Traffic Engineering for the Modern MPLS Backbone

Extending PCEP for Stateful Control of MPLS RSVP-TE Attributes

Edward Crabbe, Google
Jan Medved, Juniper
Robert Varga, Juniper

"...it is generally desirable to ensure that subsets of network resources do not become over utilized and congested while other subsets along alternate feasible paths remain underutilized. Bandwidth is a crucial resource in contemporary networks. Therefore, a central function of Traffic Engineering is to efficiently manage bandwidth resources.

Minimizing congestion is a primary traffic and resource oriented performance objective. The interest here is on congestion problems that are prolonged rather than on transient congestion resulting from instantaneous bursts. Congestion typically manifests under two scenarios:

1. When network resources are insufficient or inadequate to accommodate offered load.

2. When traffic streams are inefficiently mapped onto available resources; causing subsets of network resources to become over-utilized while others remain underutilized."

[RFC 2702 Requirements for Traffic Engineering Over MPLS]

"...it is generally desirable to ensure that subsets of network resources do not become over utilized and congested while other subsets along alternate feasible paths remain underutilized. Bandwidth is a crucial resource in contemporary networks. Therefore, a central function of Traffic Engineering is to efficiently manage bandwidth resources.

Minimizing congestion is a primary traffic and resource oriented performance objective. The interest here is on congestion problems that are prolonged rather than on transient congestion resulting from instantaneous bursts. Congestion typically manifests under two scenarios:

1. When network resources are insufficient or inadequate to accommodate offered load.

2. When traffic streams are inefficiently mapped onto available resources; causing subsets of network resources to become over-utilized while others remain underutilized."

[RFC 2702 Requirements for Traffic Engineering Over MPLS]

Topics

- MPLS TE Today
- Example Use Cases
- Stateful PCE Protocol Proposal

Topics

- **MPLS TE Today**
- Example Use Cases
- Stateful PCE Protocol Proposal

Online MPLS Control Mechanisms

Online: computation and control of RSVP-TE parameters by local network element

- **auto bandwidth**
  - defacto 'standard' in online control mechnisms
  - not widely deployed (but the few networks that do use it are quite large)
  - provides independent, asynchronous control of device-local LSPs
    - unsynchronized, fixed rate per device timers
    - local empirical measurement of demands with little hysteresis

Offline MPLS Control Mechanisms

Offline: computation by system outside network element and control of RSVP-TE parameters via northbound API

- **Config**
- **PCE**
- **Openflow**

Offline MPLS Control Mechanisms

Offline: computation by system outside network element and control of RSVP-TE parameters via northbound API

- **Config**
  - simplest' offline control method
  - relies on heavyweight config database changes for updates a bit heavy weight for transient forwarding state
  - may lock config sections for duration of changes potentially problematic
  - platform dependent interface
  - Does not trigger RFC3209 section 2.5 reroute behavior on most platforms
  - config length effects compilation time (∴ boot time)

  Why not just do everything with static routes? :P

- **PCE**
- **Openflow**

Offline MPLS Control Mechanisms

Offline: computation by system outside network element and control of RSVP-TE parameters via northbound API

- **Config**
- **PCE**
  - No way to control timing of updates (without inefficiency in control plane as a result of directionality)
  - No way to control sequence of updates across devices
  - no way to collect results of PCEP PCRep messages (RSVP-TE error codes)
  - no way to collect LSP state from devices 'in-band'
- **Openflow**

Topics

- MPLS TE Today
- **Example Use Cases**
- Stateful PCE Protocol Proposal

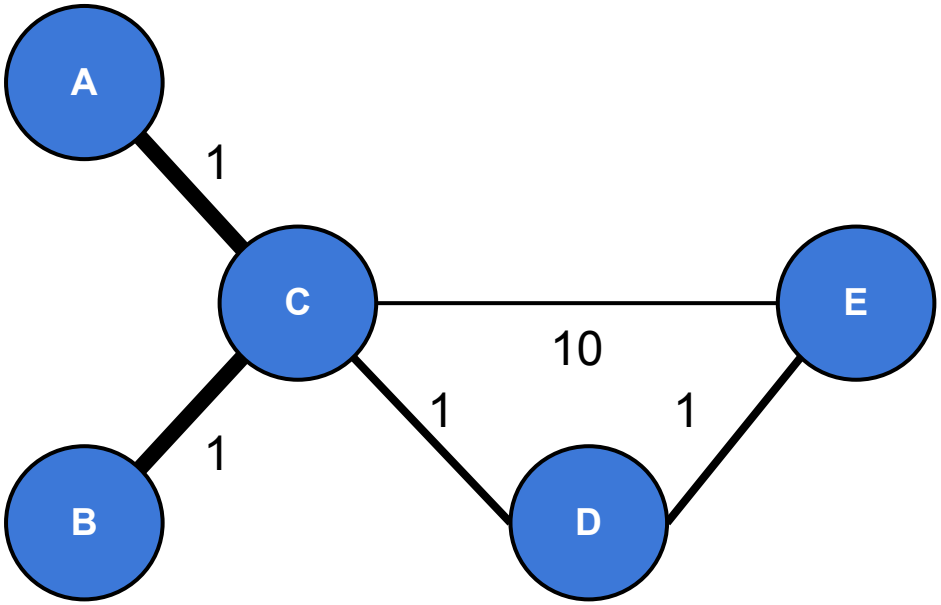
Example Stateful PCEP Use Cases

- Deadlock Resolution
- Bin Packing
- Scheduling / Calendaring
- Predictability
- Adaptive Timescales
- Constraint Relaxation
- GCO
- ⋮

Example Stateful PCEP Use Cases

- Deadlock Resolution
- Bin Packing
- Scheduling / Calendaring
- Predictability
- Adaptive Timescales
- Constraint Relaxation
- GCO
- ⋮

Deadlock



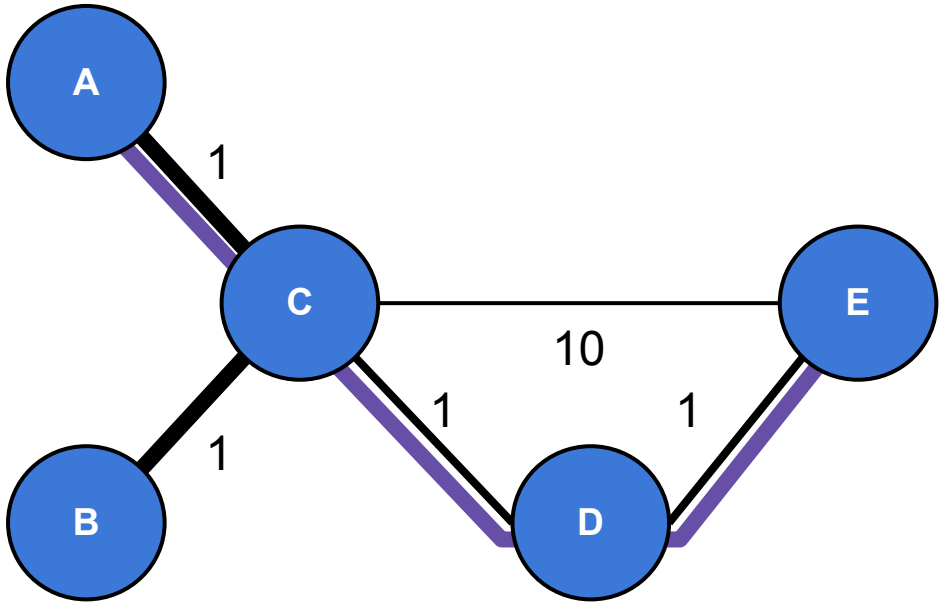
causes:

- control / dataplane decoupling
- rfc3209 implies no teardown on reservation increase failure
  - demand will be miss signaled for long periods
- lack of global LSP state
- lack of LSP level ingress admission control
  - would require another online or offline control mechanism
  - tension between overprovisioning level and transport elasticity

Link	Metric	Capacity
A-C	1	20
B-C	1	20
C-E	10	5
C-D	1	10
D-E	1	10

Time	LSP	Src	Dst	Demand
1	1	A	E	2
2	2	B	E	2
3	1	A	E	20

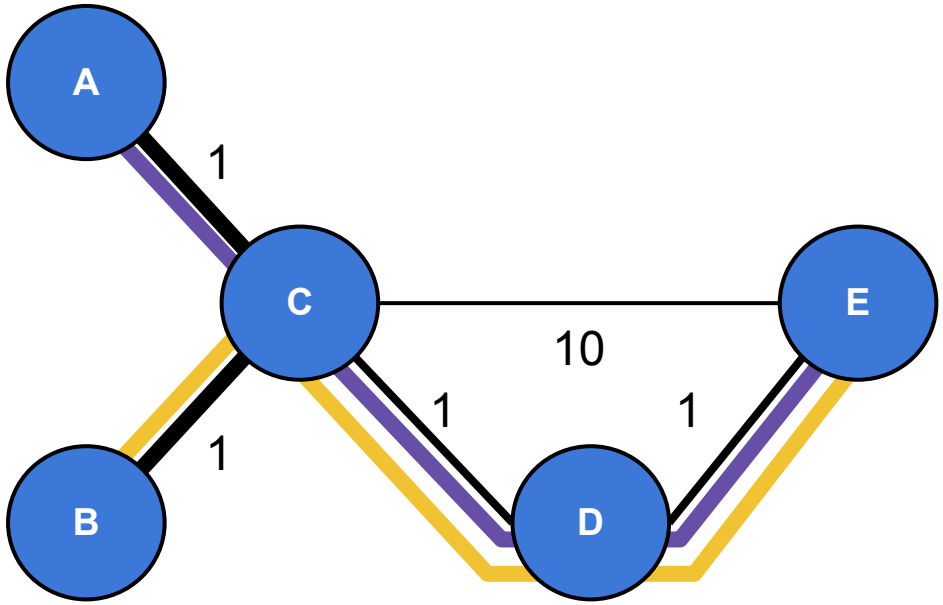
Deadlock



Link	Metric	Capacity
A-C	1	20
B-C	1	20
C-E	10	5
C-D	1	10
D-E	1	10

Time	LSP	Src	Dst	Demand
1	1	A	E	2
2	2	B	E	2
3	1	A	E	20

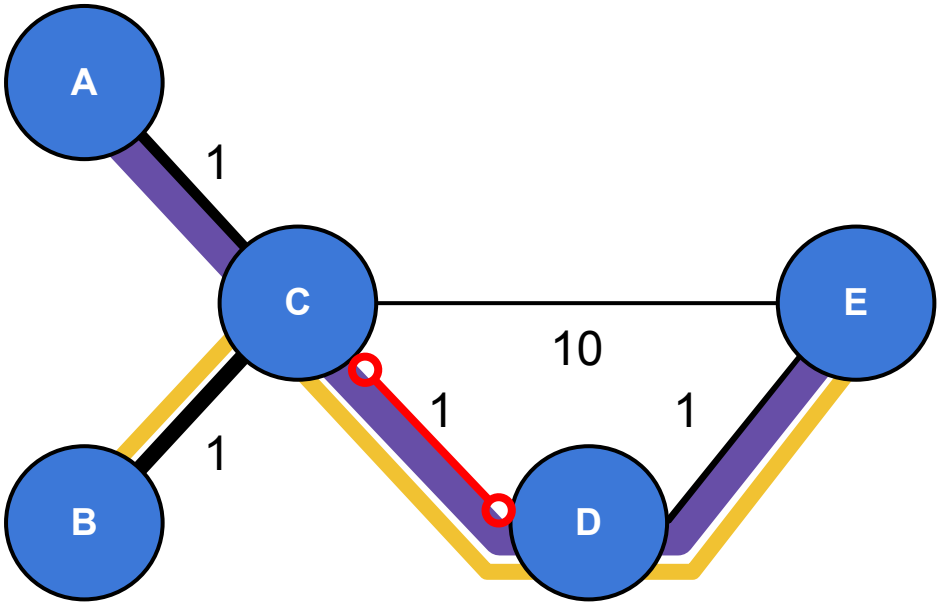
Deadlock



Link	Metric	Capacity
A-C	1	20
B-C	1	20
C-E	10	5
C-D	1	10
D-E	1	10

Time	LSP	Src	Dst	Demand
1	1	A	E	2
2	2	B	E	2
3	1	A	E	20

Deadlock

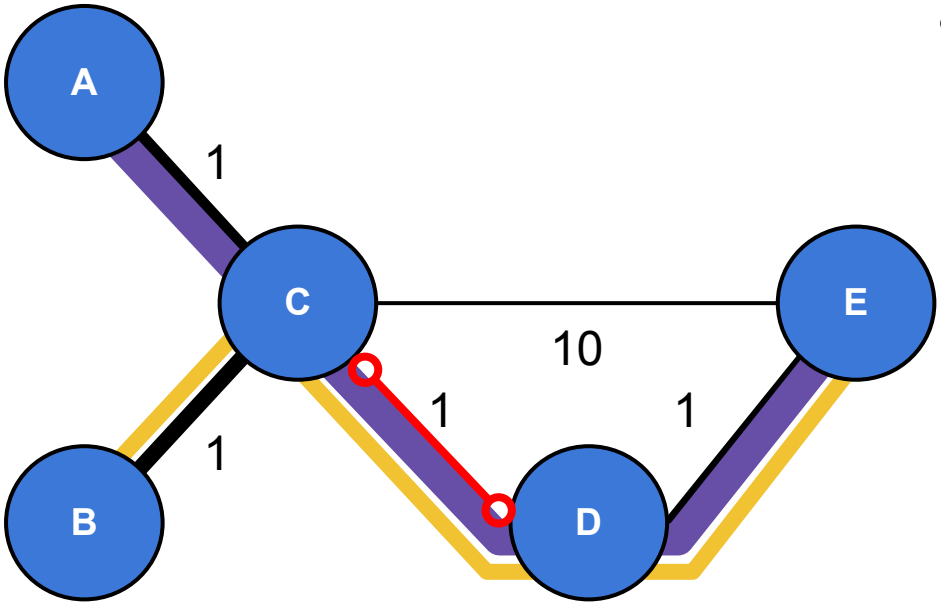


- LSP 1:
  - demand cannot be satisfied
  - LSP not torn down due to 3209
  - usage controlled due to control/data plane decoupling
  - ⇒ information in IGP, RSVP is inaccurate
- LSP 2
  - lack of visibility w/r/t LSP 1 misbehavior results in unecessary, potentially prolongued degradation in service
  - could be rerouted along C-E link modulo flow performance constraints

Link	Metric	Capacity
A-C	1	20
B-C	1	20
C-E	10	5
C-D	1	10
D-E	1	10

Time	LSP	Src	Dst	Demand
1	1	A	E	2
2	2	B	E	2
3	1	A	E	20

Deadlock

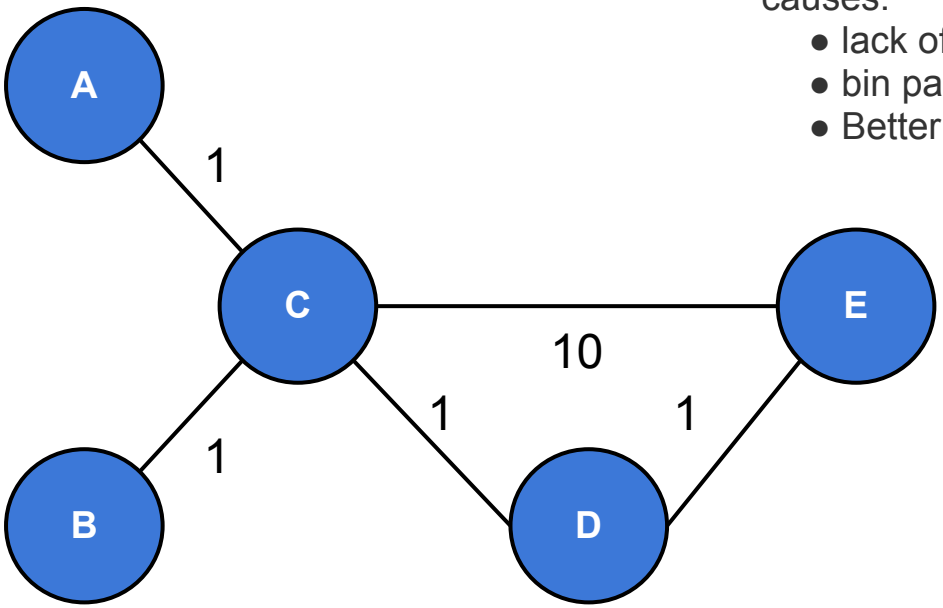


- lack of LSP level ingress admission control
  - would require another online or offline control mechanism
    - offline: need northbound API
    - online: back to autopbw issues
  - tension between overprovisioning level and transport elasticity

Link	Metric	Capacity
A-C	1	20
B-C	1	20
C-E	10	5
C-D	1	10
D-E	1	10

Time	LSP	Src	Dst	Demand
1	1	A	E	2
2	2	B	E	2
3	1	A	E	20

Bin Packing



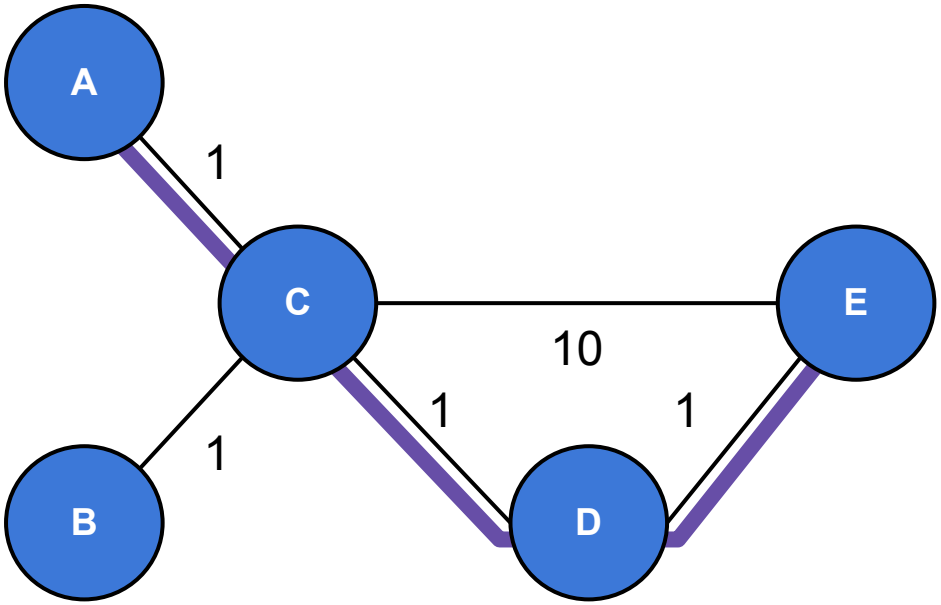
causes:

- lack of global LSP state
- bin packing is a sequencing problem - NP-Hard
- Better to solve w/ throughput optimization

Link	Metric	Capacity
A-C	1	10
B-C	1	10
C-E	10	5
C-D	1	10
D-E	1	10

Time	LSP	Src	Dst	Demand
1	1	A	E	5
2	2	B	E	10

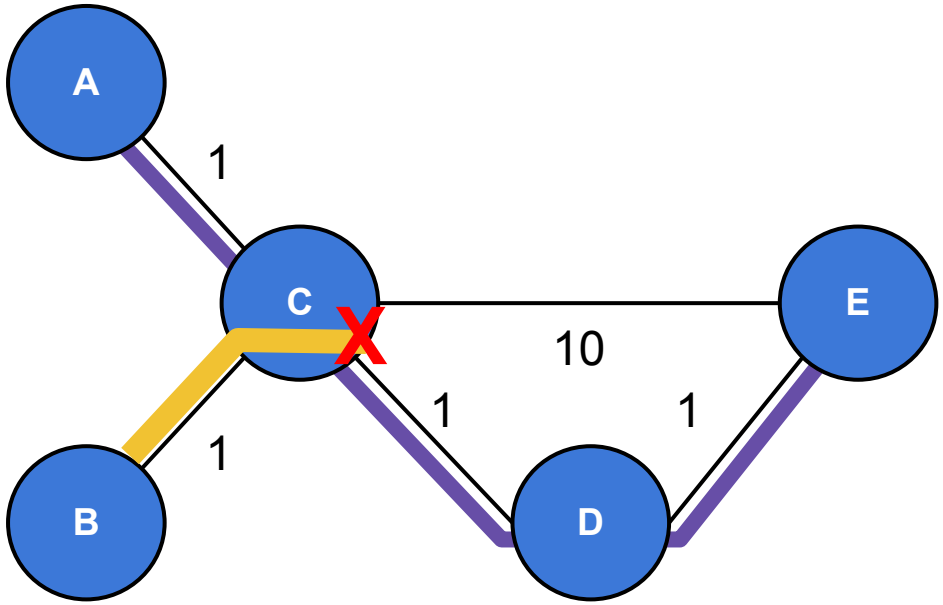
Bin Packing



Link	Metric	Capacity
A-C	1	10
B-C	1	10
C-E	10	5
C-D	1	10
D-E	1	10

Time	LSP	Src	Dst	Demand
1	1	A	E	5
2	2	B	E	10

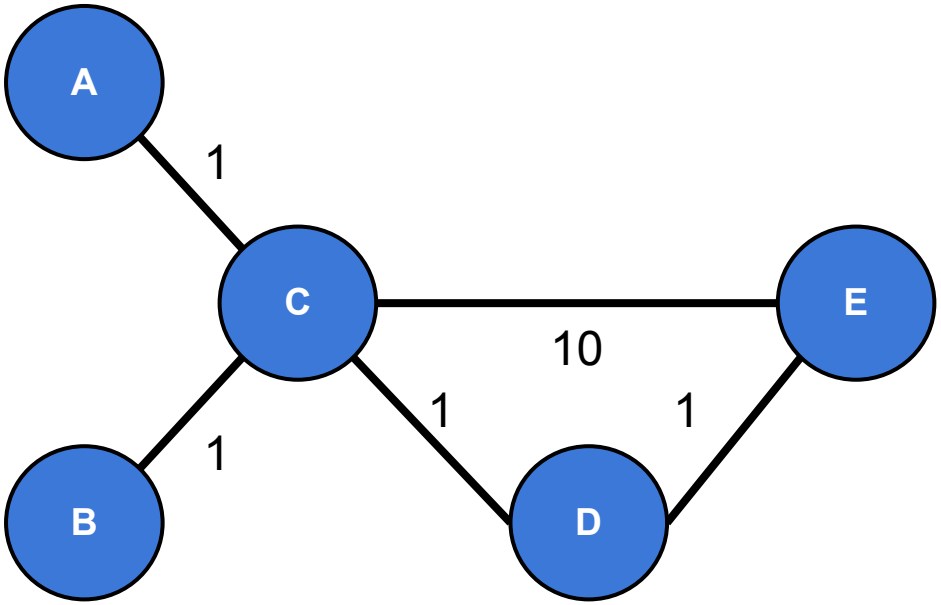
Bin Packing



Link	Metric	Capacity
A-C	1	10
B-C	1	10
C-E	10	5
C-D	1	10
D-E	1	10

Time	LSP	Src	Dst	Demand
1	1	A	E	5
2	2	B	E	10

Scheduling



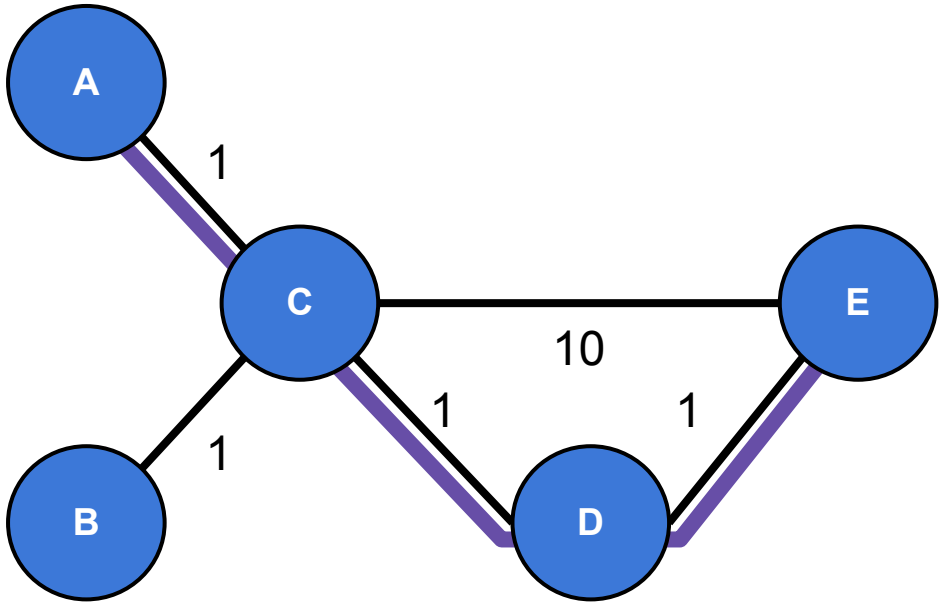
causes:

- autobw empirically derives demand with single period hysteresis

Link	Metric	Capacity
A-C	1	10
B-C	1	10
C-E	10	10
C-D	1	10
D-E	1	10

Time	LSP	Src	Dst	Demand
1	1	A	E	2
2	2	B	E	7
3	1	A	E	7

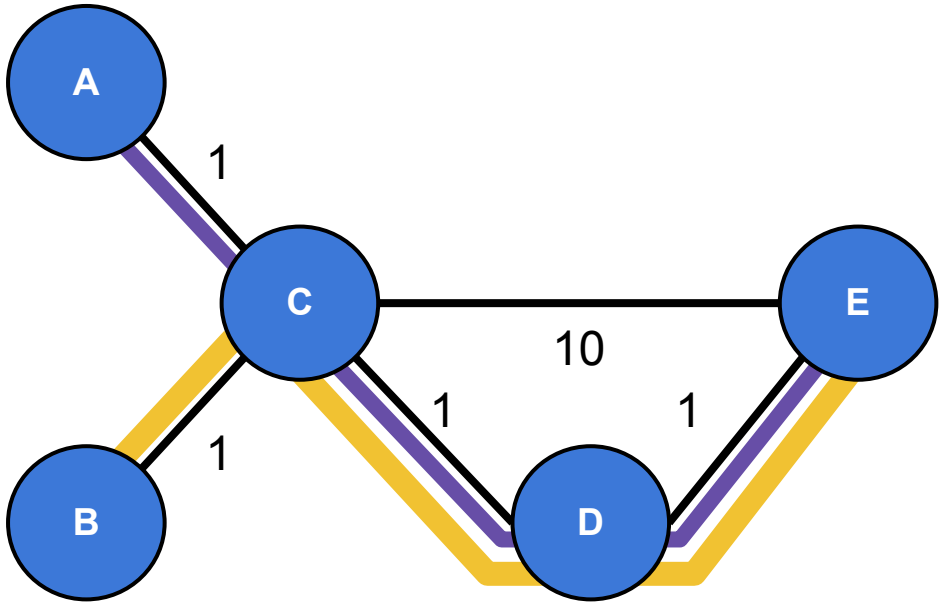
Scheduling



Link	Metric	Capacity
A-C	1	10
B-C	1	10
C-E	10	10
C-D	1	10
D-E	1	10

Time	LSP	Src	Dst	Demand
1	1	A	E	2
2	2	B	E	7
3	1	A	E	7

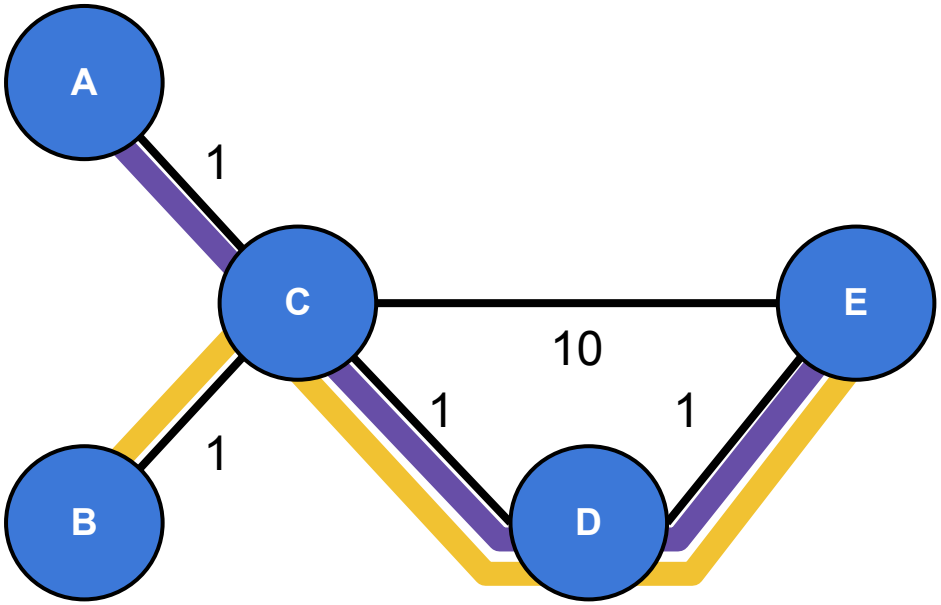
Scheduling



Link	Metric	Capacity
A-C	1	10
B-C	1	10
C-E	10	10
C-D	1	10
D-E	1	10

Time	LSP	Src	Dst	Demand
1	1	A	E	2
2	2	B	E	7
3	1	A	E	7

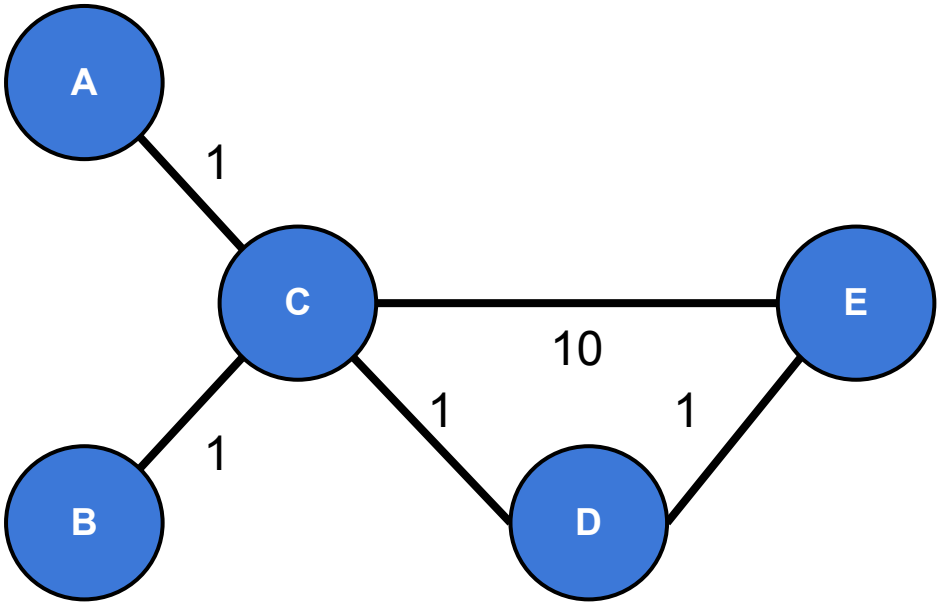
Scheduling



Link	Metric	Capacity
A-C	1	10
B-C	1	10
C-E	10	10
C-D	1	10
D-E	1	10

Time	LSP	Src	Dst	Demand
1	1	A	E	2
2	2	B	E	7
3	1	A	E	7

Predictability



causes:

- routers act independently and asynchronously ⇒ path dictated by order of event arrival

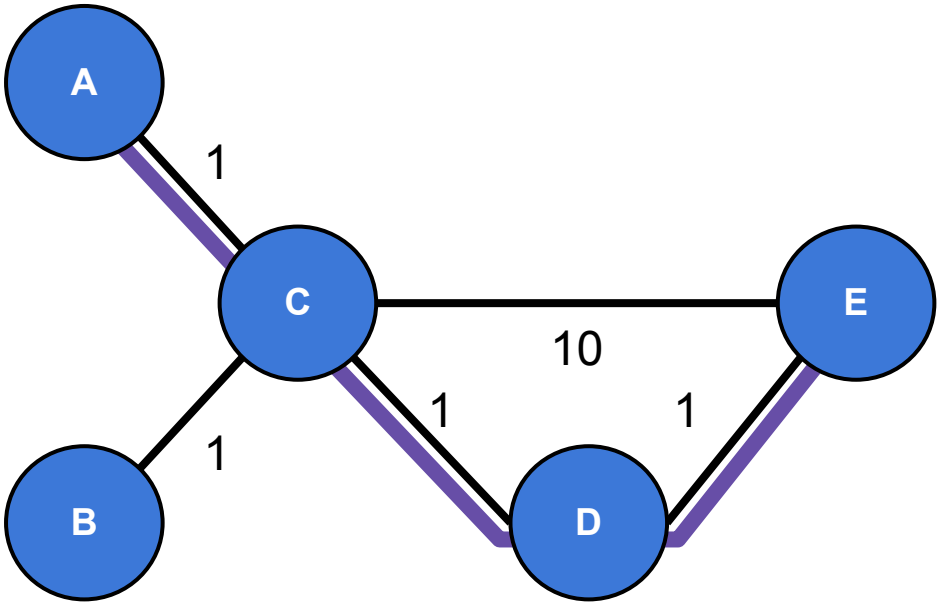
Link	Metric	Capacity
A-C	1	10
B-C	1	10
C-E	1	10
C-D	1	10
D-E	1	10

Time	LSP	Src	Dst	Demand
1	1	A	E	7
2	2	B	E	7

VS

Time	LSP	Src	Dst	Demand
1	2	B	E	7
2	1	A	E	7

Predictability



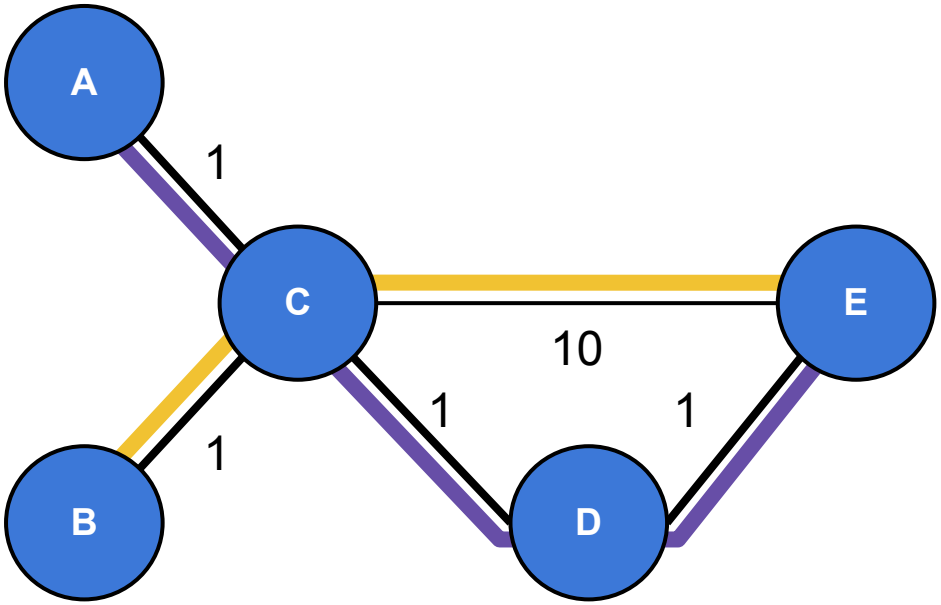
Link	Metric	Capacity
A-C	1	10
B-C	1	10
C-E	1	10
C-D	1	10
D-E	1	10

Time	LSP	Src	Dst	Demand
1	1	A	E	7
2	2	B	E	7

VS

Time	LSP	Src	Dst	Demand
1	2	B	E	7
2	1	A	E	7

Predictability



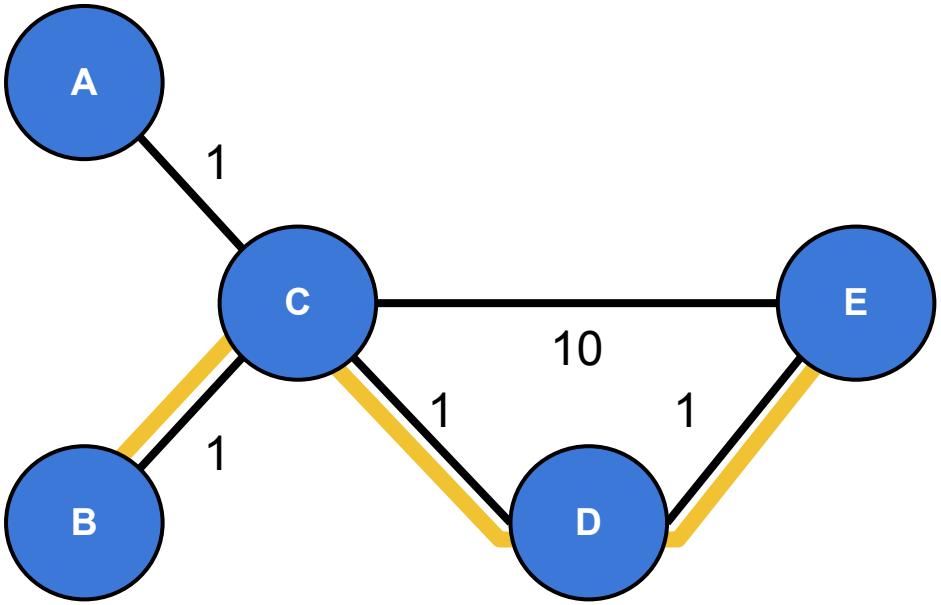
Time	LSP	Src	Dst	Demand
1	1	A	E	7
2	2	B	E	7

Link	Metric	Capacity
A-C	1	10
B-C	1	10
C-E	1	10
C-D	1	10
D-E	1	10

VS

Time	LSP	Src	Dst	Demand
1	2	B	E	7
2	1	A	E	7

Predictability



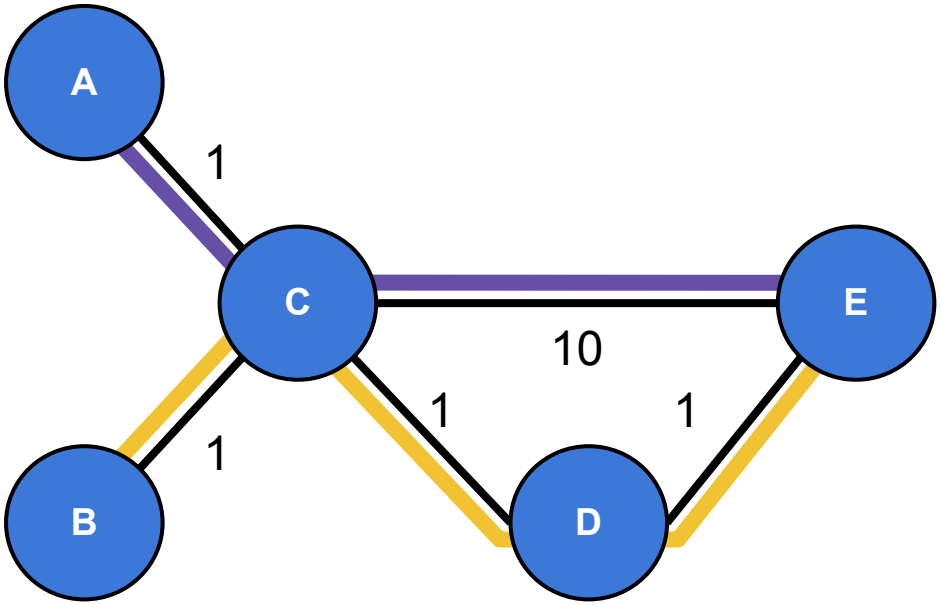
Link	Metric	Capacity
A-C	1	10
B-C	1	10
C-E	1	10
C-D	1	10
D-E	1	10

Time	LSP	Src	Dst	Demand
1	1	A	E	7
2	2	B	E	7

VS

Time	LSP	Src	Dst	Demand
1	2	B	E	7
2	1	A	E	7

Predictability



Link	Metric	Capacity
A-C	1	10
B-C	1	10
C-E	1	10
C-D	1	10
D-E	1	10

Time	LSP	Src	Dst	Demand
1	1	A	E	7
2	2	B	E	7

VS

Time	LSP	Src	Dst	Demand
1	2	B	E	7
2	1	A	E	7

Topics

- MPLS TE Today
- Example Use Cases
- **Stateful PCE Protocol Proposal**

Stateful PCE (RSVP-TE focused)

Will Provide:

- Efficient Distribution of LSP State to PCE
  - Global view of network state to LSP level
  - minimal distributed state - much more efficient than distributed protocol for accomplishing similar objectives
- Delegation of control of LSP attributes to PCE
  - Support for redundant PCEs with near-hitless failover
- PCE driven control of LSP attribute changes
  - Ordering and synchronization of attribute changes across devices
  - opens the way for all manner of optimization

Stateful PCE for RSVP-TE Objectives

- Allow a single PCC to interact with a mix of stateless and stateful PCEs simultaneously using the same PCEP.
- Allow stateful PCEs to learn about the state of LSPs in a PCC
- Allow a PCC to delegate control of its LSPs to an active stateful PCE on a per LSP basis.
- Allow PCE to control of computation timing and sequence across all LSPs that have been delegated to it.
- Enable uninterrupted operation of PCC's LSPs in the event PCE failure or while control of LSPs is being transferred between PCEs.

PCEP Modifications & Additions

Capability Additions

Function	Direction	Message & RFC Updated
Capability negotiation	E<->C	PCRpt [RFC5440]
ISIS stateful capability advertisement	E->C	ISIS PCE-CAP-FLAGS sub-TLV [RFC 5089]
OSPF stateful capability advertisement	E->C	OSPF RI LSA, PCE TLV, PCE-CAP-FLAGS sub-TLV [RFC 5088]

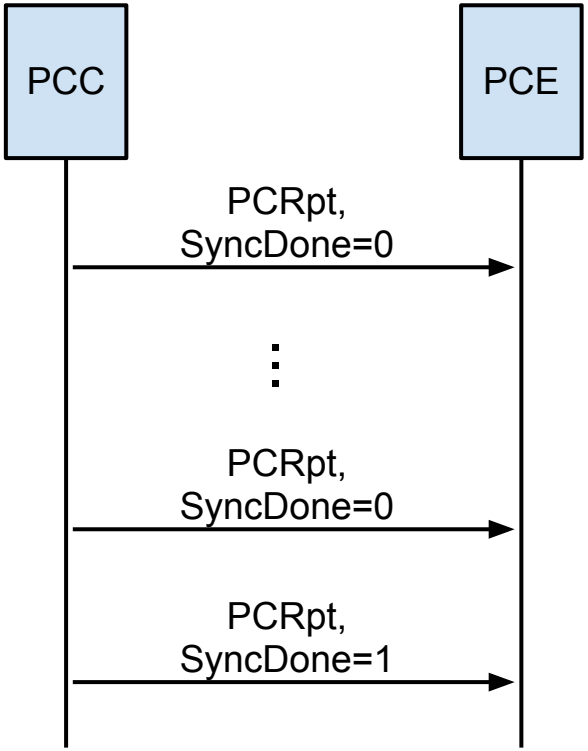
New Functions

Function	Direction	PCEP Message
LSP state synchronization	C->E	PCRpt
LSP status request	E->C	PCSrq
LSP status report/notification	C->E	PCRpt
LSP control delegation	C<->E	PCRpt
LSP modification	E->C	PCUpd

State Synchronization

Uses new Messages: PCRpt

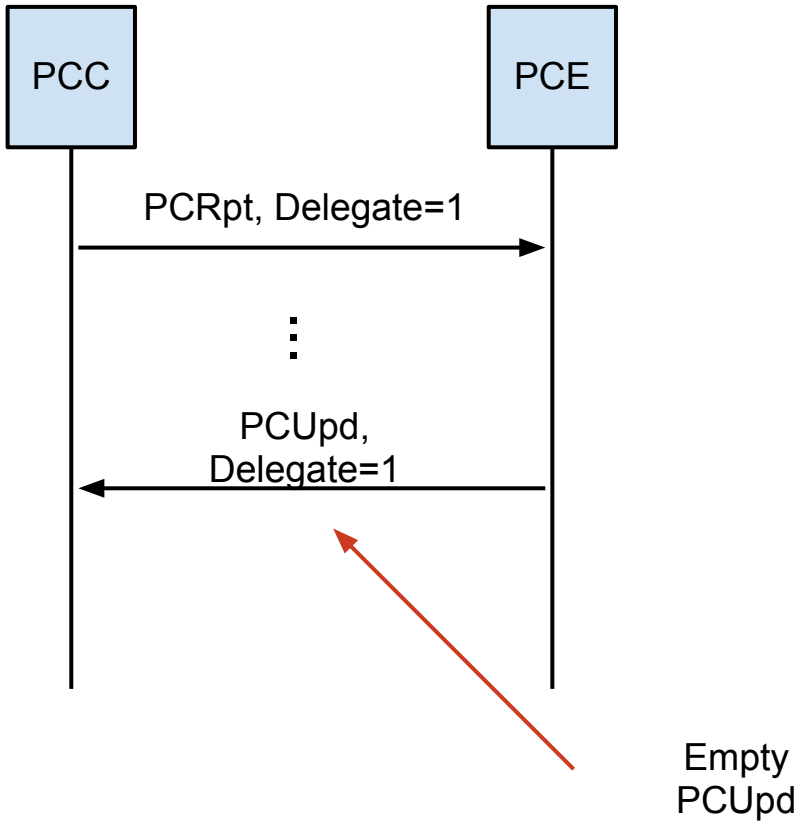
New object class and type: LSP-Object



LSP Delegation

Uses new Messages: PCRpt, PCUpd

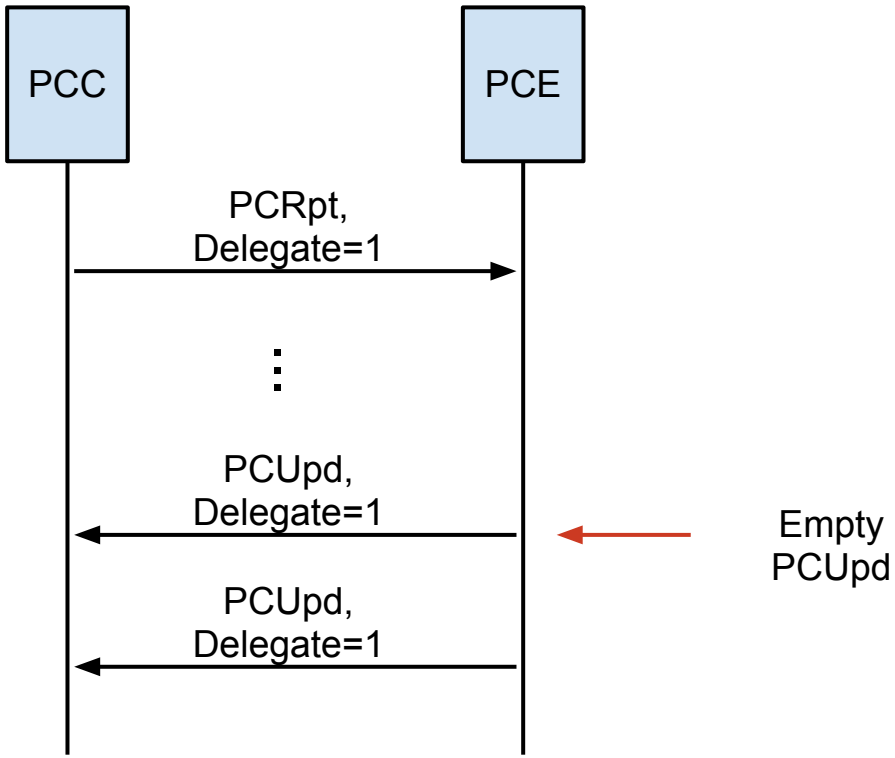
New object class and type: LSP-Object



PCE Driven Control

Uses new Messages: PCRpt, PCUpd

New object class and type: LSP-Object



Next Steps

- Discussion...
- WG Adoption?