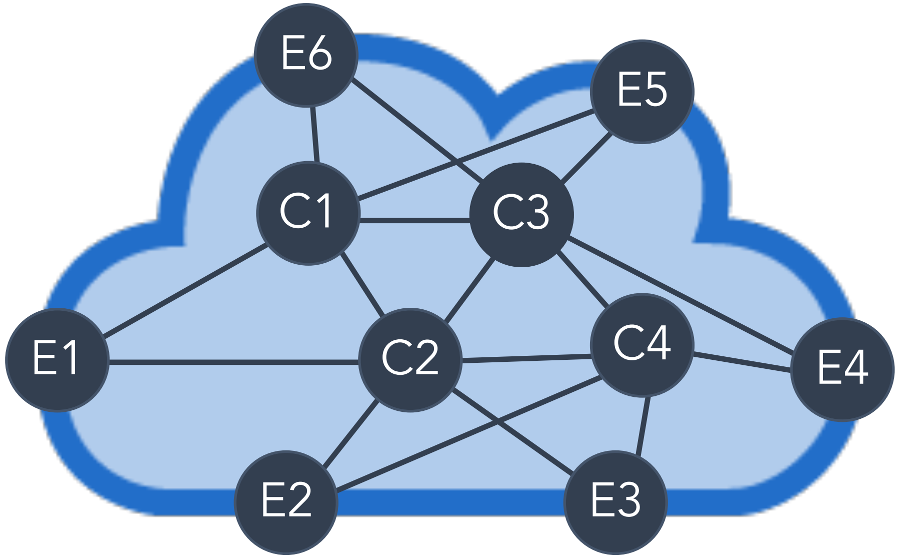


Physical bandwidth	400 Gbps
Reserved bandwidth	180 Gbps
Available bandwidth	220 Gbps

Other resources

- In classical networks, link bandwidth is the most important resource
  - Link bandwidth is tracked separately for each traffic priority
  - Router throughput is usually limited by the links (“non-blocking”)
- Administrative group (“color”) for each link
  - This is a high latency (satellite link)
  - The country in which the link is located (Canada, US, …)
- Shared Risk Link Group (SRLG)
  - This link is in the same physical conduit as those other links

Discovery of resources: OSPF-TE or ISIS-TE



Each node floods information about resource availability across the whole network:

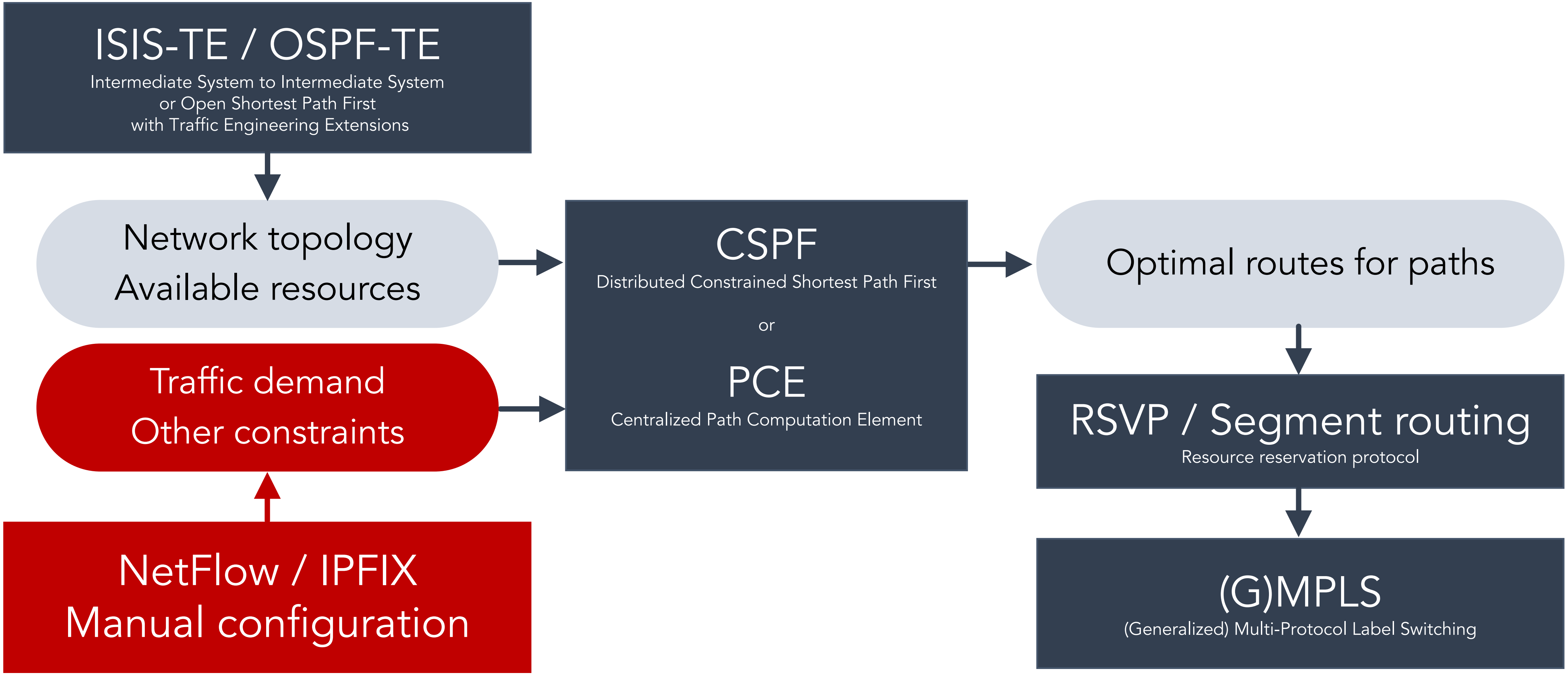
Traffic engineering attributes of leak link:
Traffic engineering metric
Maximum bandwidth
Maximum reservable bandwidth
Unreserved bandwidth
Administrative group (aka class or color)

Traffic Engineering (TE) extensions to link-state routing protocols: OSPF and ISIS

Topology discovery lessons for quantum networks

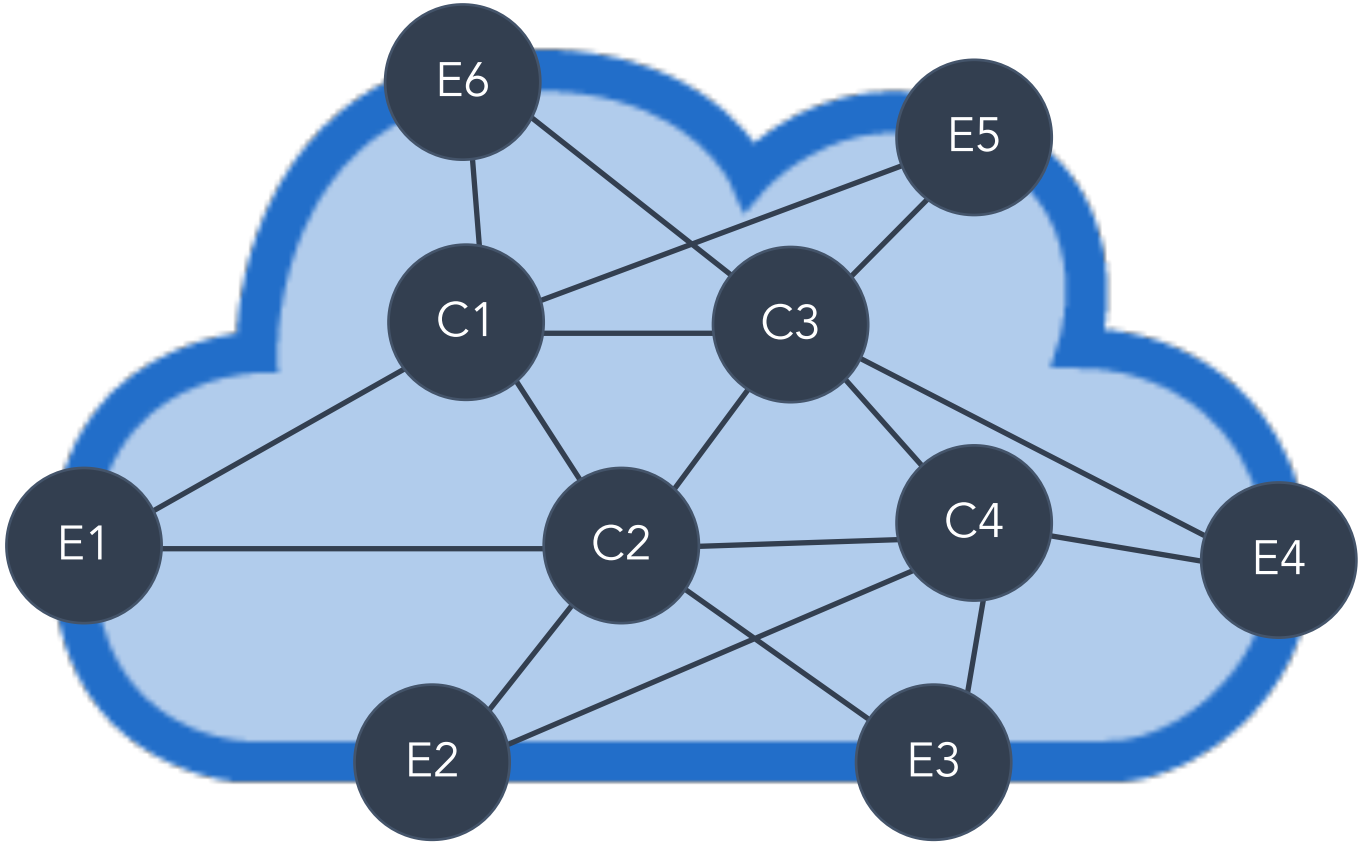
- Classical networks use Traffic Engineering extensions to OSPF or ISIS to flood information about resource availability in the network.
- There are new types of network resources that are unique to quantum networks (Bell pairs per second, number of memory qubits, non-deterministic swap success probability, …)
- <u>It makes sense to use / generalize OSPF-TE or ISIS-TE for resource discovery in Quantum Networks</u>
- <u>This work has already started:</u>
  IETF draft "Advertising Entanglement Capabilities in Quantum Networks" https://datatracker.ietf.org/doc/draft-kaws-qirg-advent/

Traffic demand and constraints discovery



Traffic demand matrix

Traffic demand matrix is typically measured (e.g. using NetFlow or IPFIX)



Ingress	Egress	Demand
E1	E2	120 Gbps
E1	E3	30 Gbps
E1	E4	20 Gbps
E1	E5	20 Gbps
E1	E6	90 Gbps
E2	E1	40 Gbps
E2	E3	190 Gbps
E2	E4	80 Gbps
E2	E5	70 Gbps
E2	E6	30 Gbps
…	…	…

How to determine the traffic demand?

- **Measure demand**
  - Using NetFlow, IPFIX, etc.
  - Take peak demand for diurnal patterns
  - Adjust for expected changes
- **Schedule demand**
  - Have list of jobs to be completed (e.g. huge file transfers)
  - Each job has known bandwidth demand and deadline
  - Schedule jobs to meet deadline and maximize network utilization

Additional constraints

- Example use cases:
  - Primary LSP and backup LSP must use node-disjoint or link-disjoint paths
  - Traffic between Canadian cities cannot go through USA
  - Traffic must not use high-latency satellite links
  - Traffic must avoid router which is known to need scheduled maintenance
  - Etc.
- RSVP-TE uses link "colors" for this purpose
- Centralized controllers can handle more sophisticated constraints

Specify path requirements: configuration

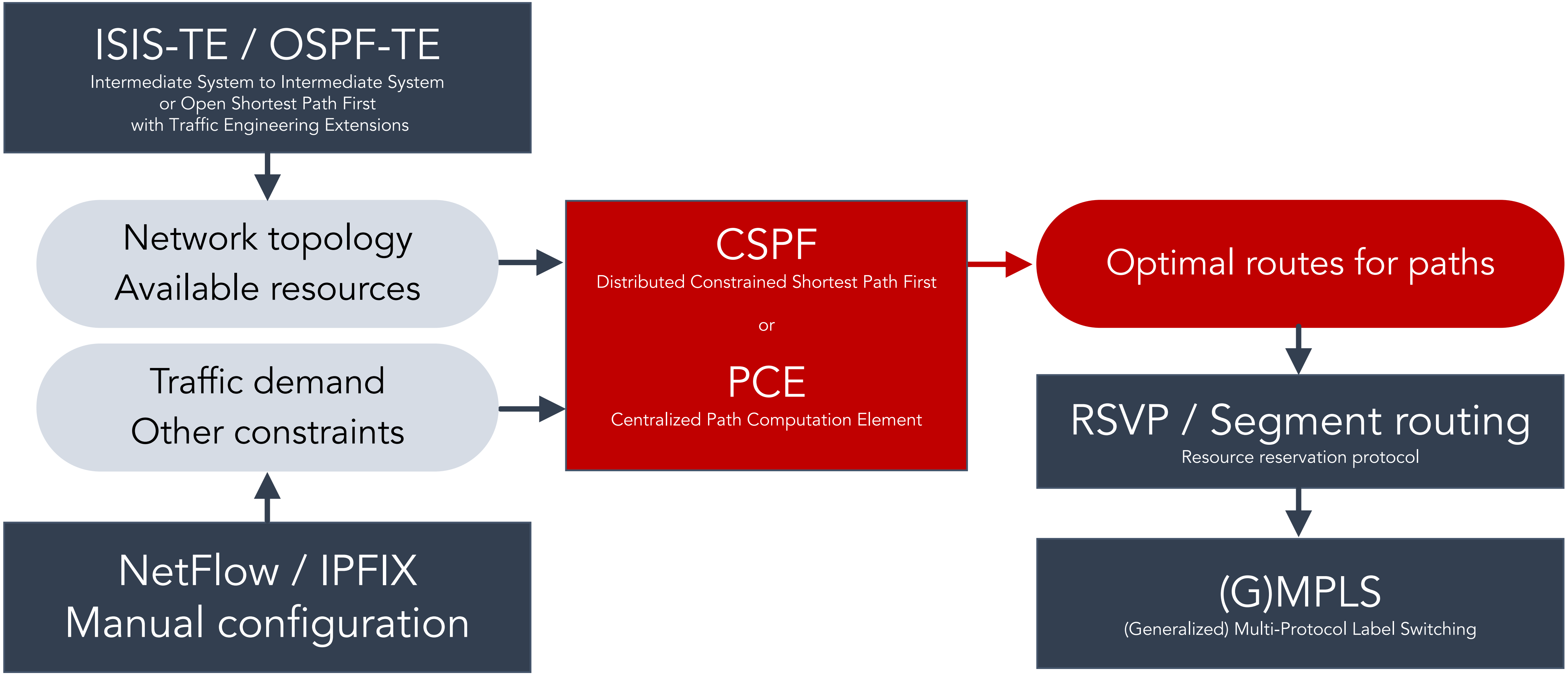
Configure path requirements on the ingress node (head) of the path

```
protocols {
  mpls {
    label-switched-path my-lsp {
      to 10.0.9.7;
      bandwidth 100m;
      exclude [ satellite-links ];
      include [ usa-links canada-links ];
    }
  }
}
```

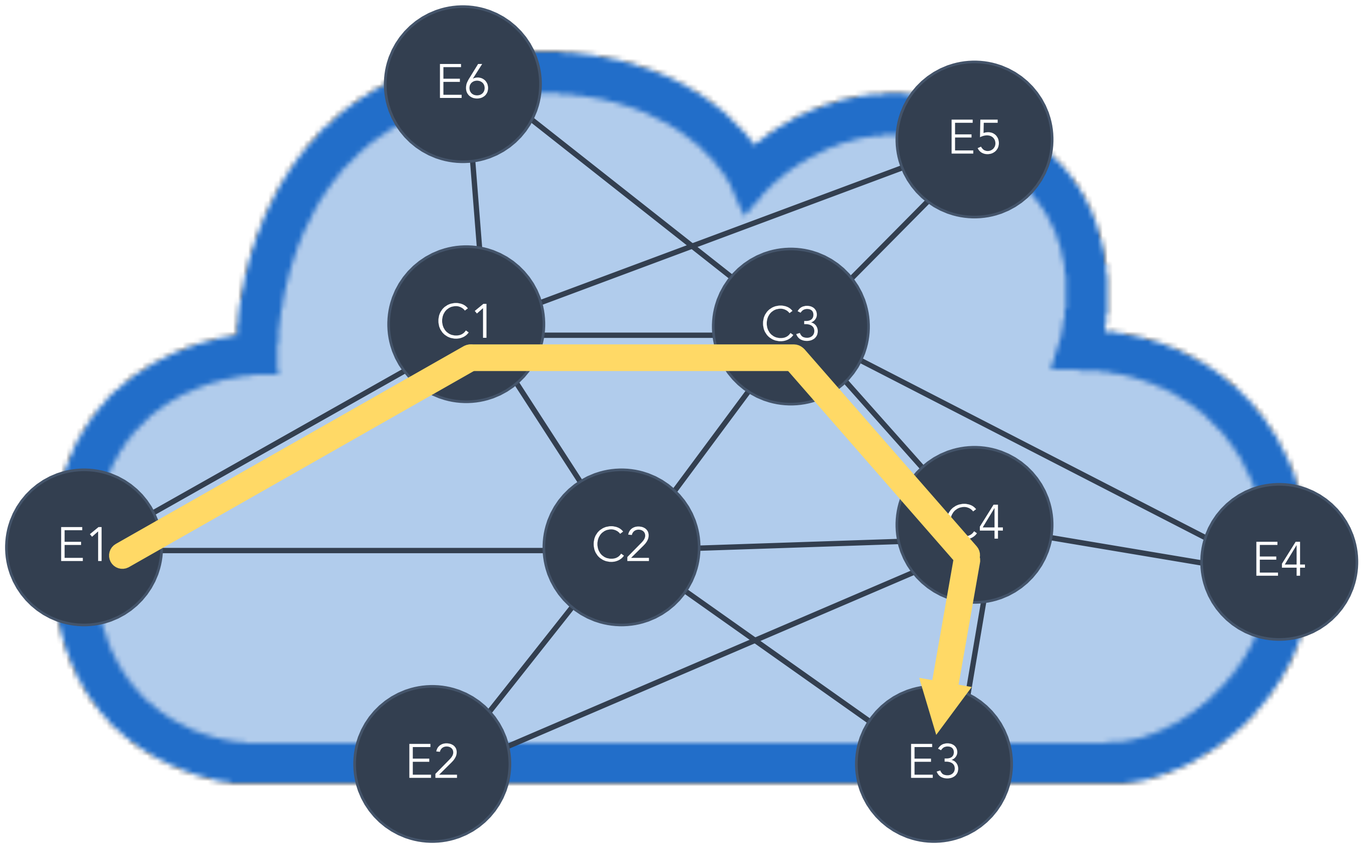
Demand discovery lessons for quantum networks

- If we want to use traffic engineering in quantum networks, we need to know the quantum “demand matrix”:
  - How many end-to-end Bell pairs per second (BPPS) between end-points?
  - What is the required fidelity of those Bell pairs (similar to QoS in classical networks).
- The applications can specify the demand when they request a path (integrated services model).
- We can use and out-of-band method to create a demand matrix for the aggregate of all quantum flows (differentiated services model).
- The bandwidth of initial quantum networks will be so scarce that the integrated services model makes sense
  (Even though the classical Internet has abandoned the integrated services model and only uses the differentiated services model now.)

Path computation



Path computation



Given:

- Required bandwidth for path
- Other constraints
- Available bandwidth on each link
- Other resource attributes

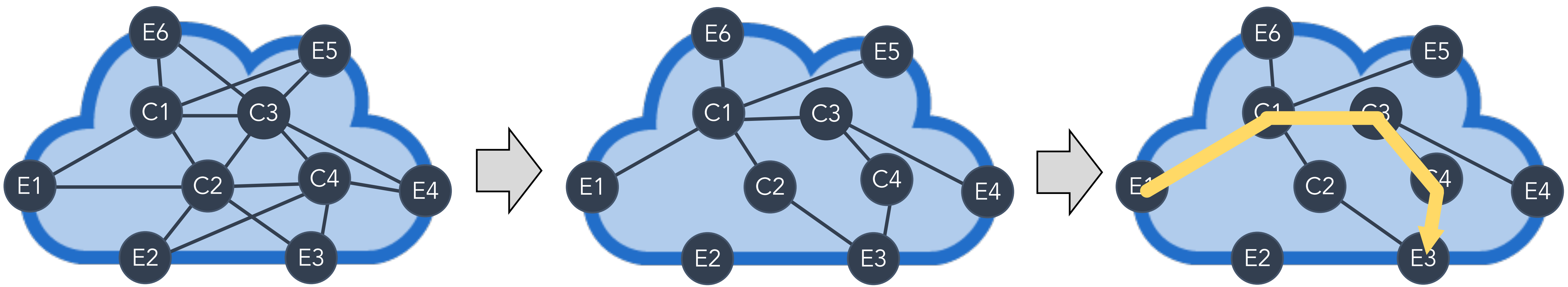
Do the following:

- Compute the best route of the path
- Signal the path in the network
- Program the path into the forwarding plane

Distributed algorithm: Constrained SPF (CSPF)

Remove links from topology that do not satisfy constraints

- Bandwidth < 100 mbps
- Remove "satellite" links
- Remove all links outside "usa" or "canada"

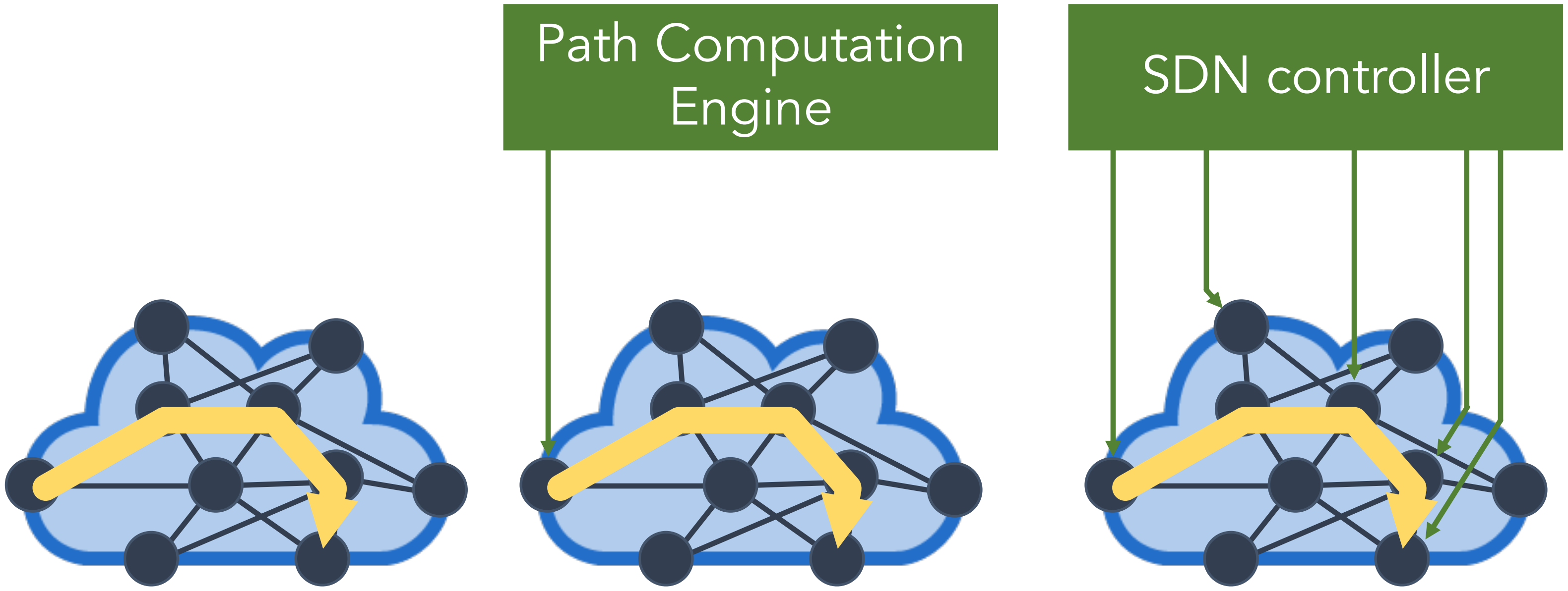


Full information in link-state database: all links

Pruned information in link-state database: only links that satisfy constraints

Computed shortest path on remaining topology: E1→C1 →C3 →C4 →E3

Centralization of path computation

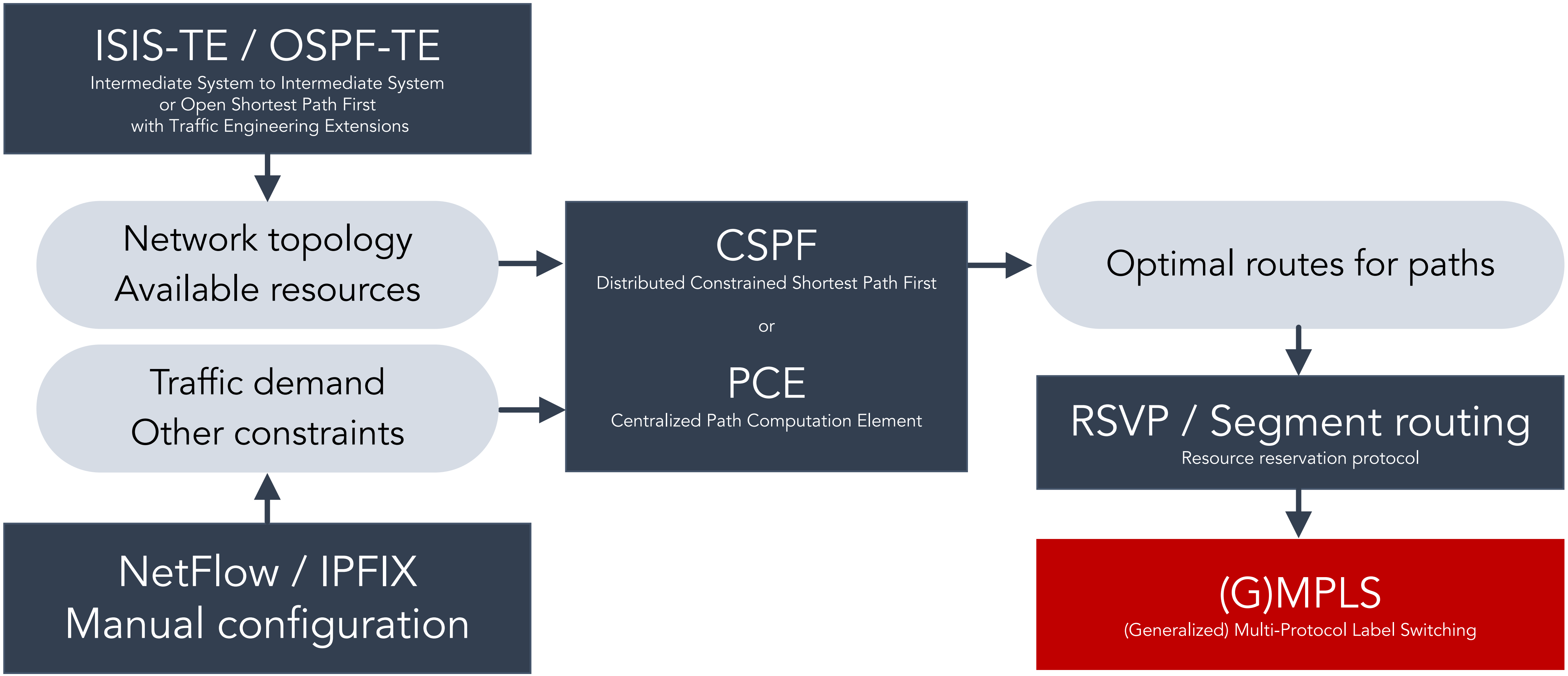


Path Computation	Distributed	Centralized	Centralized
Path Signaling	Distributed	Distributed	Path not signaled
Program forwarding plane	Distributed	Distributed	Centralized

Path computation lessons for quantum networks

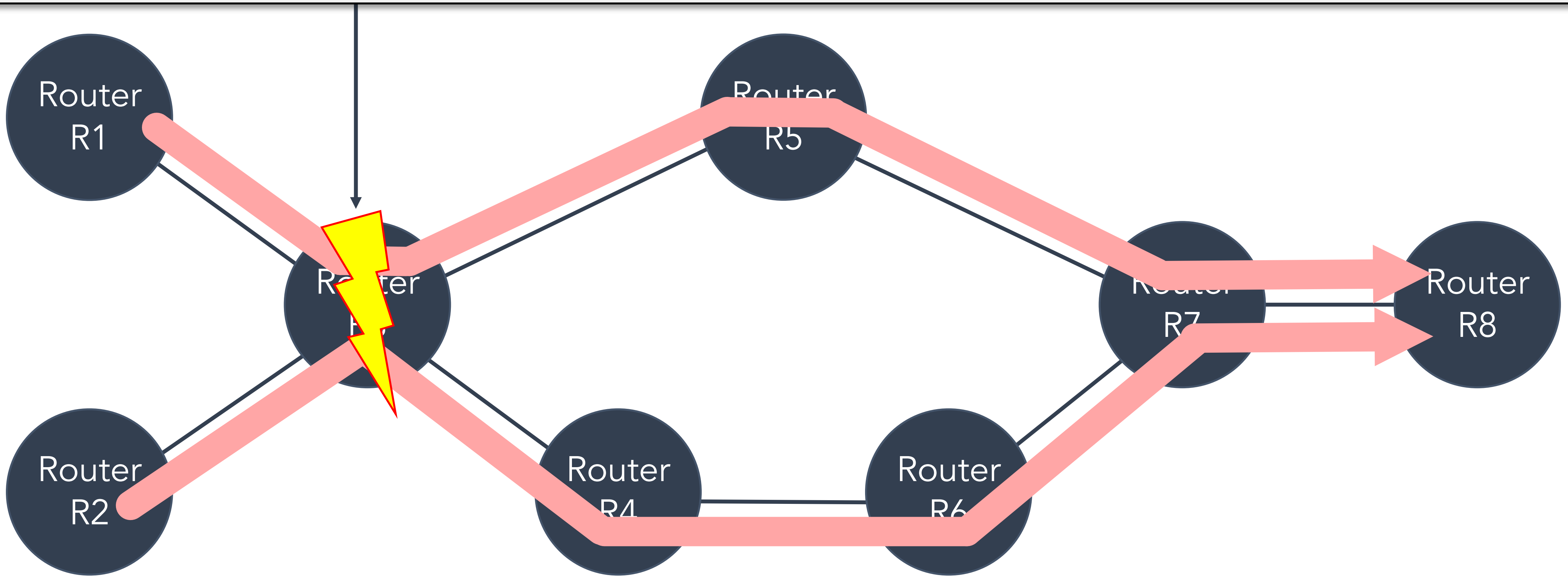
- <u>General concept of path computation applies to quantum networks</u>
- Path computation is more complex in quantum networks
  - Limited number of memory qubits
  - Non-deterministic swaps
  - Distillation
  - Etc.
- New path computation algorithms and heuristics will be needed.
- Logically centralized path computation element makes sense.

Data plan for traffic engineered networks



Difficult to traffic engineer with IP alone

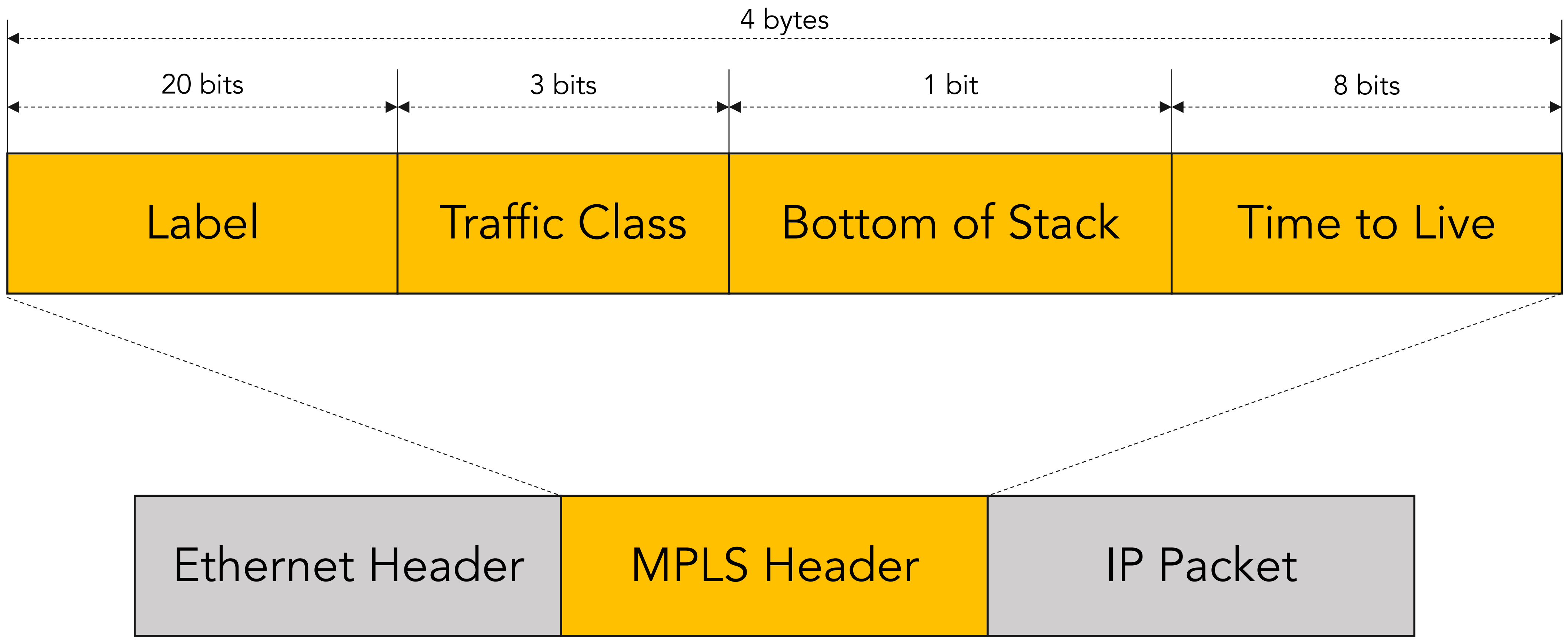
IP forwarding table can**not** have two different routes that depend on the source IP address for the same destination IP address.



Multi-Protocol Label Switching (MPLS)

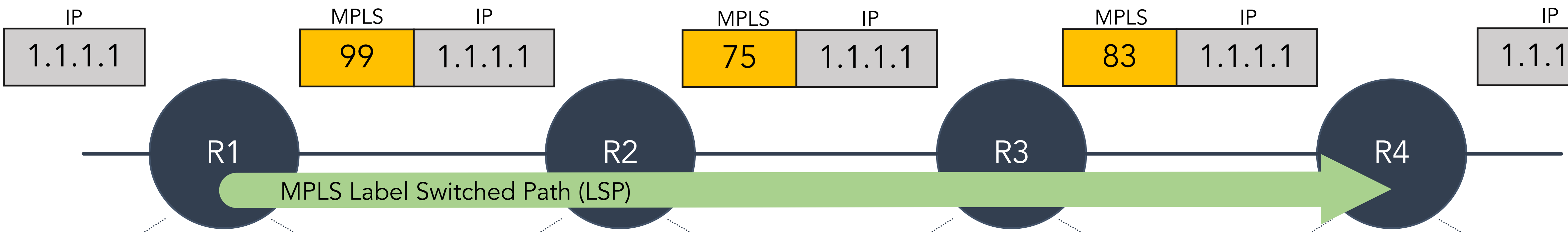
- A new data plane protocol that uses "MPLS labels"
- Used to create Label Switched Paths (LSPs)
- Enables Traffic Engineering (TE) and many other services
- Routers that support MPLS are called Label Switched Routers (LSRs)
- MPLS runs over any layer 2 protocol (not just Ethernet)
- Any protocol can run over MPLS (not just layer 3 but also layer 2)
- Referred to as a layer 2½ protocol
- Works with many control plane protocols (RSVP, LDP, BGP, SDN, …)

MPLS header



4 bytes
20 bits
3 bits
1 bit
8 bits
Label
Traffic Class
Bottom of Stack
Time to Live
Ethernet Header
MPLS Header
IP Packet

MPLS label switching



Label	Next-hop

MPLS forwarding table

Label	Next-hop
99	Swap 75, R3

MPLS forwarding table

Label	Next-hop
75	Swap 83, R4

MPLS forwarding table

Label	Next-hop
83	Pop

MPLS forwarding table

IP	Next-hop
1.1.0.0/8	Push 99, R2

IP forwarding table

IP	Next-hop

IP forwarding table

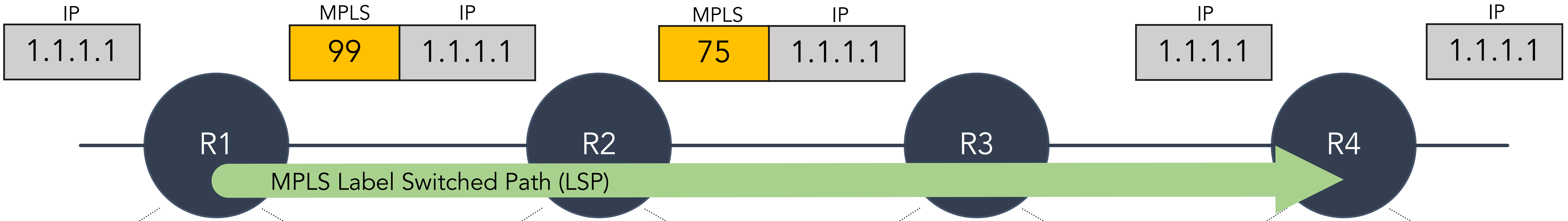
IP	Next-hop

IP forwarding table

IP	Next-hop
1.1.0.0/8	R5

IP forwarding table

Penultimate Hop Popping (PHP)



R1

Label	Next-hop

MPLS forwarding table

IP	Next-hop
1.1.0.0/8	Push 99, R2

IP forwarding table

R2

Label	Next-hop
99	Swap 75, R3

MPLS forwarding table

IP	Next-hop

IP forwarding table

R3

Label	Next-hop
75	Pop, R4

MPLS forwarding table

IP	Next-hop

IP forwarding table

R4

Label	Next-hop

MPLS forwarding table

IP	Next-hop
1.1.0.0/8	R5

IP forwarding table

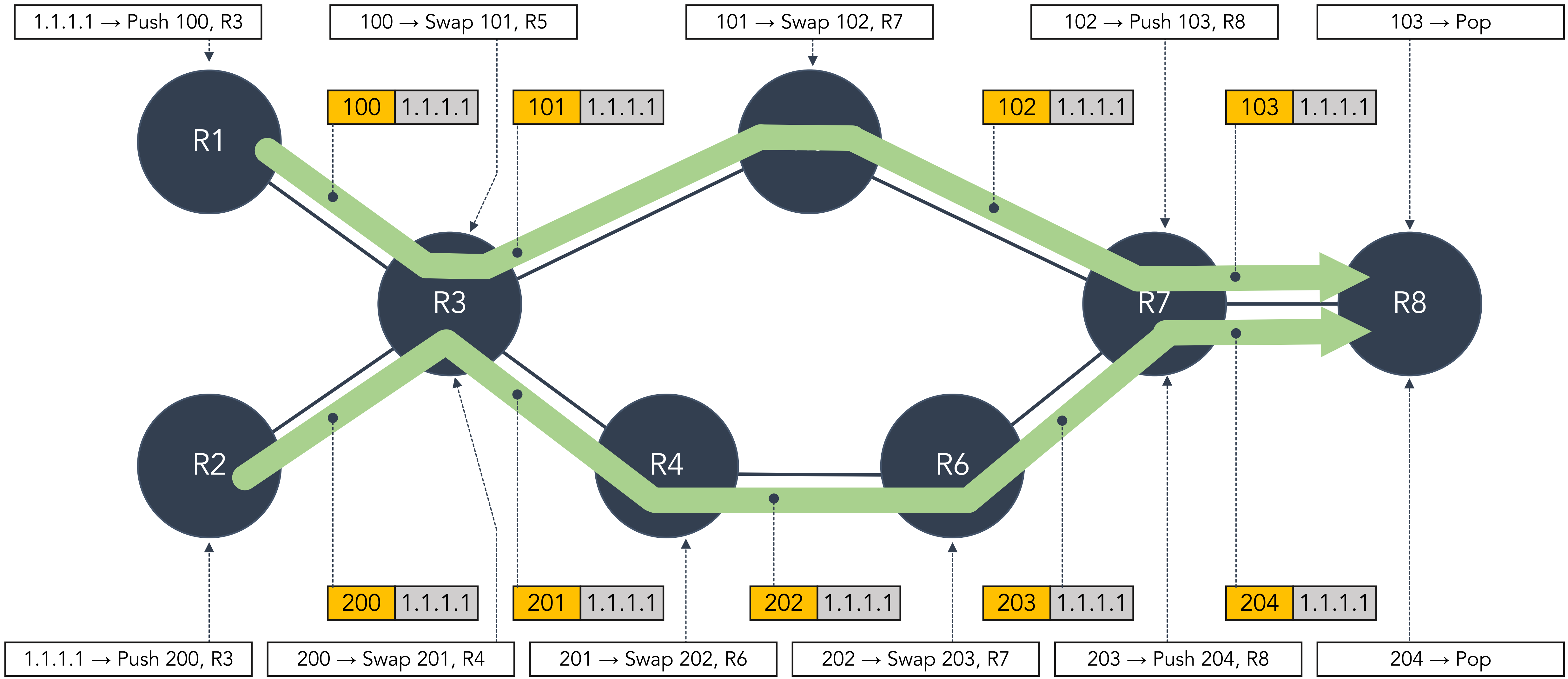
MPLS Label Stacking

MPLS label stack of depth 3

	Top label (= outer-most label)		Bottom label (= inner-most label)	
Ethernet Header	MPLS Header	MPLS Header	MPLS Header	IP Packet
Dest Ethernet Address a8:66:7f:3a:2b:1a	MPLS label 99	MPLS label 103	MPLS label 75	Dest IP Address 1.1.1.1

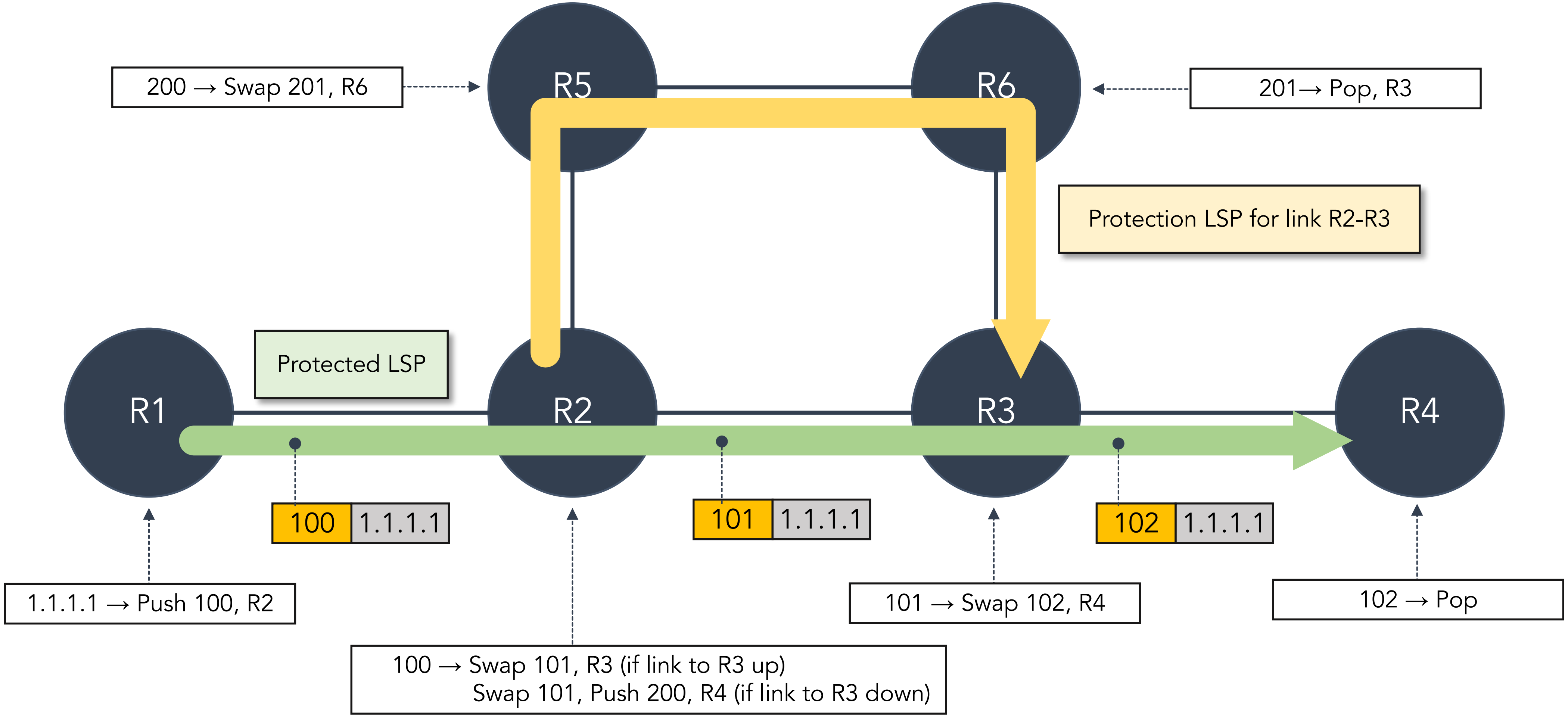
Complex services using simple building blocks

Traffic Engineering (TE)



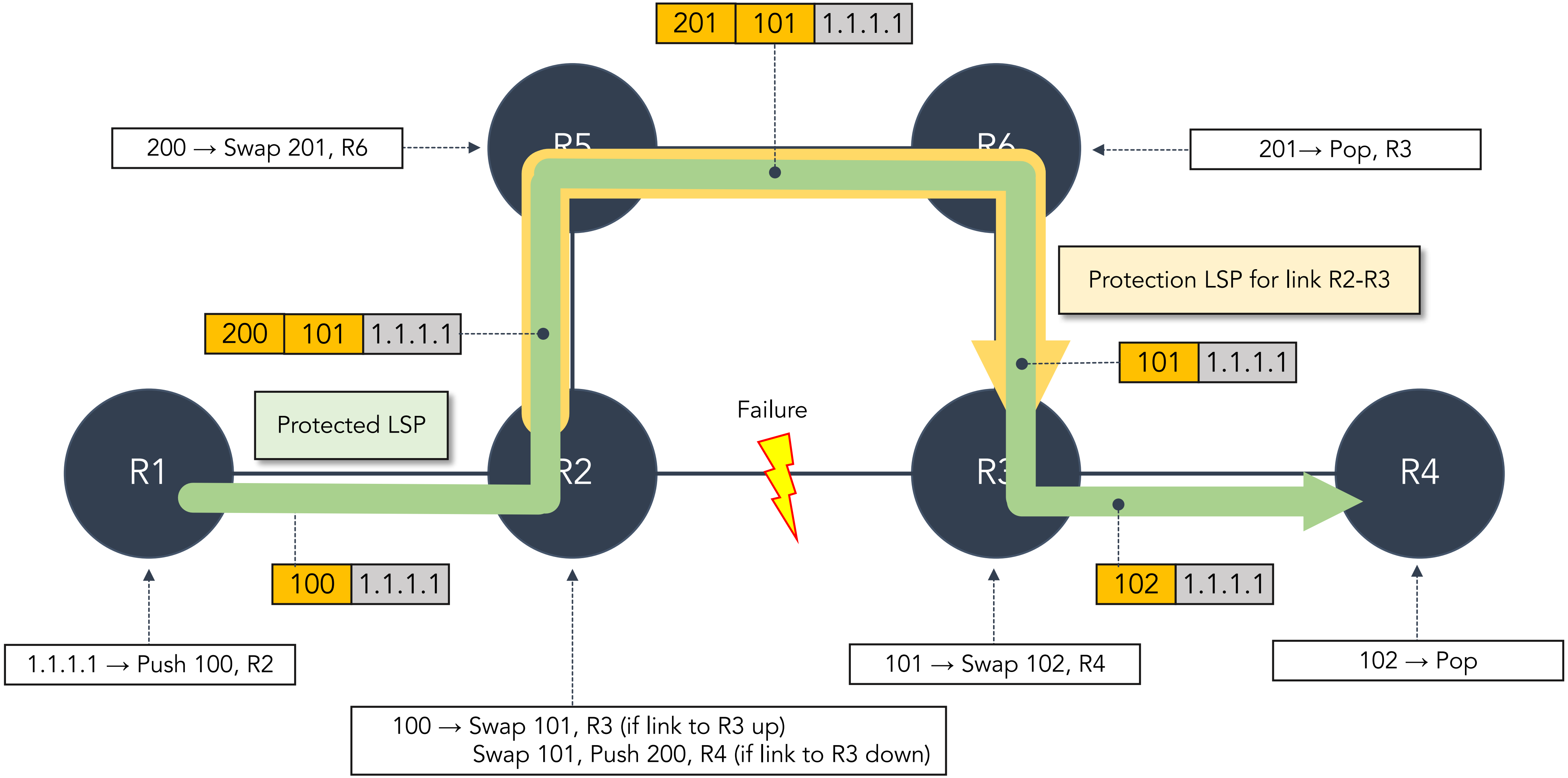
Complex services using simple building blocks

Fast Re-Route (FRR)



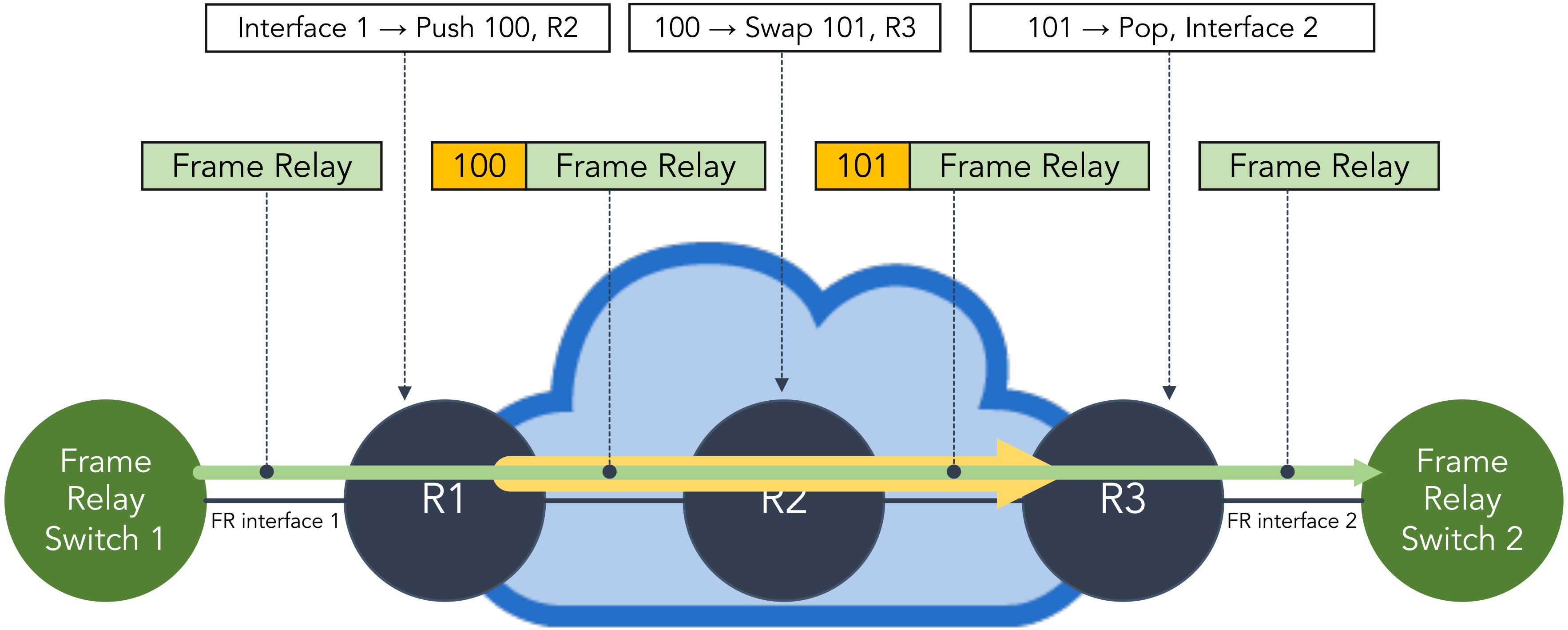
Complex services using simple building blocks

Fast Re-Route (FRR)



Complex services using simple building blocks

Pseudo Wire (PW)

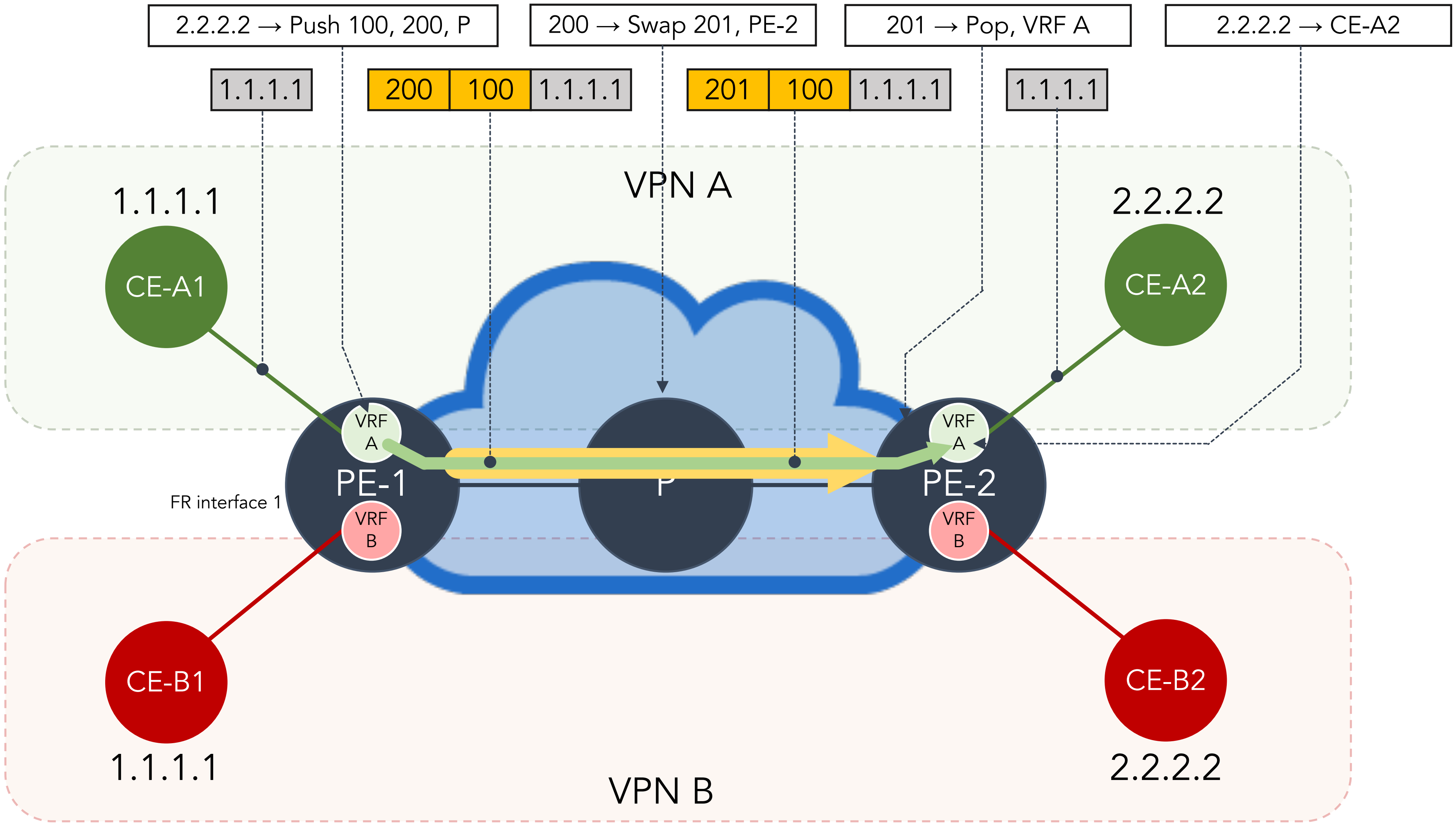


Allows service provider to offer Frame Relay service without a Frame Relay network

Complex services using simple building blocks

MPLS Virtual Private Networks (VPNs)

VPNs must be isolated and can have overlapping IP address spaces



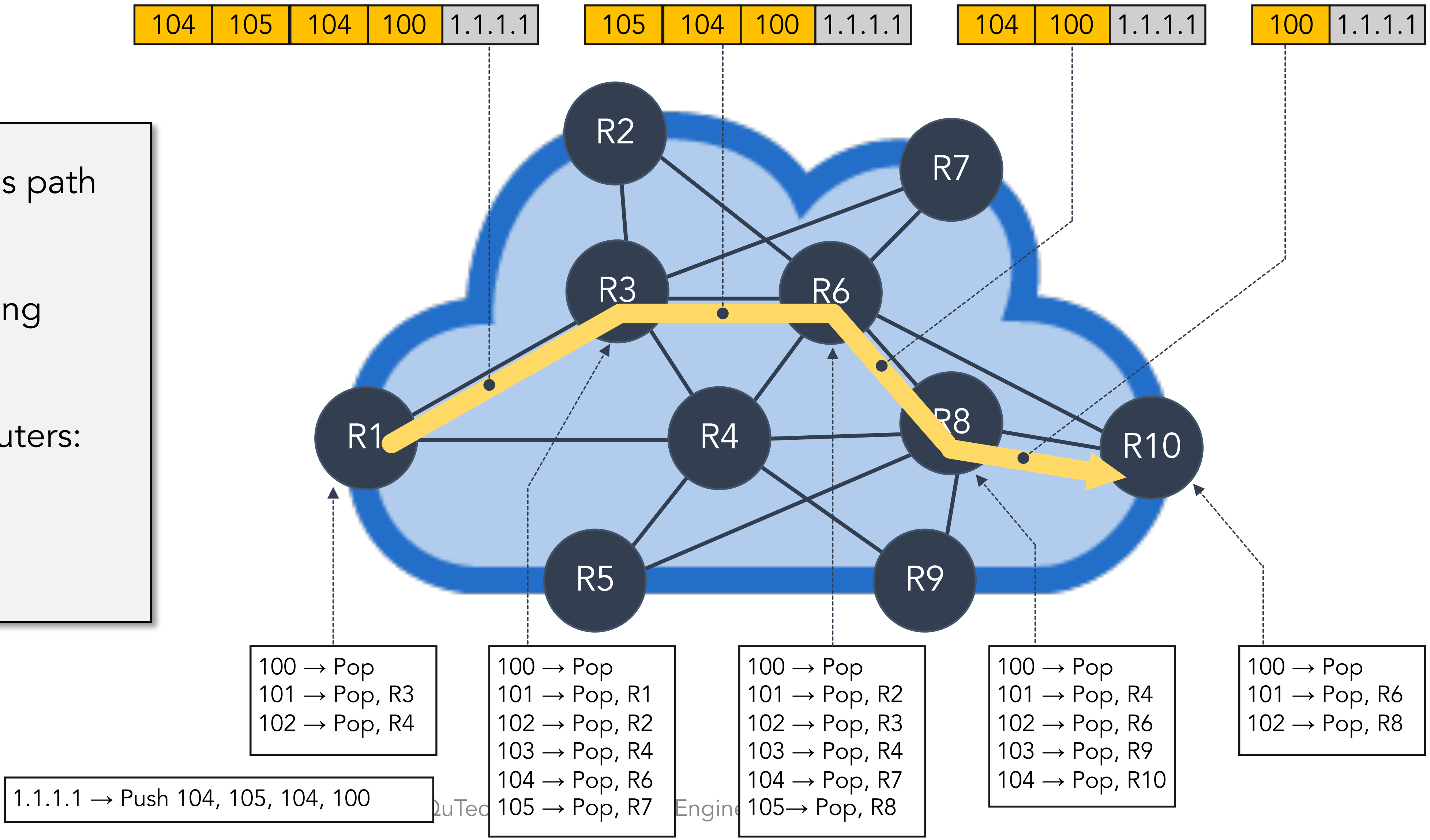
Complex services using simple building blocks

Segment Routing (SR)

Sender encodes path into packet.

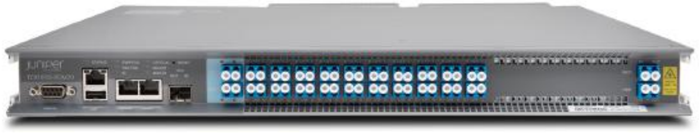
No path signaling required.

Less state in routers:
O(#neighbors)
Instead of
O(#paths)



Generalized MPLS (GMPLS)

Generalization of MPLS when data plane cannot carry explicit labels
Instead the label is implicit



ROADM: Wavelength is implicit label

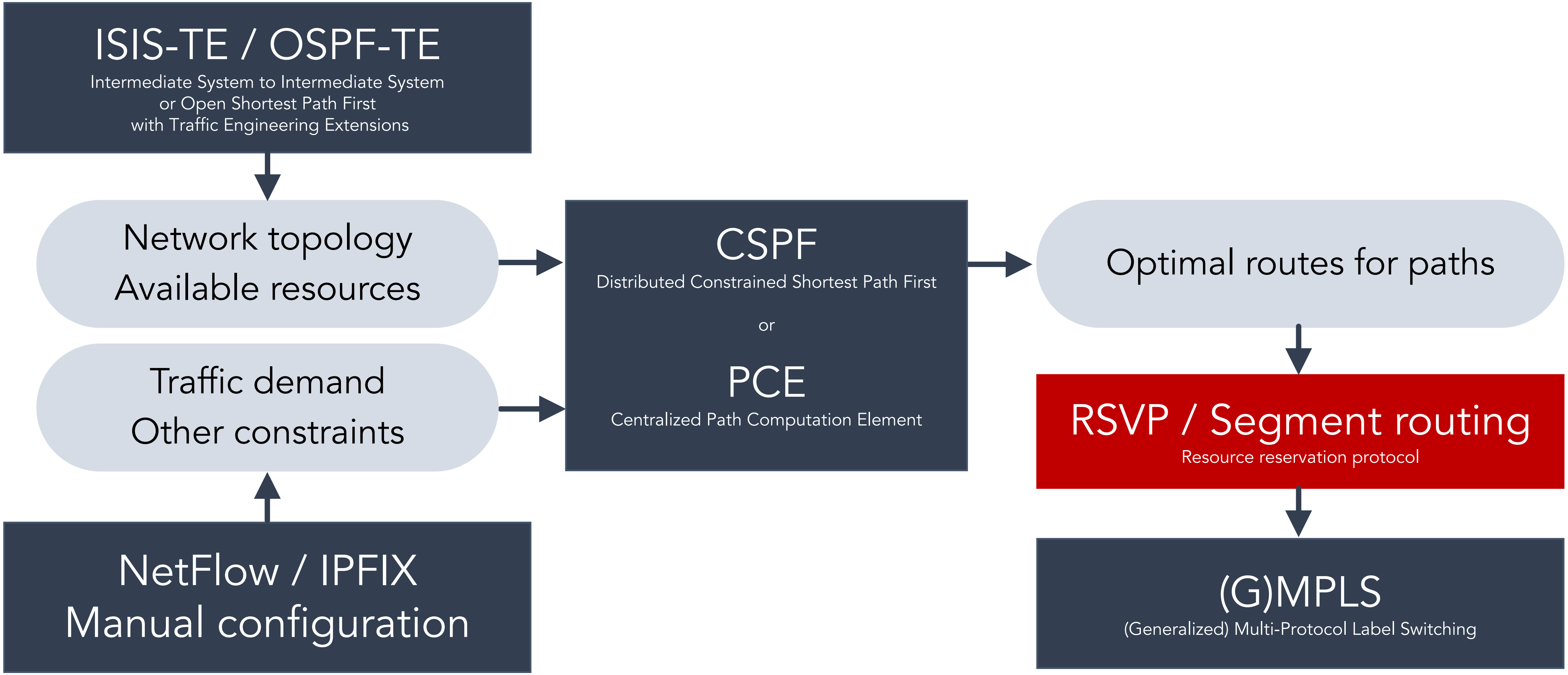
OXC: Optical port is implicit label

SONET ADM: TDM timeslot is implicit label

Data-plane lessons for quantum networks

- The data plane should use very simple but flexible primitives
  Just like MPLS push, swap, pop
- That support many different services
  Just like MPLS TE, FRR, PW, VPN, SR, …
- That are independent of the signaling protocol
  Just like MPLS is independent of RSVP, LDP, BGP, …
- That don't assume that data plane can carry control headers
  Just like GMPLS
  Qubits don't have headers; need some implicit label (e.g. timeslot)
- The quantum data plane will need to be stateful
  Unlike IP and MPLS which are stateless
  Needed because the creation of a single end-to-end Bell pair is a stateful multi-step process: multiple point-to-point Bell pairs, distillation, swaps, etc.
- Proposal: use P4 and add quantum and statefulness extensions
  In research phase, quantum protocols will change often

Signaling protocol for traffic engineered paths



MPLS signaling protocols

- Resource reservation protocol with traffic engineering extensions (RSVP-TE)
- Label distribution protocol (LDP)
- Targeted label distribution protocol (T-LDP)
- Border gateway protocol labeled unicast (BGP-LU)
- Border gateway protocol layer 3 virtual private networks (BGP L3VPN)
- Border gateway protocol layer 2 virtual private networks (BGP L2VPN)
- Border gateway protocol layer ethernet virtual private networks (BGP EVPN)
- Border gateway protocol labeled unicast (BGP LU)
- OpenFlow

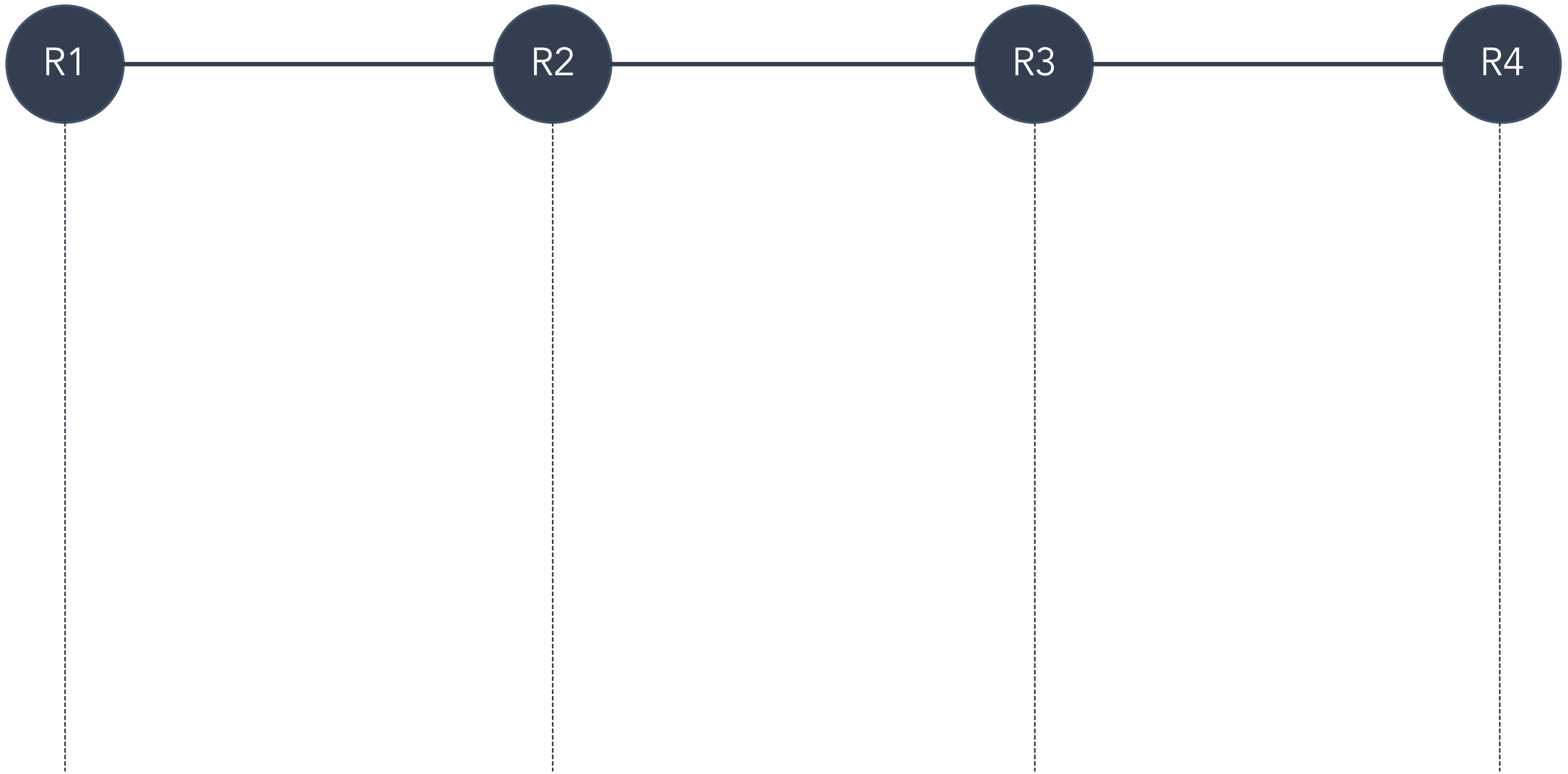
Label allocation concepts

- Downstream vs upstream
- On-demand vs unsolicited
- Ordered vs unordered
- Conservative vs liberal retention

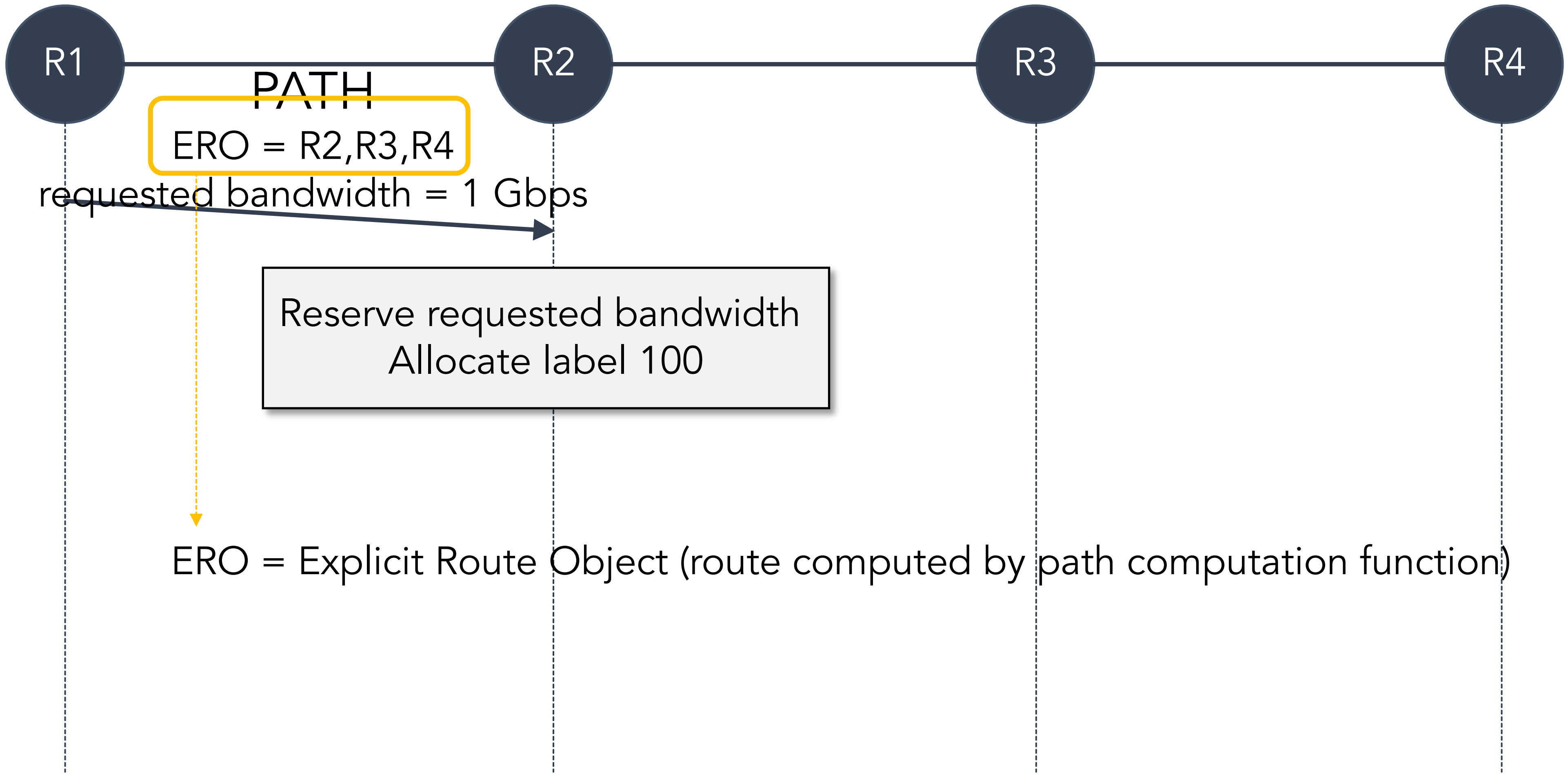
Resource reservation protocol (RSVP-TE)

- RSVP was originally designed for integrated services (IntServ)
  Bandwidth reservation for individual flows of traffic
- RSVP-TE is currently used for differentiated services (DiffServ)
  Strategic / tactical bandwidth reservation of aggregates of many flows
- Relies on OSPF-TE or ISIS-TE for path computation
- Relies on MPLS or GMPLS for data plane
- Uses downstream-on-demand label allocation with conservative retention mode

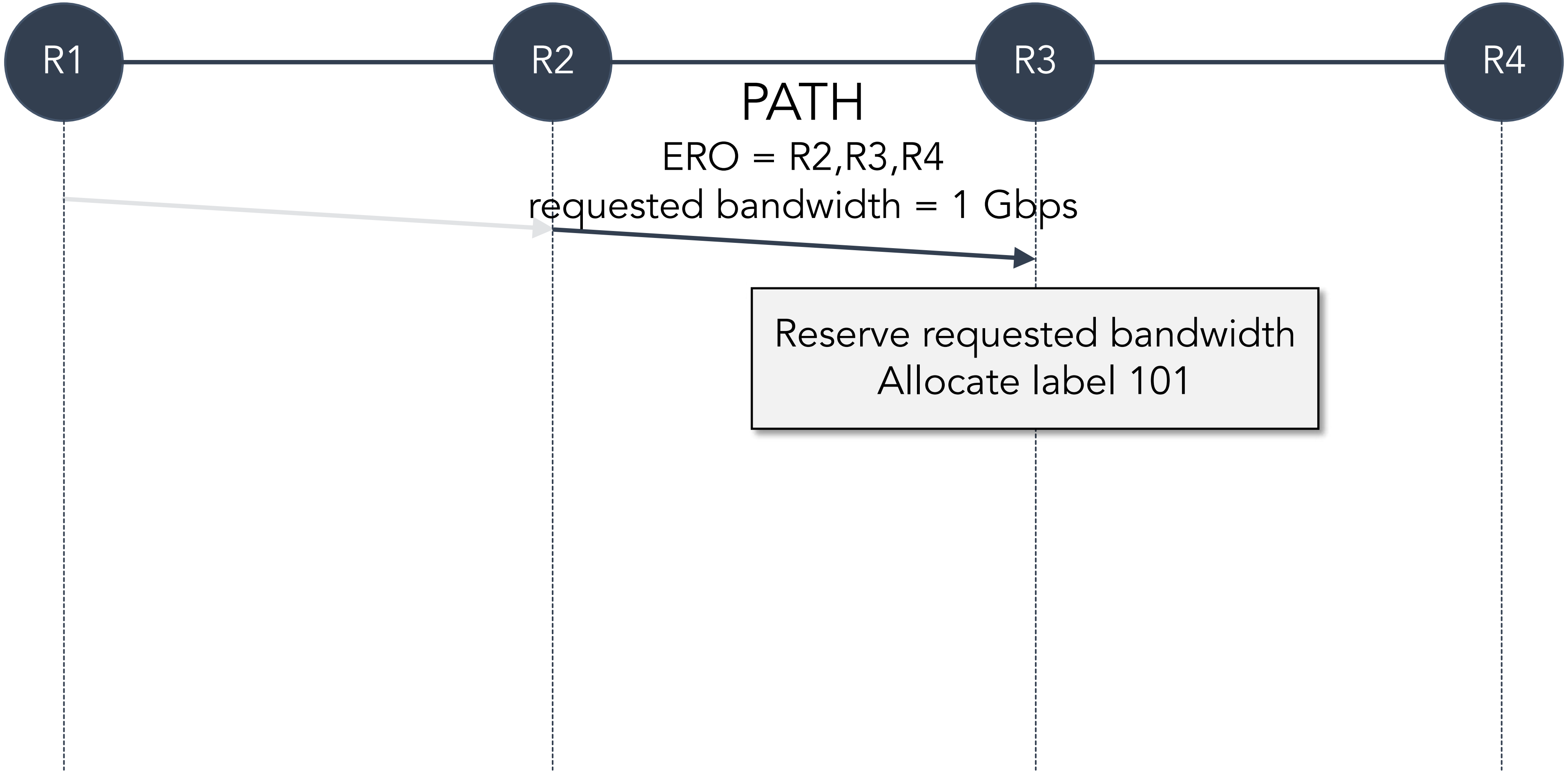
RSVP path signaling



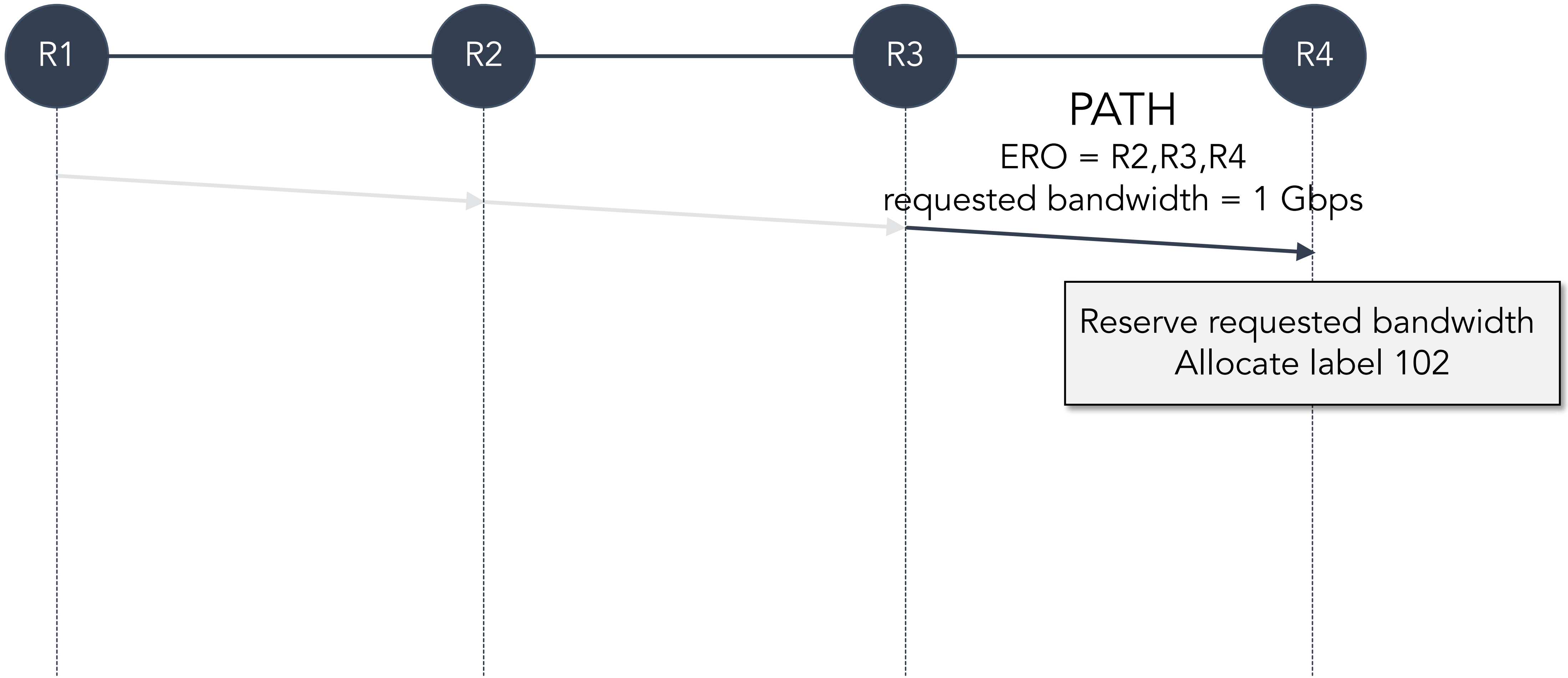
RSVP path signaling



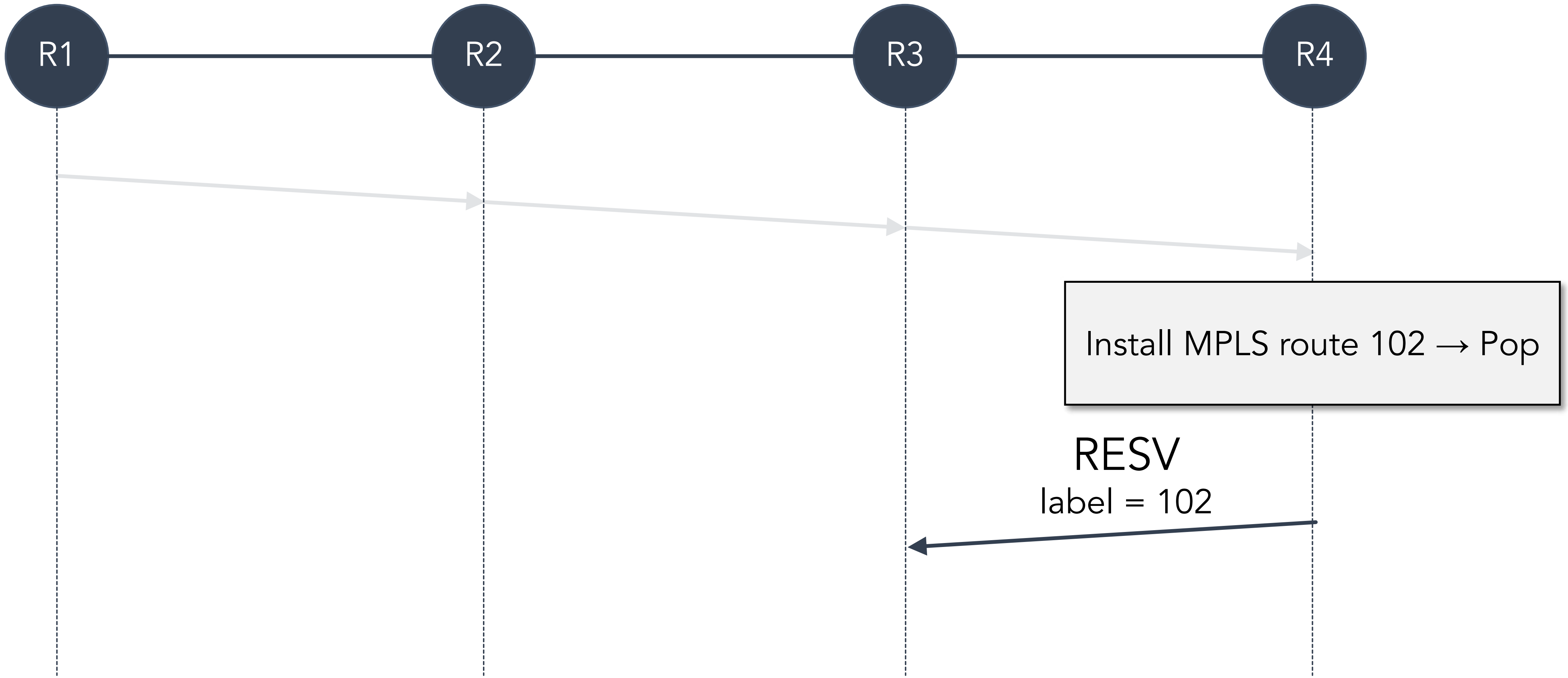
RSVP path signaling



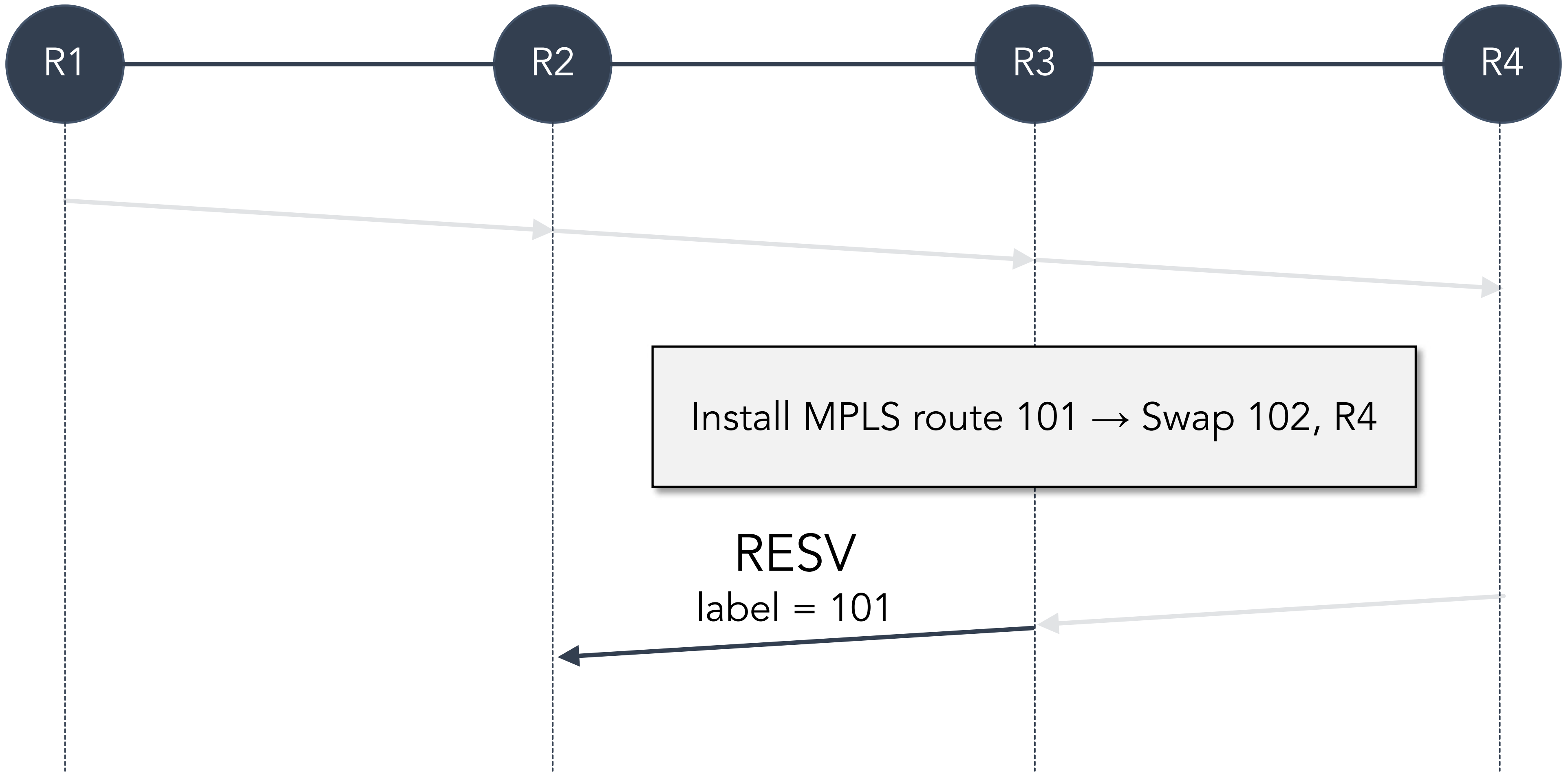
RSVP path signaling



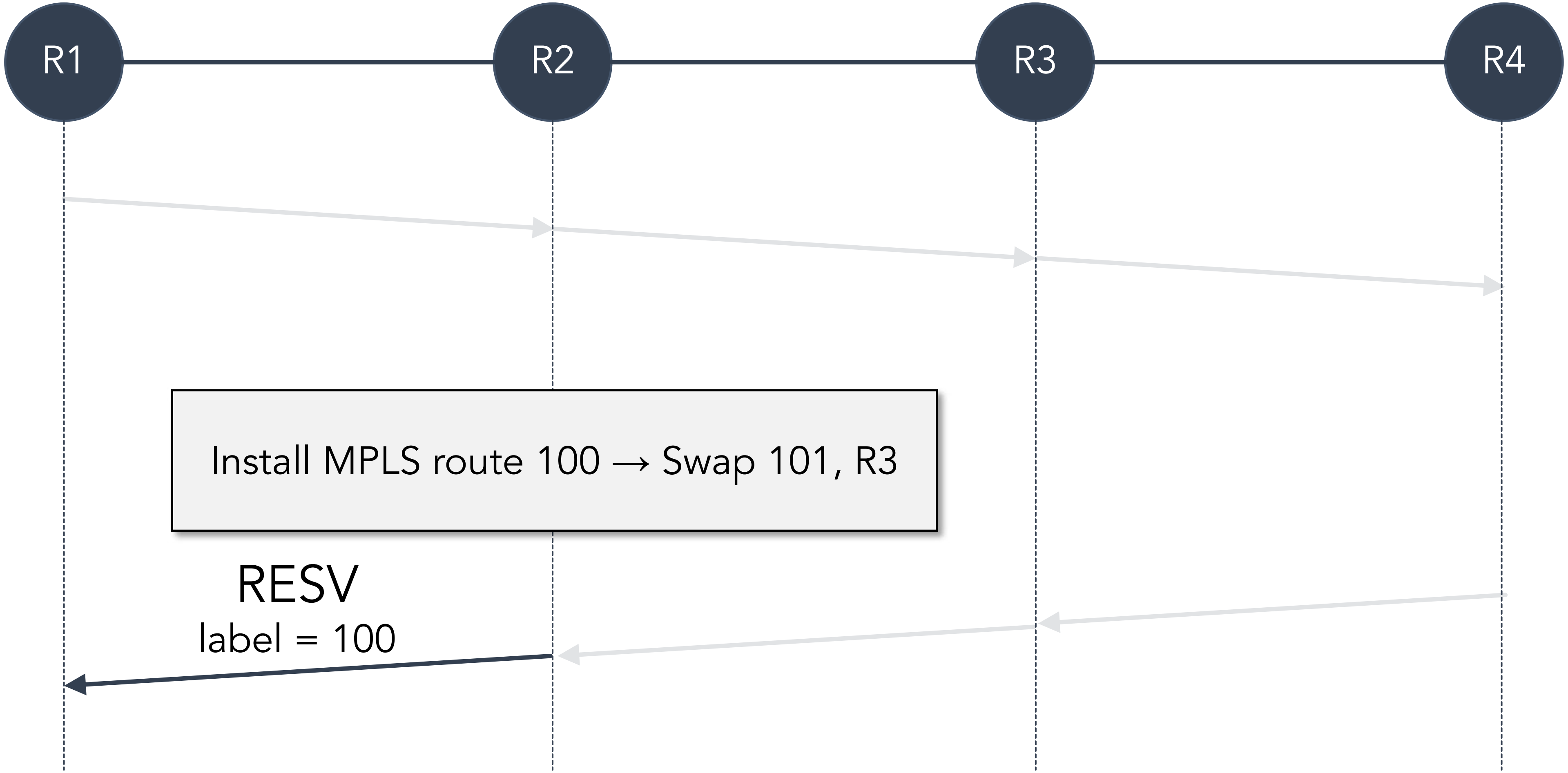
RSVP path signaling



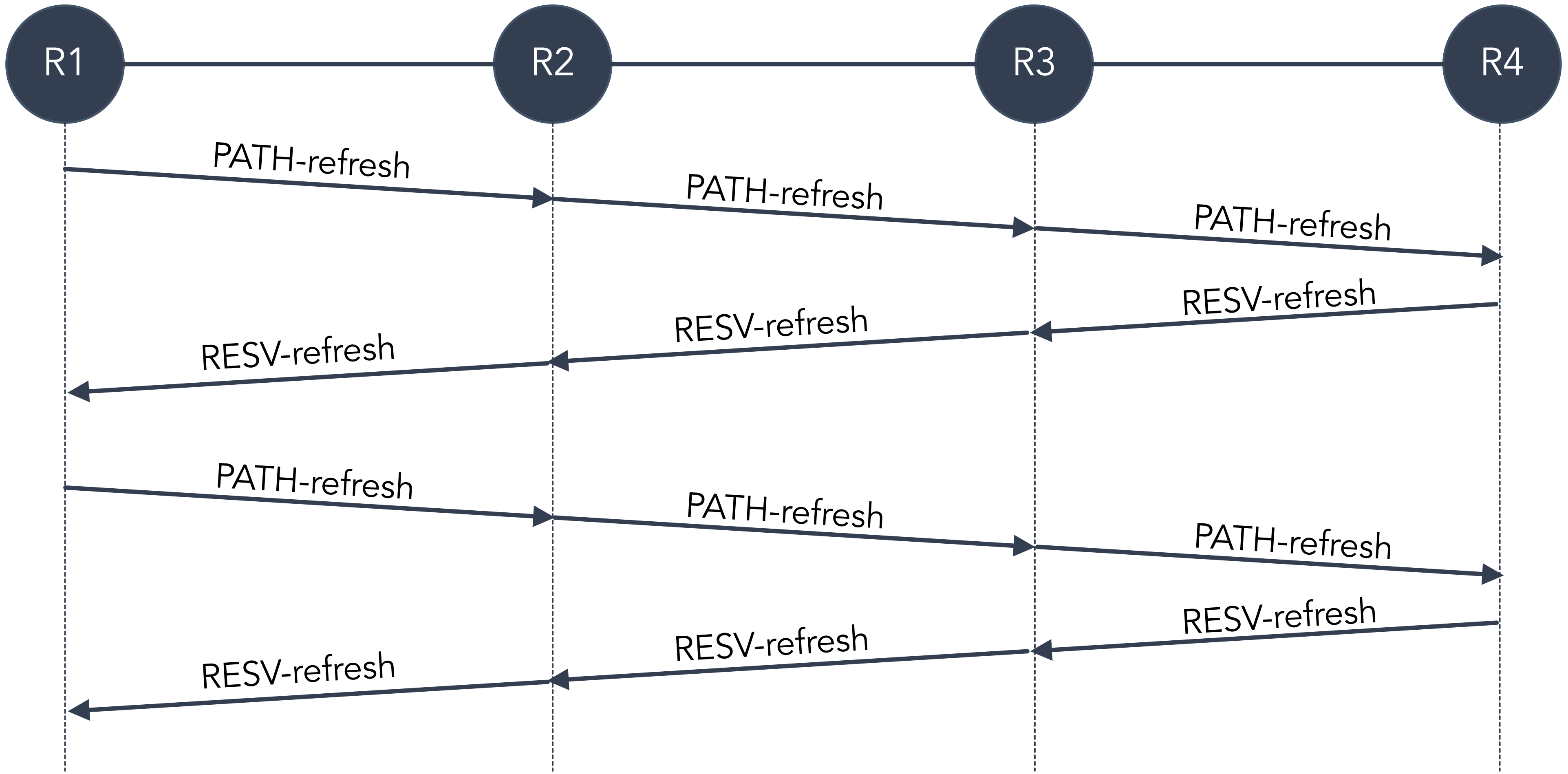
RSVP path signaling



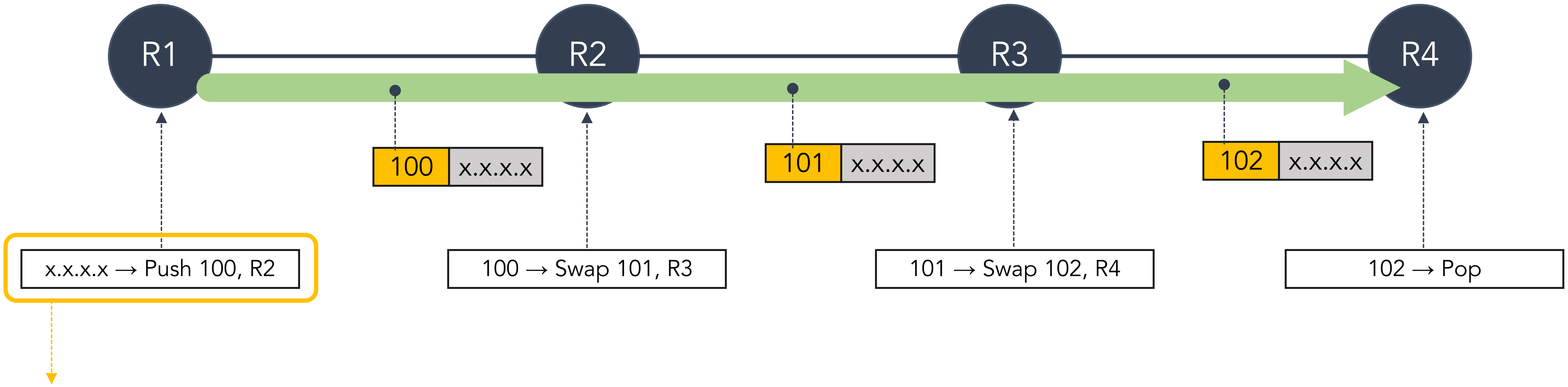
RSVP path signaling



RSVP is soft state (needs periodic refresh)



LSP installed by LSP signalling



This route is not installed by RSVP, but by some protocol that decides to "use" the LSP

Some notes on bandwidth reservations

- Reservation of the bandwidth is only a matter of control-plan “book keeping”.
- RSVP itself typically does not install any rate limiters or shapers in the data plane
- RSVP only makes sure the traffic will fit if the demand actually matches the reservations
- Typically “other mechanisms” are used to rate limit or shape the traffic that goes into LSPs
- Or, alternatively, ”auto-bandwidth” is used to grow or shrink the LSPs according to the actual demand

Challenges with RSVP

- Traffic engineer full mesh of LSPs between edge routers
  Based on traffic demand matrix
- N edge routers require N^2 LSPs
  100 edge routers means 10,000 LSPs
  Core routers may have 10,000 transit LSPs
- High RSVP signaling load in control plane
  Initial signaling
  Refresh signaling
- Large amount of state in forwarding plane
  N^2 MPLS routes

Possible solutions to RSVP challenges

- Reduce scope of traffic engineering
  Reduce N in N^2
  Only traffic engineer the core of the core
  Run LDP-over-RSVP edge-to-edge
- RSVP refresh reduction
  Various mechanisms to reduce the amount refresh messages
  Personally I liked CR-LDP which is hard-state (lost the race due to non-technical reasons)
- Use Segment Routing (SR)
  Less state in control and data plane: label per neighbor instead of label per path
  More state in packet (introduces problems of its own due to hardware limitations)
  No signaling of LSPs
  Puts much more control in the hands of the applications

Signaling lessons for quantum networks

- We already decided that a centralized path computation element makes sense for quantum networks.
- In that case, we can initially do without a router-to-router signaling protocol such as RSVP.
- Instead, we can use a controller-to-router protocol (e.g. a generalization of PCEP or P4 APIs)
- We can always add a generalization of RSVP or SR later

Summary of lessons for quantum networks

- Traffic engineering makes sense for quantum networks
- Existing classical protocols can be re-used and/or generalized for quantum networks:
  - **Topology discovery**: use OSPF or ISIS multi-topology
  - **Resource discovery**: generalize TE extensions for OSPF or ISIS and introduce new types of resources *(already work-in-progress)*
  - **Path computation**: use centralized path computation engine and develop new path computation algorithms that are aware of the specifics of quantum repeaters *(just started an open source project on GitHub)*
  - **Path signaling**: at first use SDN approach that eliminates the need for a router-to-router signaling protocol. Later potentially generalize RSVP-TE or SR.
  - **Forwarding plane**: use existing mechanisms (e.g. P4) for classical messages, develop new (stateful) primitives for quantum qubits


Thank you!