SDN Workshop

Contact: training@apnic.net

Issue Date: [Date]

Revision: [xx]

APNIC

Path Computation Element

SDN Workshop

APNIC

WSDN01_v0.1

Issue Date: [Date]

Revision: [xx]

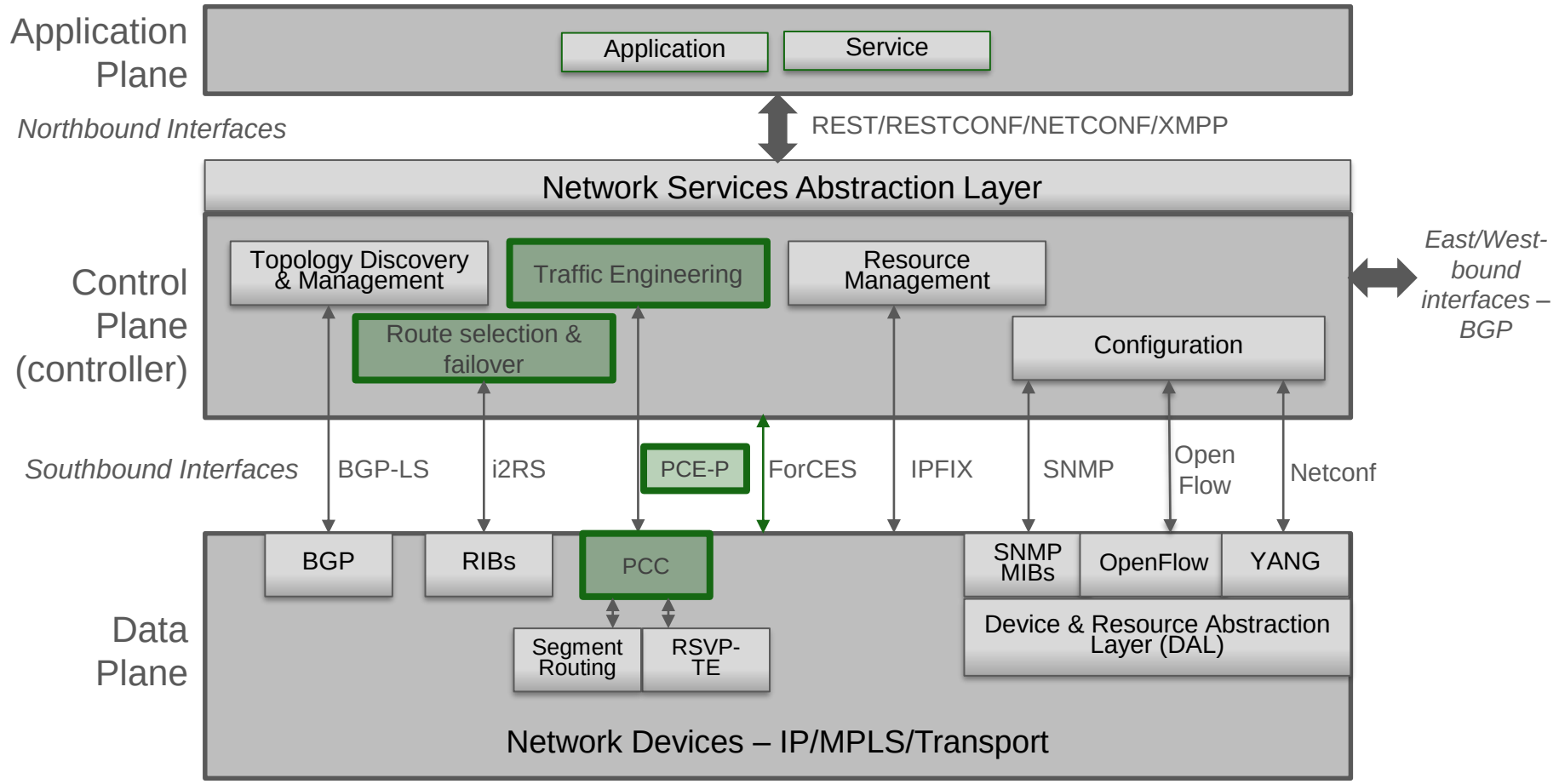
Overview

- In a nutshell
- Introduction
- Motivations
- PCE architecture
- Stateful PCE
- PCE-initiated LSPs
- PCEP
- PCEP extensions for stateful PCEs
- PCEP extensions for PCE-initiated LSPs
- SR extensions
- Operational considerations
- References

In a nutshell

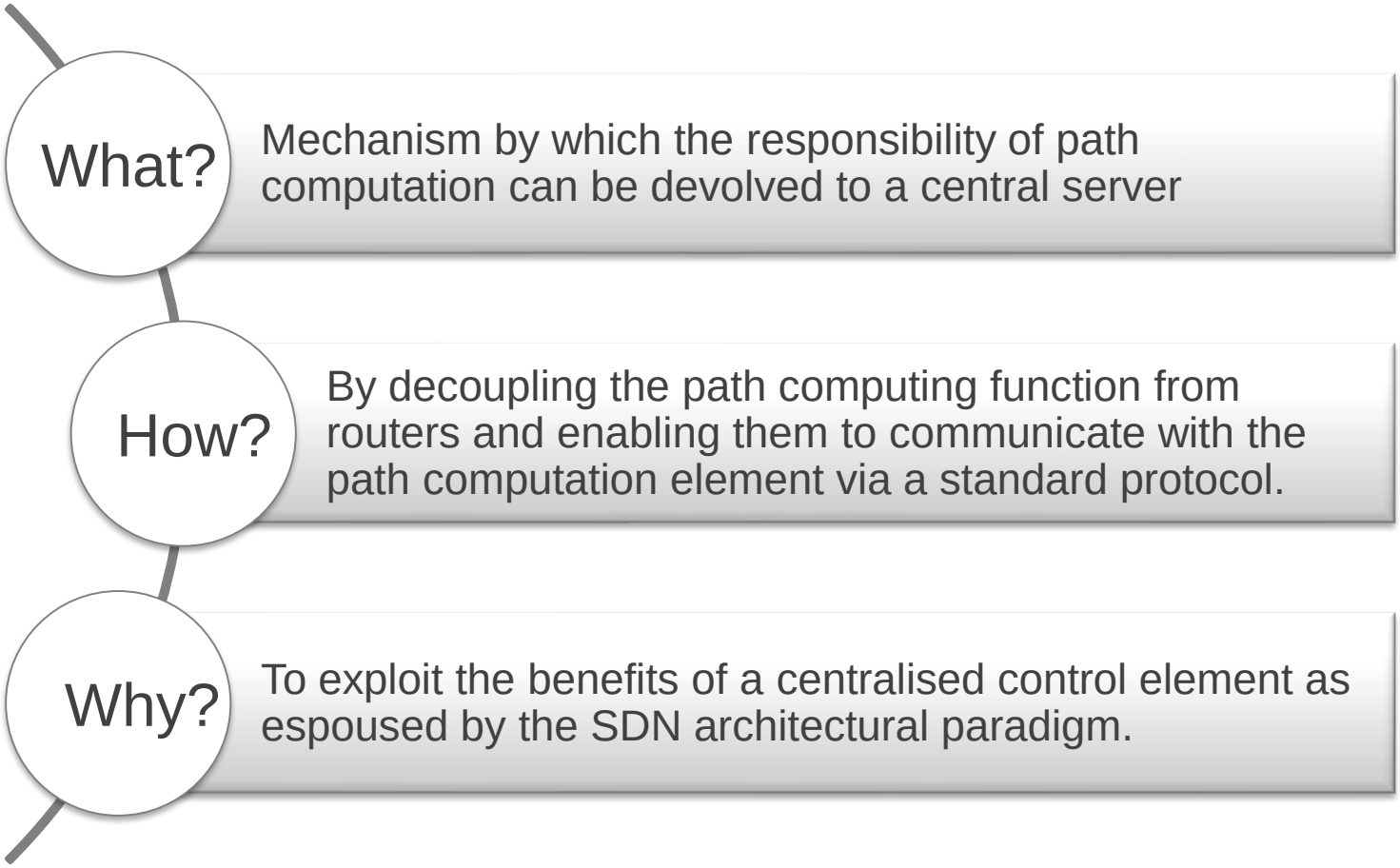
APNIC

SDN architectural framework



Note: designations of north-bound and south-bound are relative to the control plane ("controller")

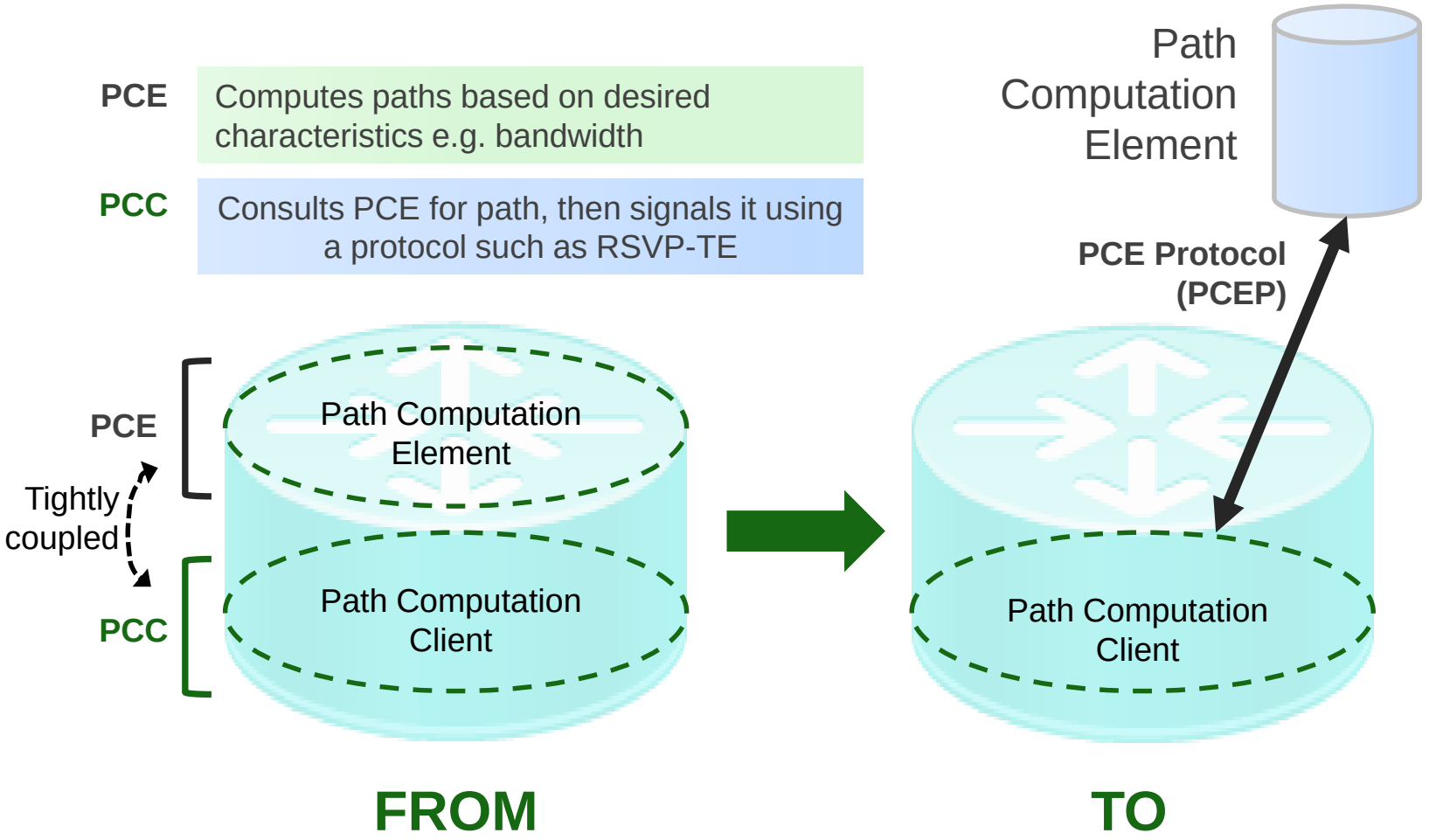
PCE: in a nutshell…



The PCE split

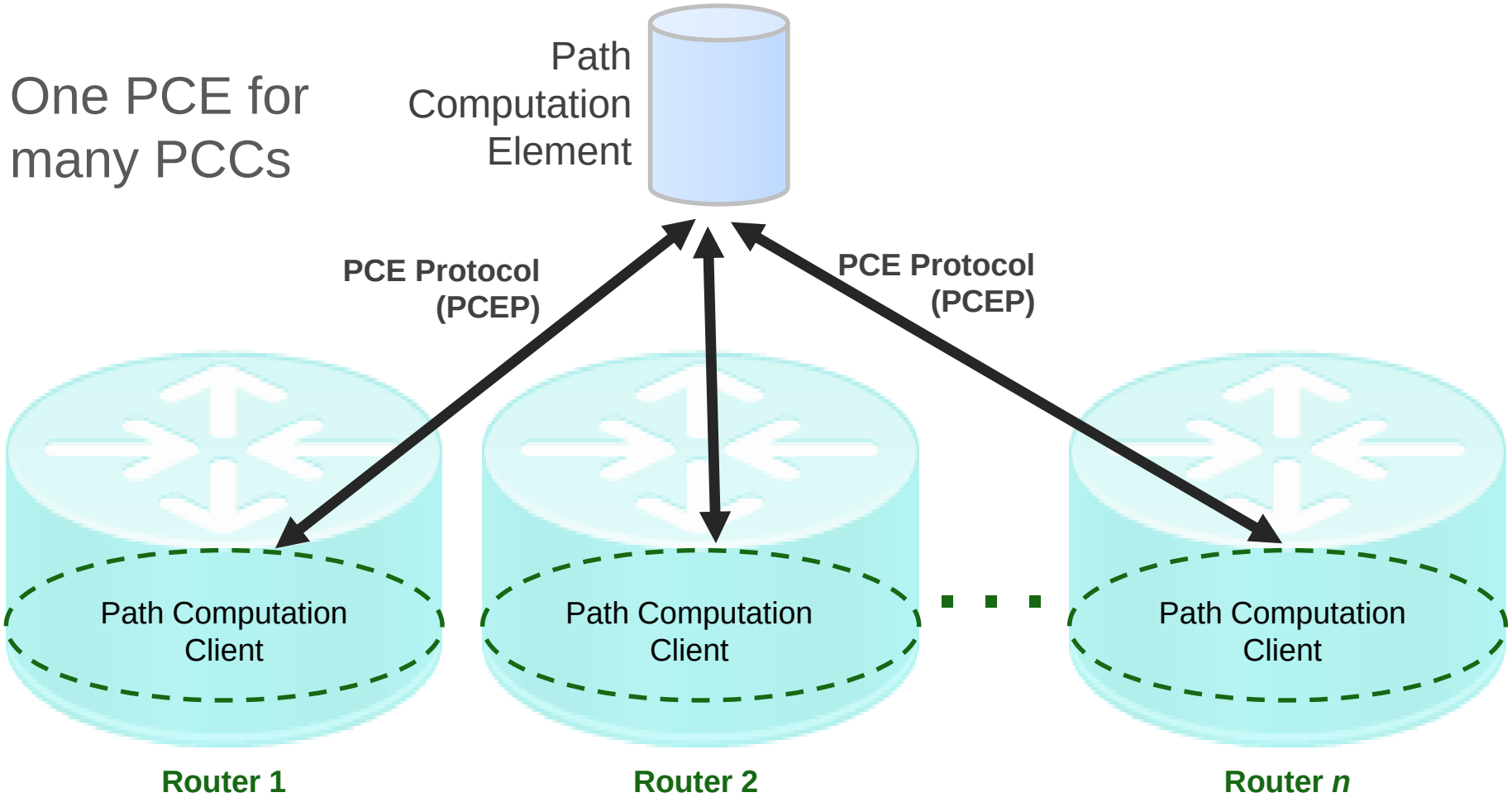
PCE Computes paths based on desired characteristics e.g. bandwidth

PCC Consults PCE for path, then signals it using a protocol such as RSVP-TE



Centralised path computation

One PCE for many PCCs



Introduction

MPLS traffic-engineering objectives

Key objectives of MPLS TE

Elements of MPLS-TE solution

Routing
IGP traffic-engineering extensions

Path Computation
Placement of LSP by the head-end LSR based on CSPF algorithm

Signaling
Establishes the LSP by signaling it along the computed path

MPLS TE is concerned with explicitly routing MPLS LSPs whose paths satisfy a set of traffic-engineering constraints

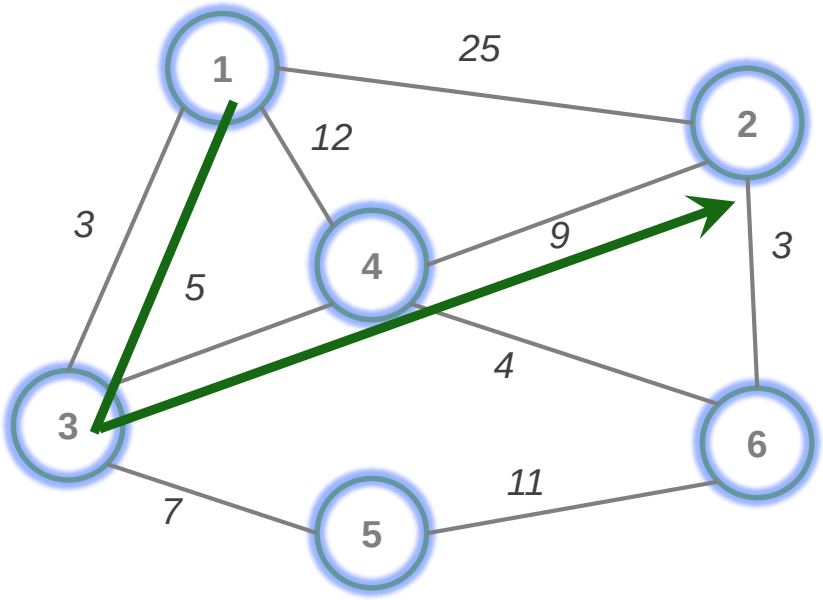
Path computation

- In a traffic-engineering context:

> *The determination of an ordered sequence of nodes and/or links that a traffic path should traverse while satisfying specified constraints*

Path computation algorithms (1)

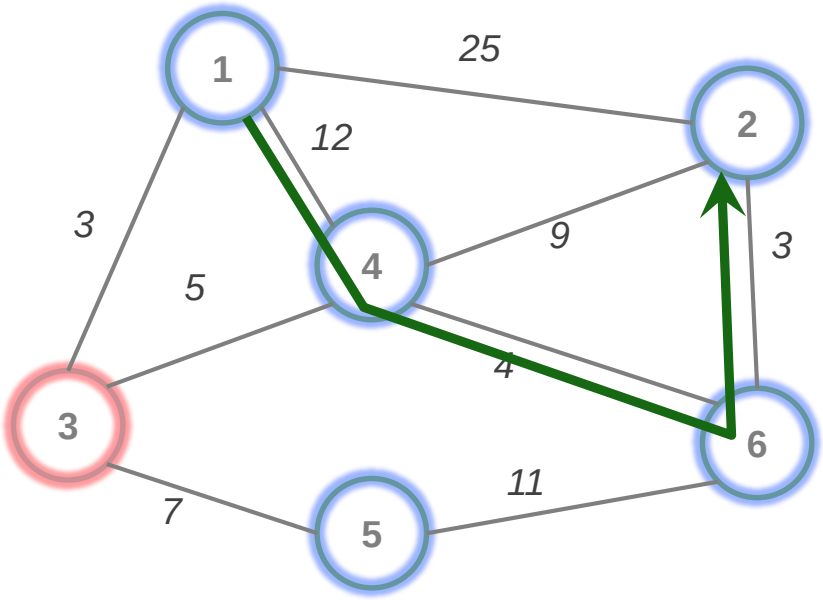
- Dijkstra's algorithm (SPF):
  - Given a network graph comprised of vertices (nodes) and edges (links), the algorithm is able to determine the shortest path between any pair of nodes.
  - The concept of shortest is based on the 'length' of each graph edge



Shortest path from 1 to 2:
1 -> 3 -> 4 -> 2

Path computation algorithms (2)

- Constrained SPF
  - Given a network graph comprised of vertices (nodes) and edges (links), the algorithm is able to determine the shortest path between any pair of nodes, after pruning nodes and links that do not satisfy the required constraints.
  - Used by RSVP-TE in MPLS networks

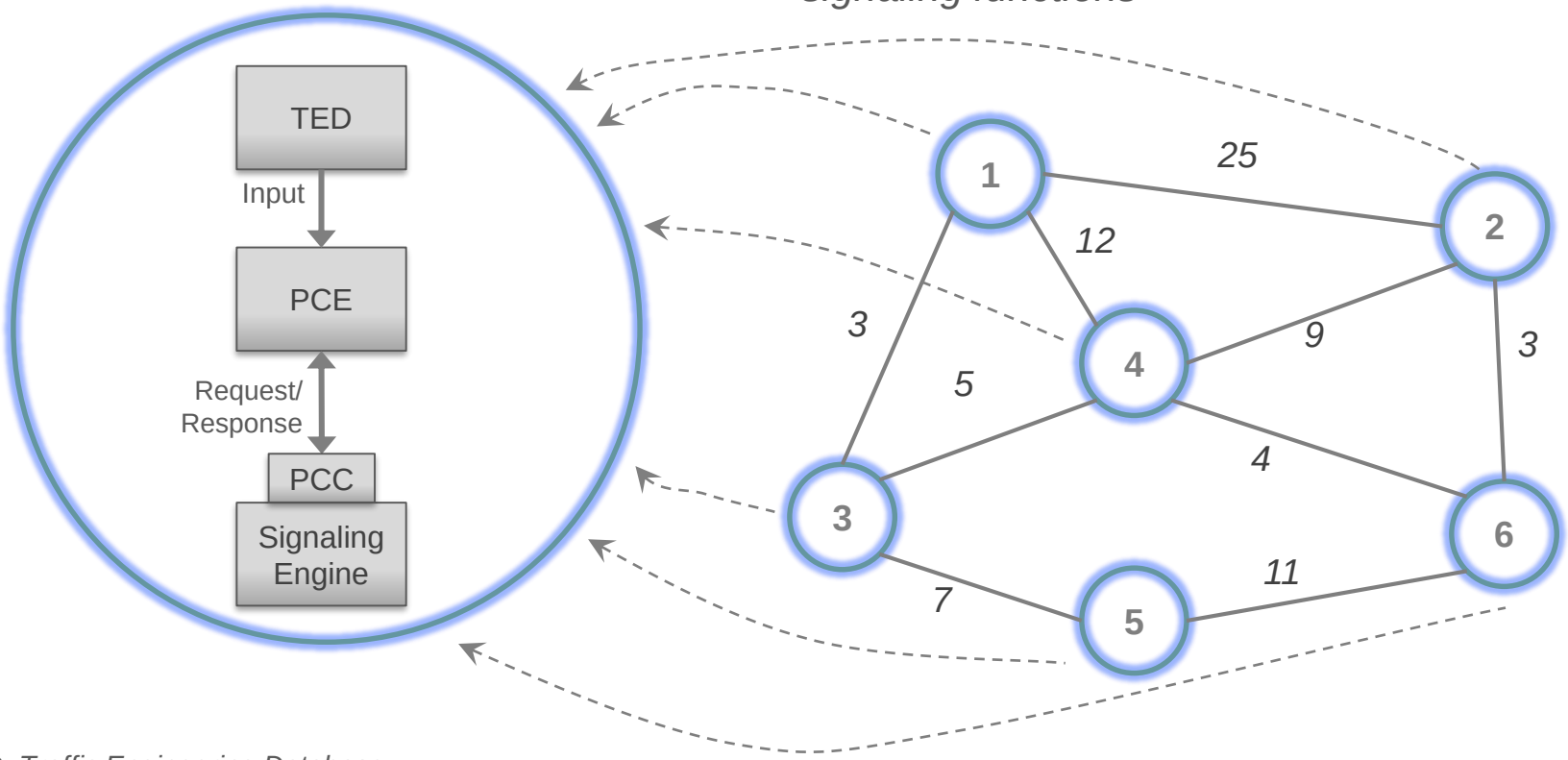


Constraint: only 'blue nodes'

Shortest path from 1 to 2:
1 -> 4 -> 6 -> 2

Current mode of operation

All routers are 'Composite PCE' nodes, having both path computation and path signaling functions



TED: Traffic Engineering Database

Motivations

Key motivations (1)

- CPU-intensive path computations:
  - e.g. complex algorithms that attempt to minimise the maximum link utilisation
- Partial visibility due to TE information not being advertised across domain boundaries:
  - e.g. in case of LSPs that start in one domain and end in another.
- Absence of a traffic-engineering database (TED):
  - e.g. when using devices that may not have the required memory or processing capability to handle large TEDs

Key motivations (2)

- Lack of control plane capability on network elements:
  - e.g. in the case of optical networks
- Backup path computation for bandwidth protection:
  - e.g. co-ordinated computation of backup path resources across multiple LSPs
- Multi-layer networks:
  - e.g. an IP/optical network where the IP network needs to understand resource dependencies in the optical network

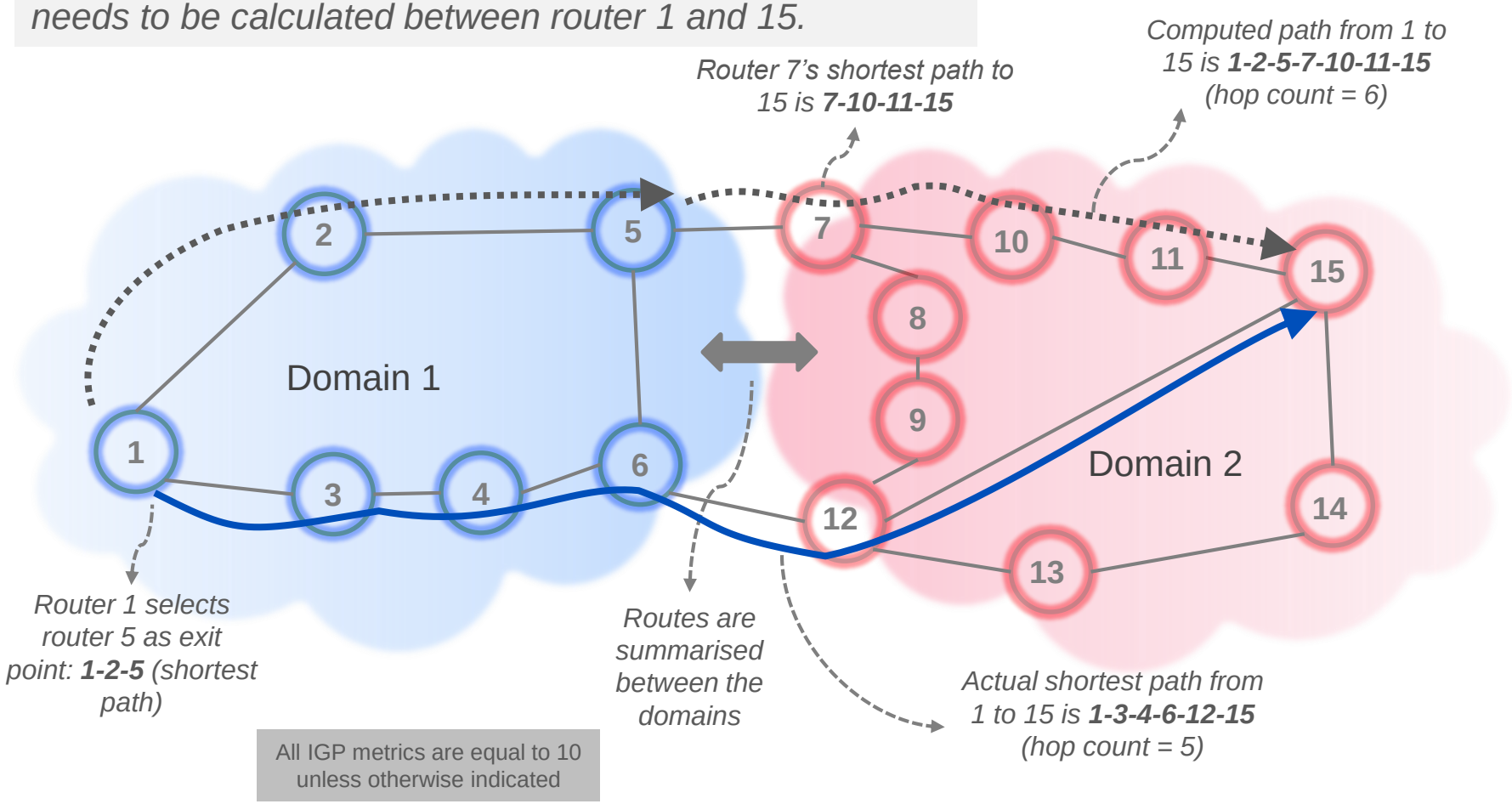
Scenario: inter-domain LSP (1)

- Challenges:
  - Routing and traffic-engineering info is not distributed outside of domain boundaries
  - If we cannot see all links and nodes along the path from a source to destination, how can we construct an optimal end-to-end path ?
  - How can a node select an exit point from its domain ?

- Solutions:
  - Per-domain path computation
  - Simple, co-operating PCEs
  - Backward Recursive Path Computation

Scenario: inter-domain LSP (2)

Requirement: An optimal (based on hop count) TE path needs to be calculated between router 1 and 15.



PCE Architecture

Terminology: information sources

CSPF

Constrained Shortest Path First. Shortest path algorithm that first prunes the network graph of elements that do not satisfy computational constraints.

LSDB

Link-state database with information on nodes within the domain and links inter-connecting them.

Traffic Engineering Database (TED)

Database of topology (nodes and links connecting them) and resource information of the domain.

Terminology: TE-LSPs

Domain

(RFC4655) “Any collection of network elements within a common sphere of address management or path computation responsibility”

LSP

Label Switched Path

Inter-domain TE LSP

A traffic-engineered LSP whose head-end LSR and tail-end LSR do not reside in the same domain, where a domain is either an IGP area, autonomous system or a BGP confederation

Terminology: PCE

Path Computation Client (PCC)

A client application (typically a router) that requests a path computation to be performed by the Path Computation Element (PCE)

Path Computation Element (PCE)

A network entity that is capable of computing a network path given a network graph while taking into account computational constraints.

Path Computation Element Protocol (PCEP)

Protocol used for communication between non-collocated PCCs and PCEs.

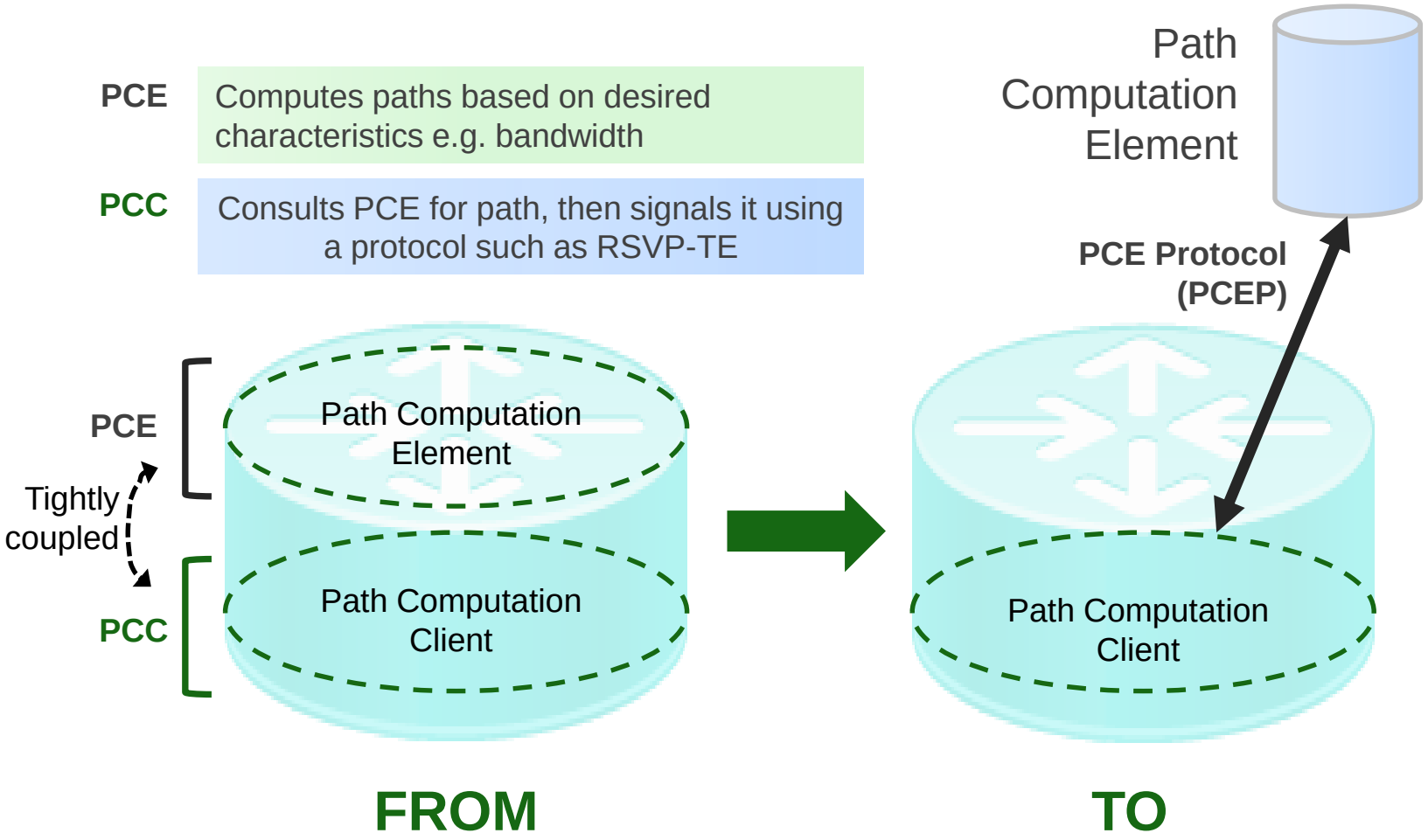
MPLS-TE with PCE

- A Path Computation Element (PCE) can be used to compute MPLS-TE paths:
  - within a “domain” (e.g. an IGP area)
    - May be used to provide greater computational capability, ability to impose additional policy and more complex algorithms
  - across multiple domains (such as a multi-area AS or multiple ASs)
    - Since the TED only includes information for a single area, routers cannot make optimal decisions for constructing end-to-end traffic-engineered paths.
    - One solution to this problem is per-domain path computation (RFC5152). However, it suffers from several limitations.
  - across layers
    - One use case is creating diverse MPLS-TE LSPs at the IP layer while ensuring that no Shared-Risk Link Groups (SRLGs) exist at the optical layer

The PCE split

PCE Computes paths based on desired characteristics e.g. bandwidth

PCC Consults PCE for path, then signals it using a protocol such as RSVP-TE



Path Computation Element (PCE)

- A network entity that is capable of computing a network path given a network graph while taking into account computational constraints such as:
  - required minimum bandwidth
  - maximum latency
  - maximum hop count

- PCE refers to a specific function and does not imply a separate hardware element; it can be located on:
  - a routing node
  - a separate server or,
  - be part of the network management system

PCC and PCE: responsibility split

Path Computation Element (PCE)

- Responsible for calculating paths given:
  - Source node
  - Destination node
  - Required constraints:
    - bandwidth
    - latency
    - SRLGs
    - protection etc
- Is not involved in signaling the actual path

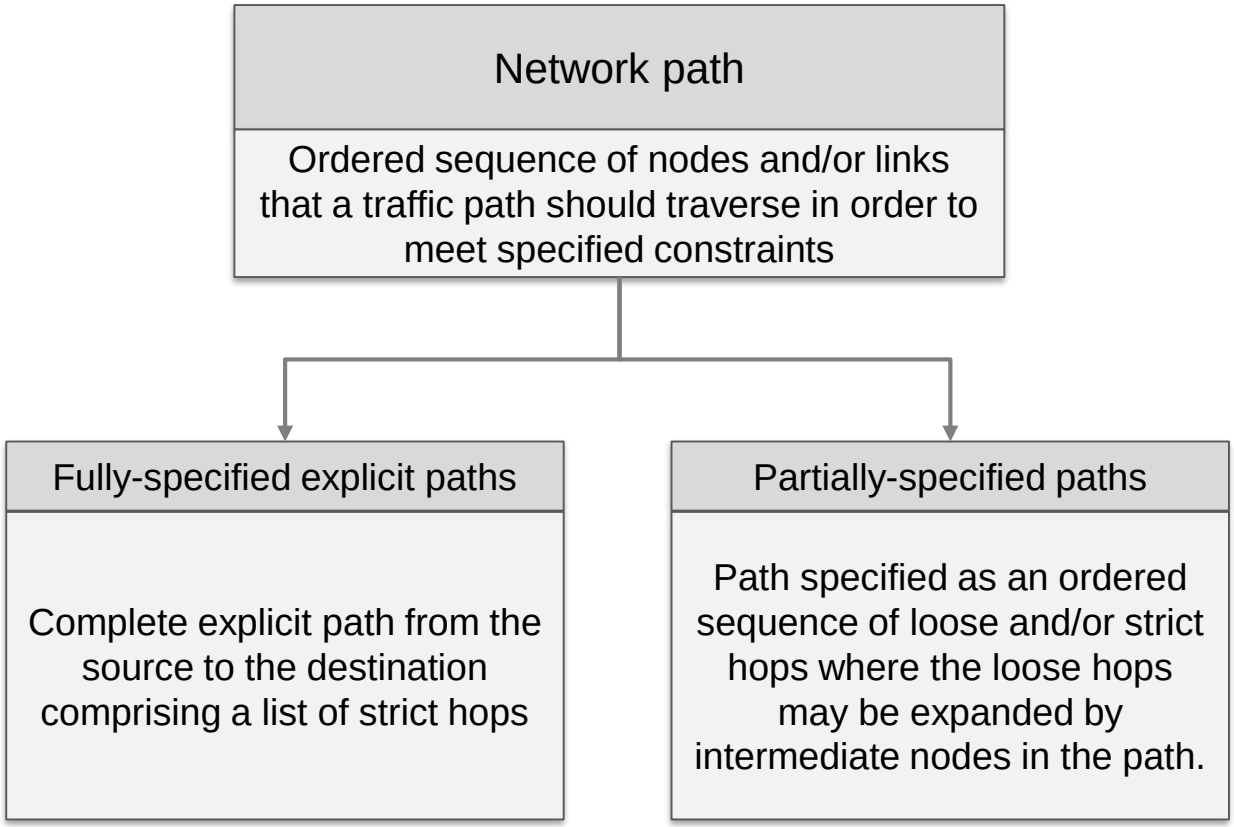
Computation

Path Computation Client (PCC)

- Network element that requests path computation from a PCE
- Supplies all necessary constraint information to the PCC
- Once it receives a path from the PCE, the PCC signals it using RSVP-TE or uses supplied segment list in the header of segment-routed packets

Signaling

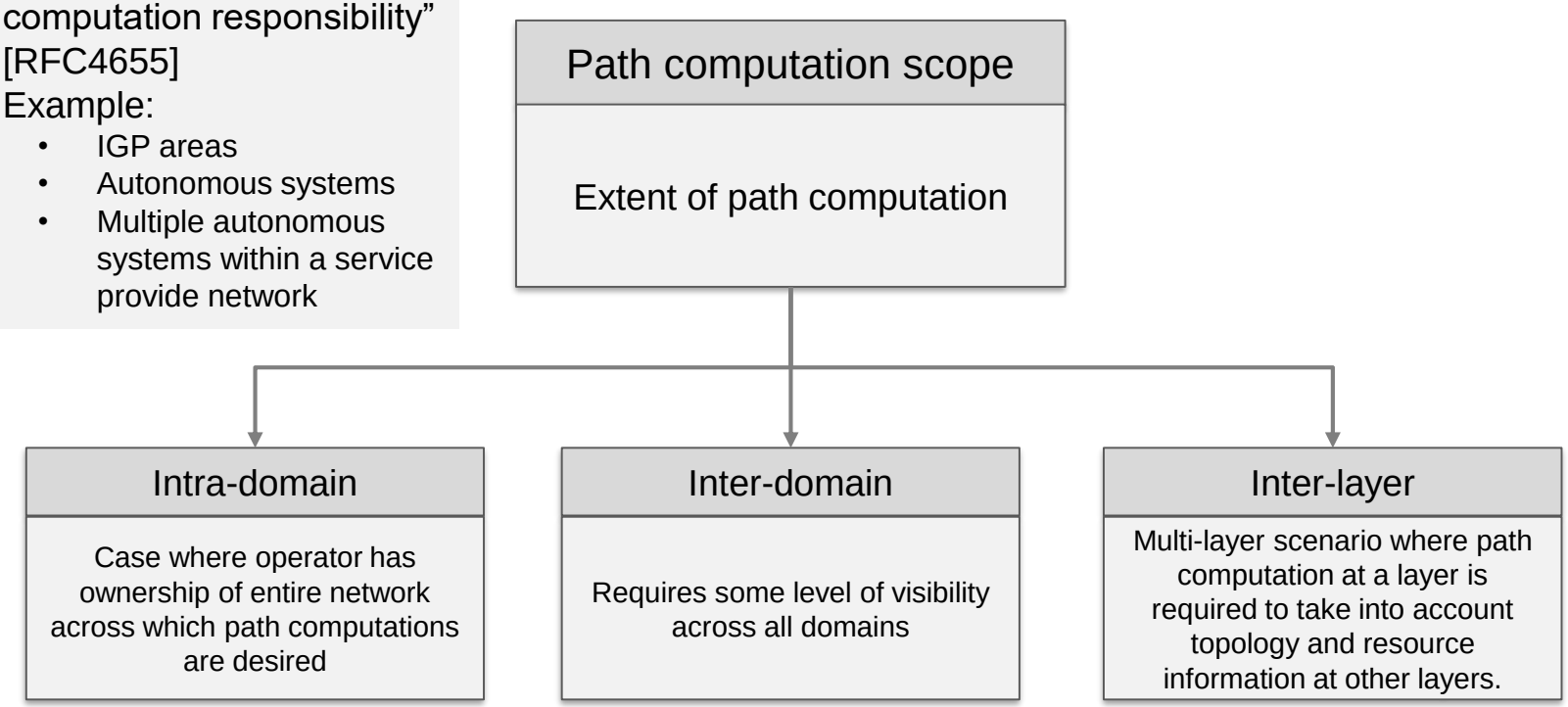
Types of paths



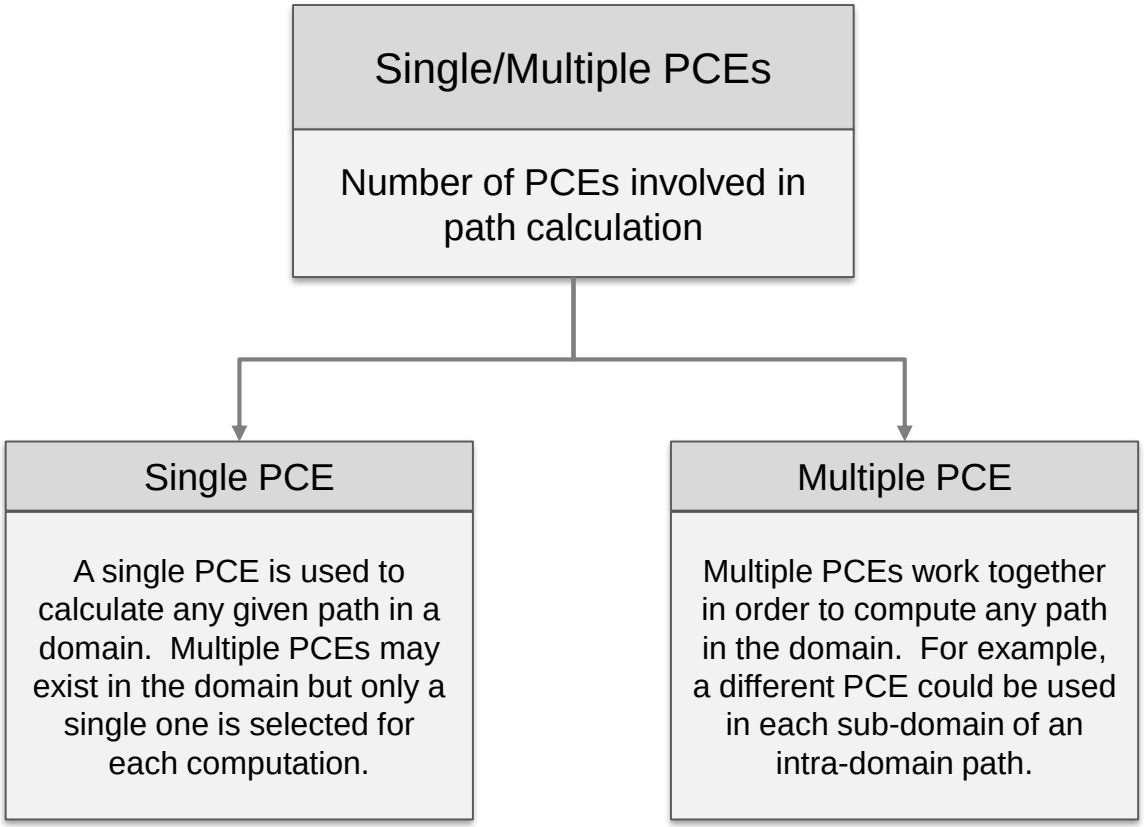
Scope of path computation

Domain

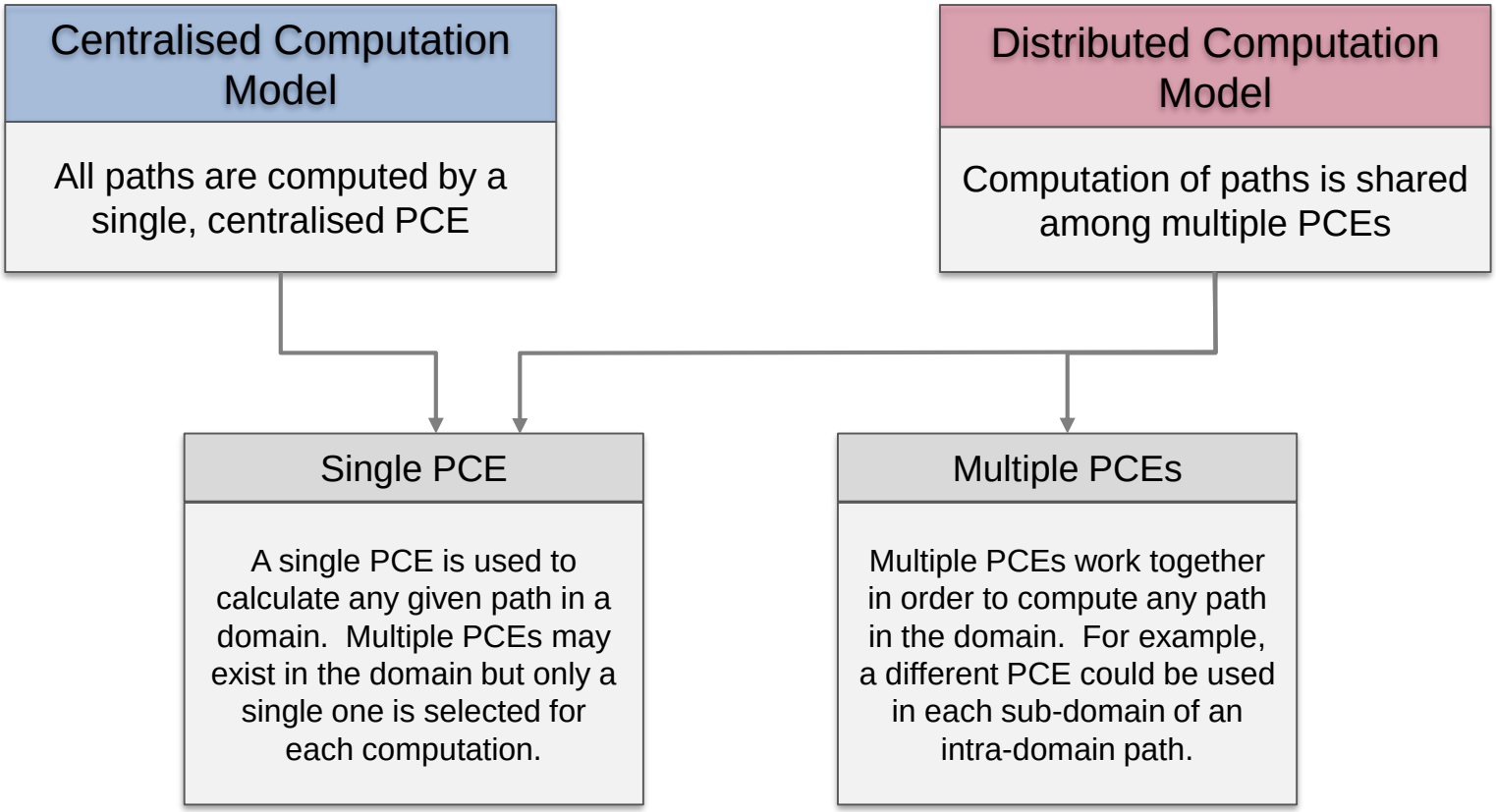
- "Any collection of network elements within a common sphere of address management or path computation responsibility" [RFC4655]
- Example:
  - IGP areas
  - Autonomous systems
  - Multiple autonomous systems within a service provide network



Computation Models (1)



Computation Models (2)



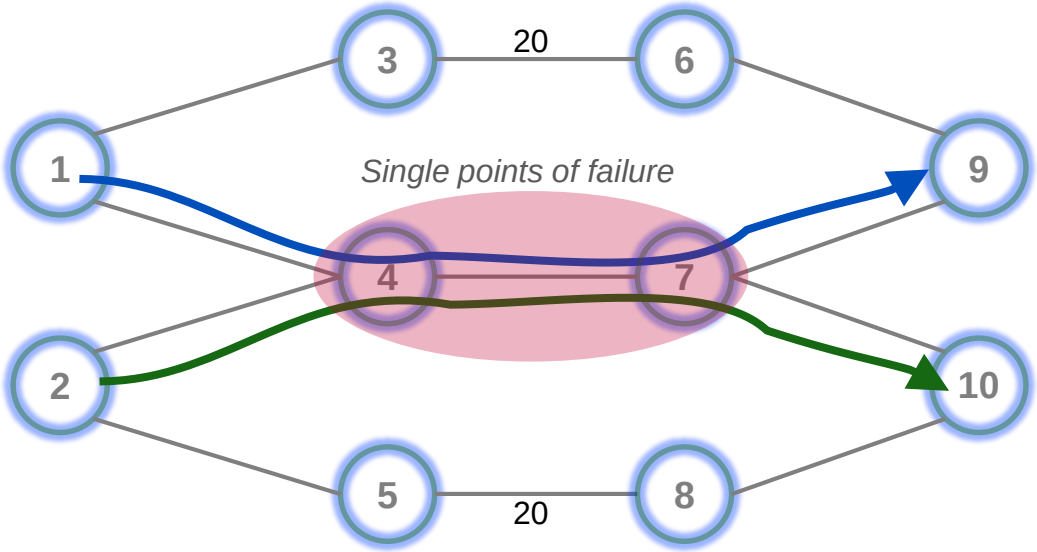
Dependent path computation

- Independent path computation requests are not related to each other.

- Path computations for dependent path requests need to consider all such requests in path calculations e.g. computation of diverse paths for two LSPs

Requirement:

- *2 LSPs:*
  - *From 1 to 9*
  - *From 2 to 10*
- *LSPs need to be path-diverse*



All IGP metrics are equal to 10 unless otherwise indicated

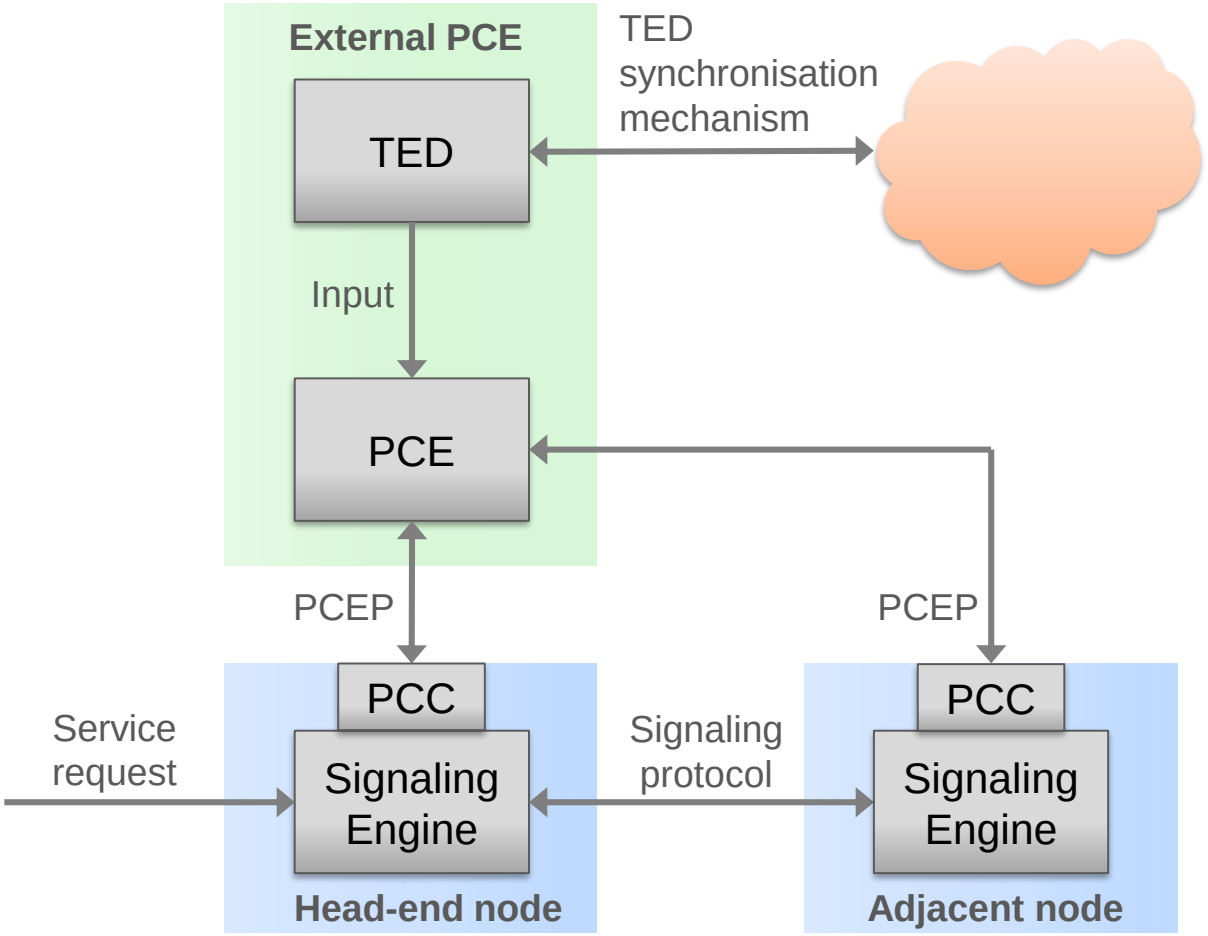
Synchronised path computation

- Path computations for non-synchronised path computation requests may be performed in a serialized manner

- Synchronisation of path computation requests requires the PCE to compute them simultaneously i.e. to consider the best way of performing the computations to allow the greatest number of requests to succeed

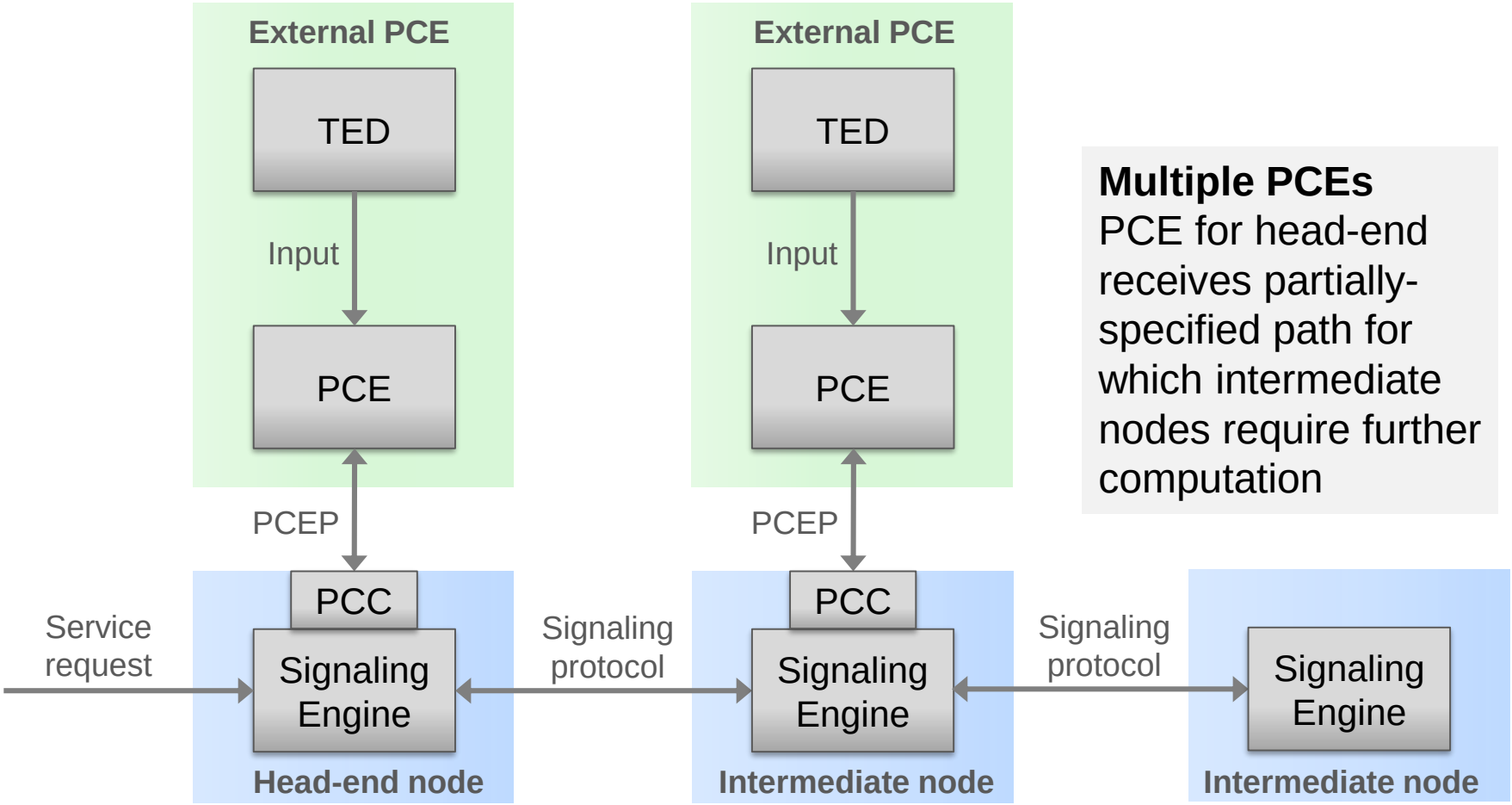
- Dependent path calculations generally tend to be synchronised

Deployment models: External PCE



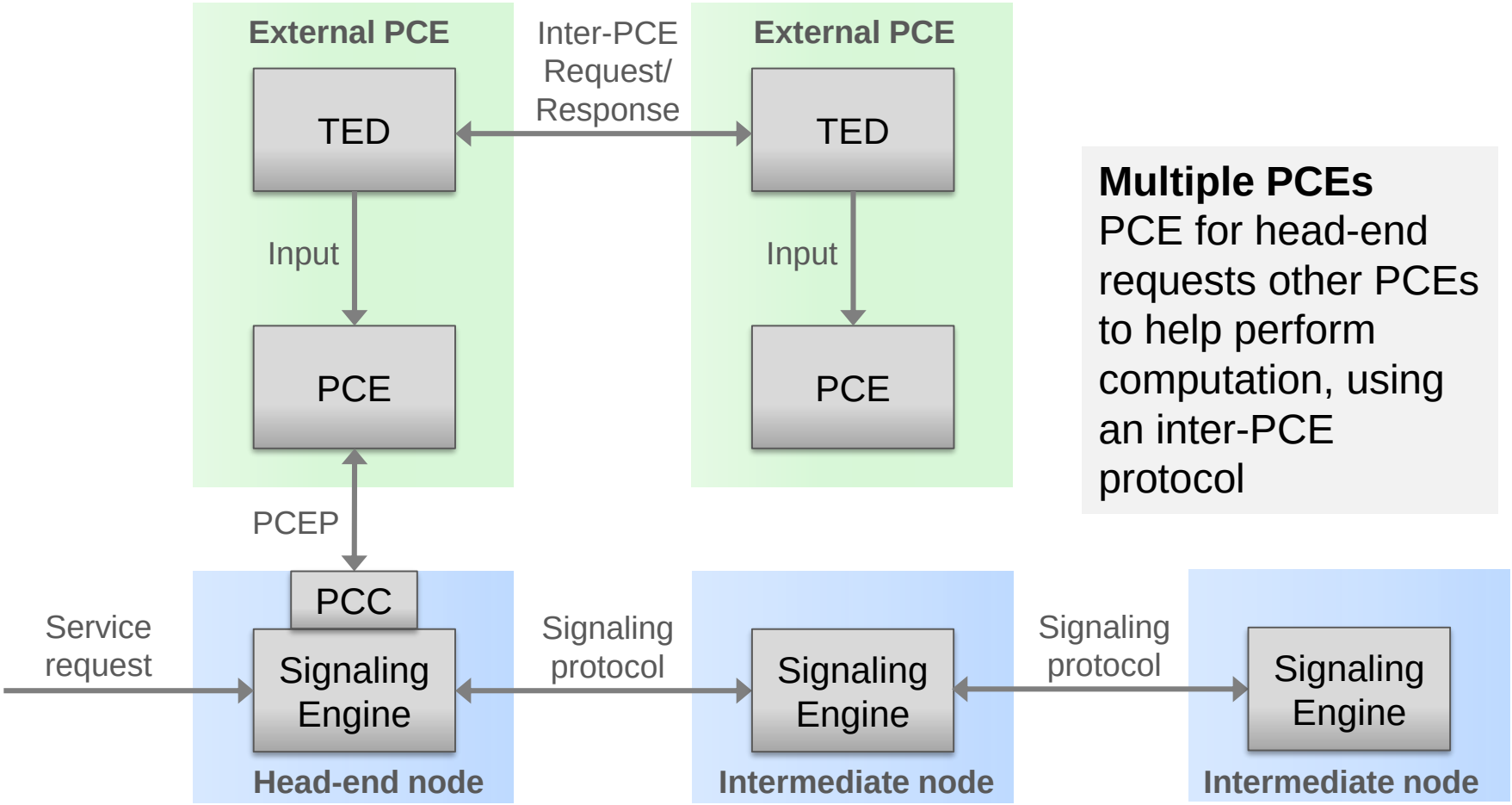
External PCE Node
PCE is physically decoupled from the routing element

Deployment models: Multiple PCEs



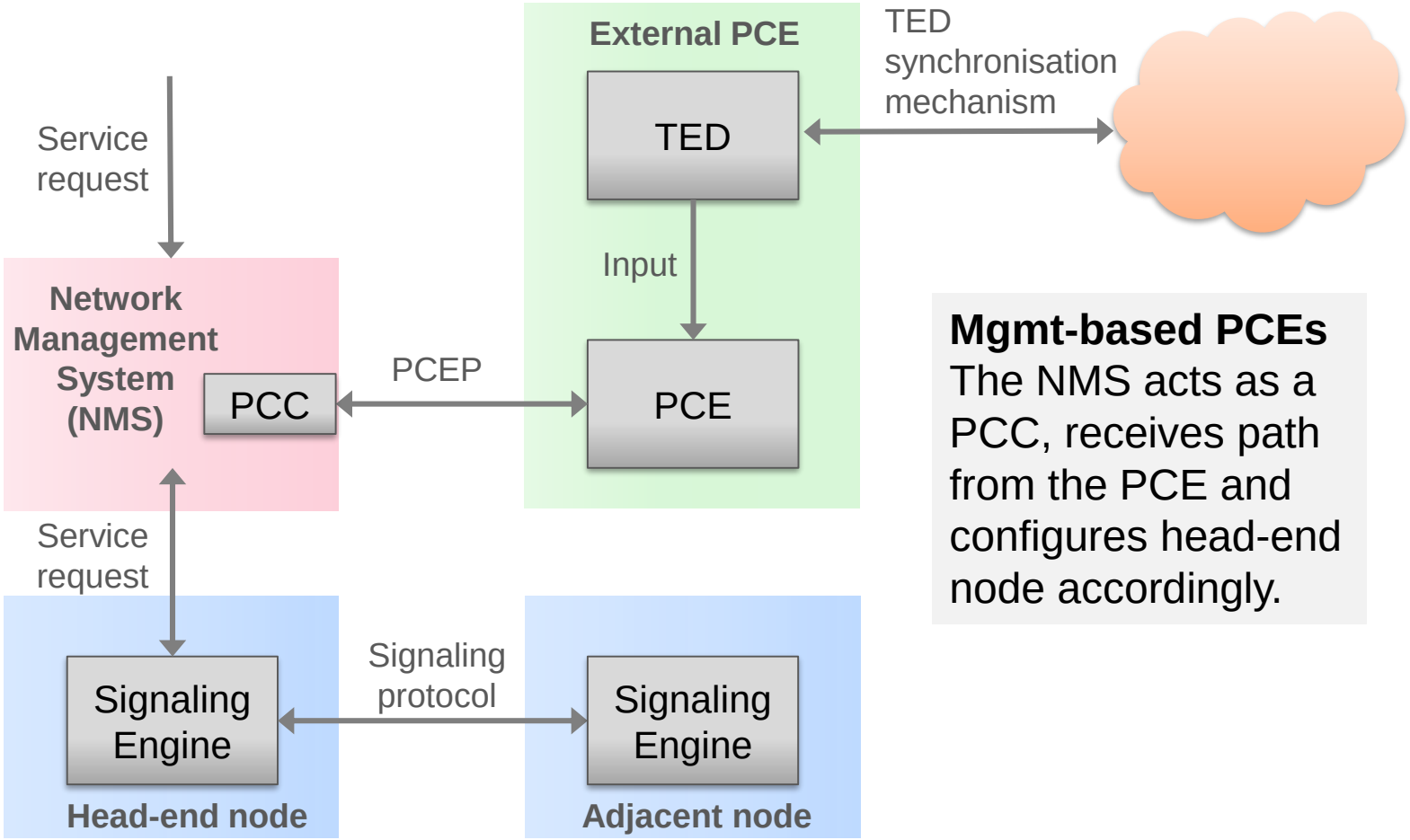
Multiple PCEs
PCE for head-end receives partially-specified path for which intermediate nodes require further computation

Deployment models: Multiple PCEs with inter-PCE comms



Multiple PCEs
PCE for head-end requests other PCEs to help perform computation, using an inter-PCE protocol

Deployment models: Management-based PCE



Mgmt-based PCEs
The NMS acts as a PCC, receives path from the PCE and configures head-end node accordingly.

Topology visibility for PCE

- There is no standards-defined mechanism mandated for a PCE to acquire network topology
- One option that is used by many implementations for the PCE is to use the services of a passive IGP listener
- BGP-LS provides an alternative option

Topology and TE information

Link-state database (LSDB)

- Router IP addresses
- Router link attributes:
  - Neighbor router identifier
  - Local Interface IP address
  - Remote interface IP address
  - IP subnet address of link
  - IGP link metric

Traffic-engineering database (TED)

- Router IP addresses
- Router link attributes:
  - Local Interface IP address
  - Remote interface IP address
  - Traffic-engineering metric
  - Maximum bandwidth
  - Maximum reservable bandwidth
  - Unreserved bandwidth
  - Administrative group
  - Per-CoS reservation state
  - Pre-emption
  - SRLGs

Stateful PCE

Stateless PCEs

- All of the material covered thus far assumed Stateless PCEs.

- Stateless PCEs have knowledge of the Link-State Database and the Traffic Engineering Database and are able to make path computation decisions based on that
  - they generally do not have knowledge of existing paths and computation requests can be processed independently of LSPs already present in the network

Stateful PCEs

- In addition to the LSDB and the TED, Stateful PCEs have complete knowledge of all paths (LSPs) in the network and the resources they have reserved in the network.
  - paths computed by Stateful PCEs also consider existing LSPs and resource reservations
  - mechanisms are required to ensure Stateful PCEs are strictly synchronised with the network

Terminology (1)

Passive Stateful PCE

A PCE that takes into account LSP state information it has acquired from PCCs when calculating new paths. It does not actively update LSP state after initial path computation.

Active Stateful PCE

A stateful PCE, in addition to considering existing LSPs in new path calculations, also issues recommendations on updating LSPs subsequent to the initial path calculation.

Terminology (2)

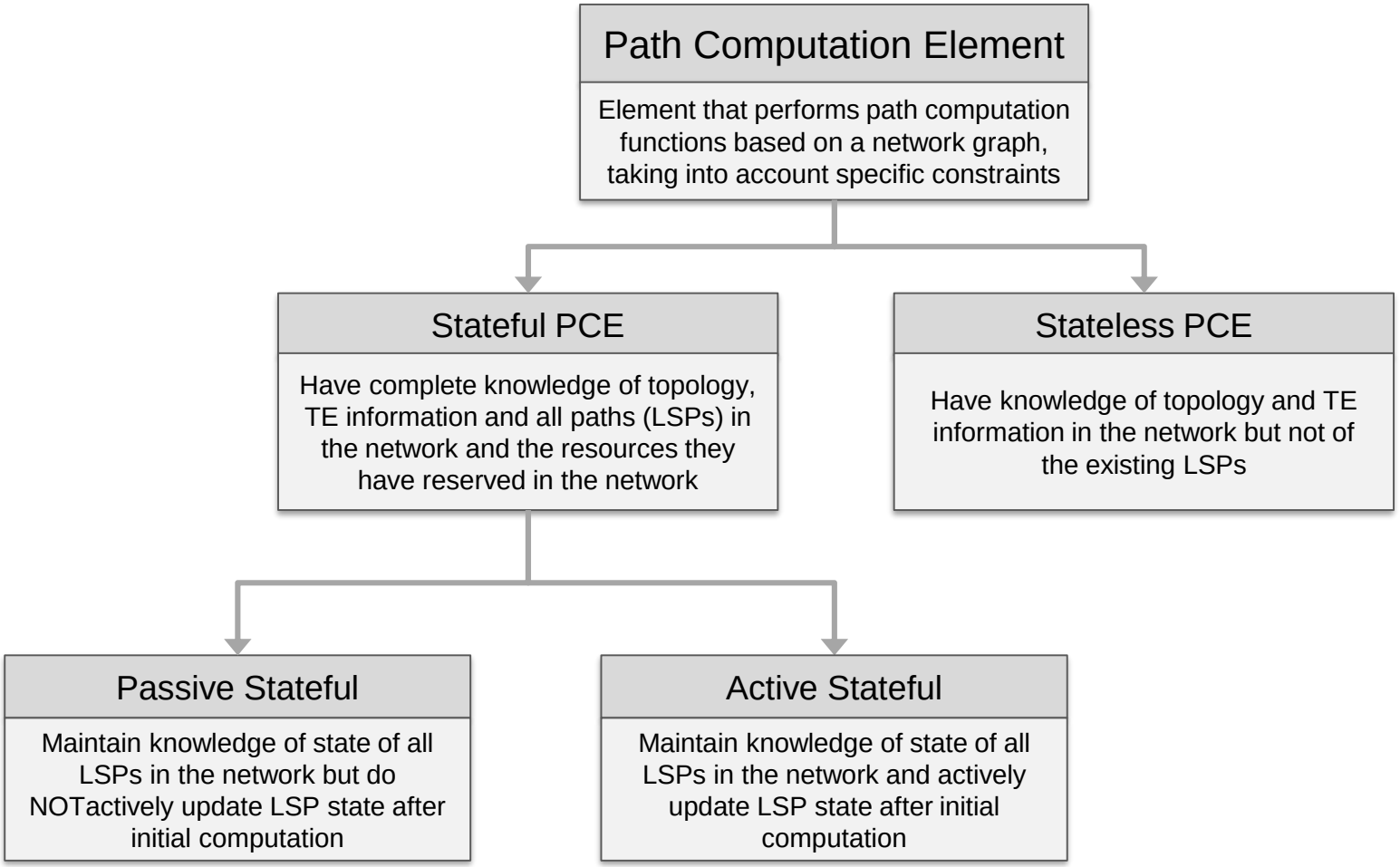
Delegation

Mechanism by which a PCC grants a PCE the right to update LSP parameters for an LSP belonging to that PCC.

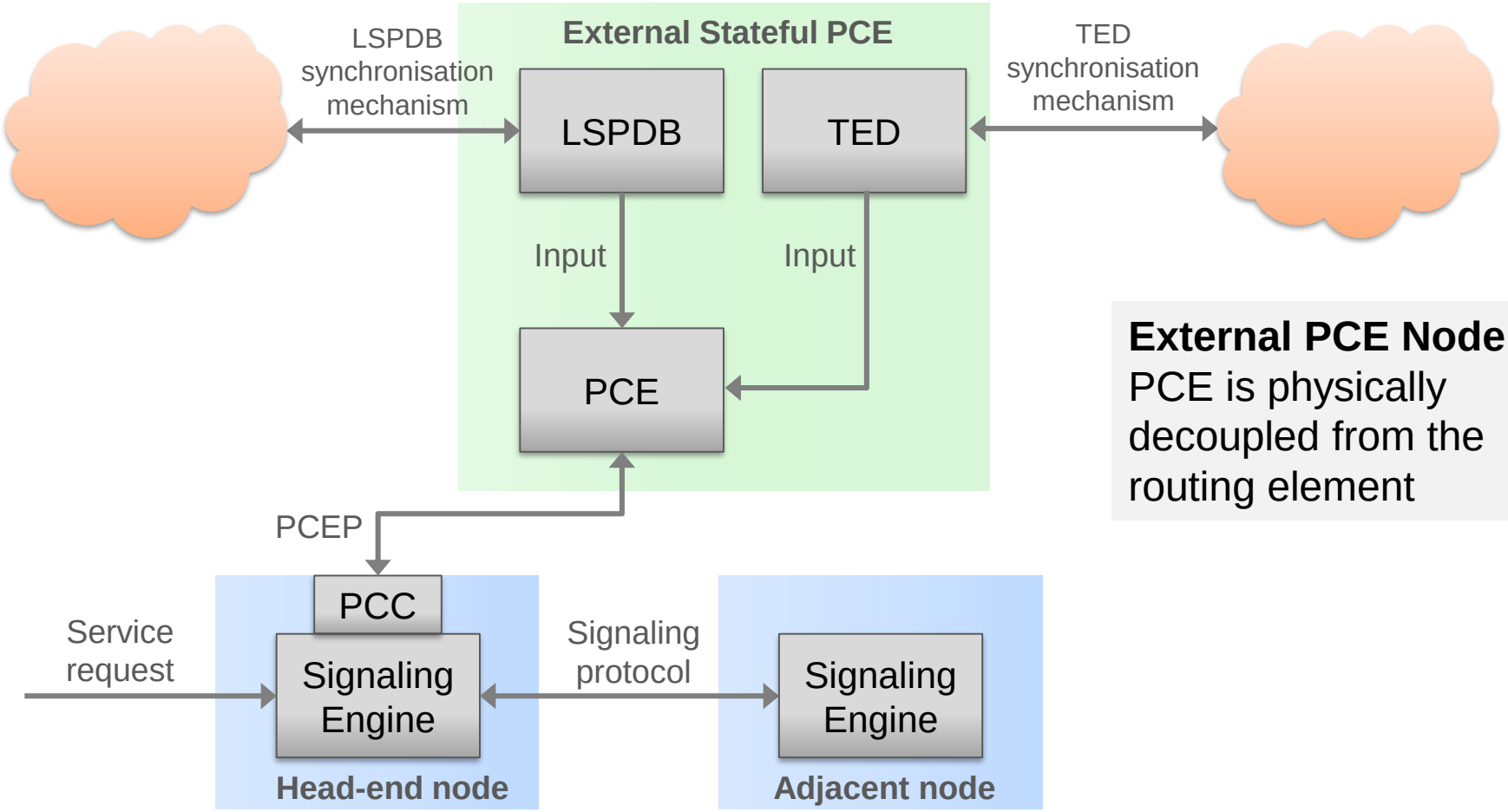
LSP DB

Database where a PCE stores information on all active LSPs in the network: complete network path, resource usage, constraints etc.

PCE taxonomy



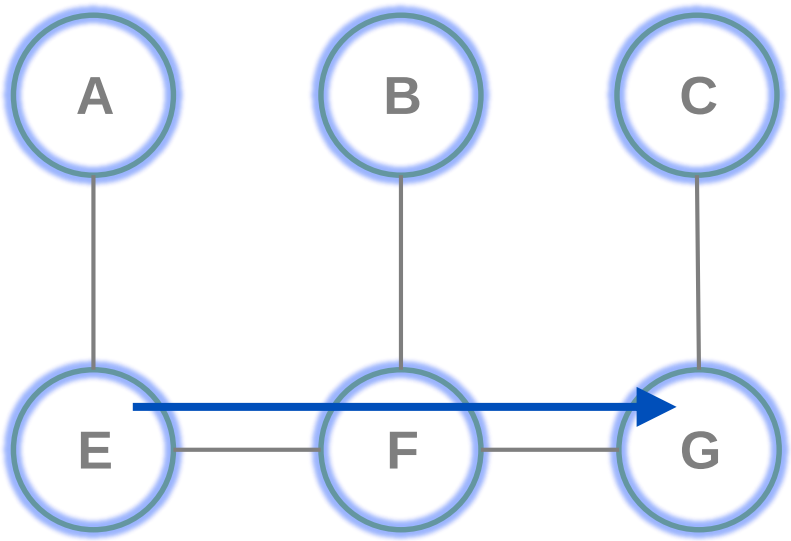
Deployment model: Stateful PCE



External PCE Node
PCE is physically decoupled from the routing element

Motivations: LSP placement (1a)

Use case: *Throughput maximisation*
Objective: *Compute LSP placement in order to maximise utilisation of the network*
PCE type: *Stateless*



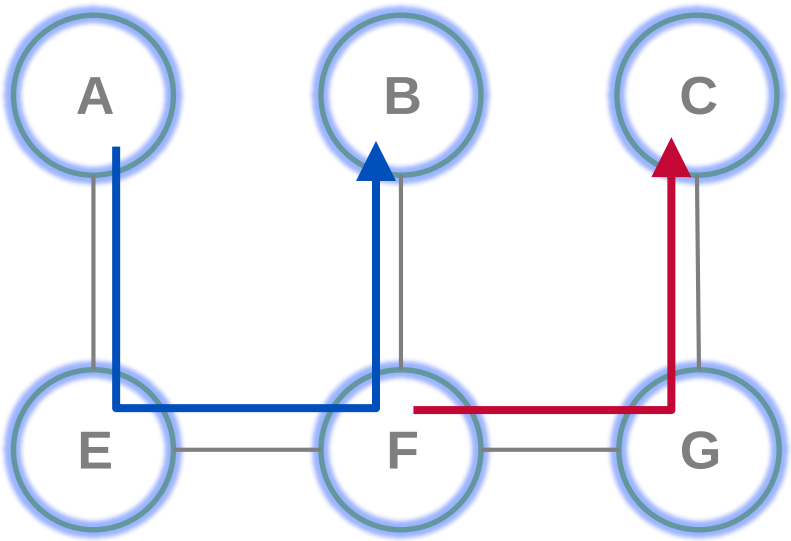
All IGP metrics are equal to 10 with bandwidth of 10Gbps unless otherwise indicated

LSP computation requests:						
Time	**LSP**	**Src**	**Dest**	**Demand**	**Routable ?**	**Path**
t1	1	E	G	10	Yes	E-F-G
t2	2	A	B	10	No	-
t3	3	F	C	10	No	-

Outcome:
Only 40% of the network throughput is utilised as LSPs 2 and 3 could not be accommodated.

Motivations: LSP placement (1b)

Use case: *Throughput maximisation*
Objective: *Compute LSP placement in order to maximise utilisation of the network*
PCE type: *Stateful*



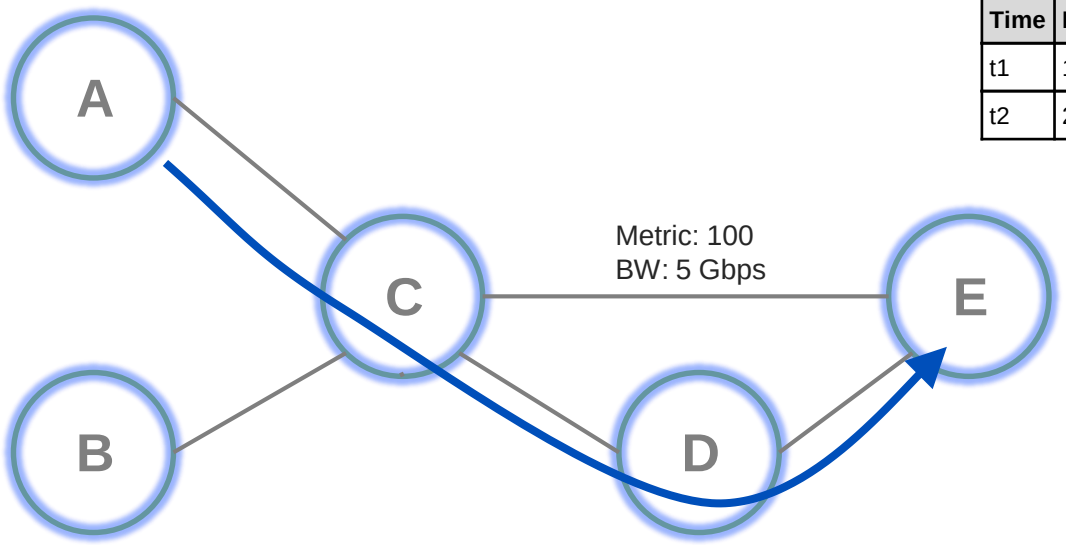
All IGP metrics are equal to 10 with bandwidth of 10Gbps unless otherwise indicated

LSP computation requests:						
Time	**LSP**	**Src**	**Dest**	**Demand**	**Routable ?**	**Path**
t1	1	E	G	10	No	-
t2	2	A	B	10	Yes	A-E-F-B
t3	3	F	C	10	Yes	F-G-C

Outcome:
100% of the network throughput is utilised if LSP1 is rejected and LSPs 2 and 3 are accepted.

Motivations: LSP placement (2a)

Use case: *Throughput maximisation*
Objective: *Compute LSP placement in order to maximise utilisation of the network*
PCE type: *Stateless*



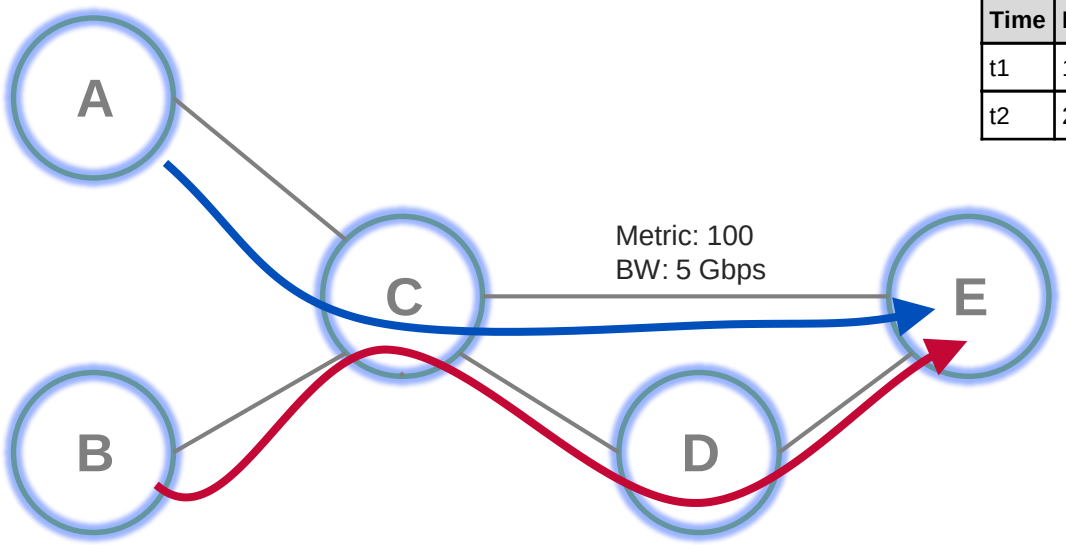
All IGP metrics are equal to 10 with bandwidth of 10Gbps unless otherwise indicated

LSP computation requests:						
Time	**LSP**	**Src**	**Dest**	**Demand**	**Routable ?**	**Path**
t1	1	A	E	5	Yes	A-C-D-E
t2	2	B	E	10	No	-

Outcome:
Only part of the network throughput is utilised as LSP 2 could not be accommodated.

Motivations: LSP placement (2b)

Use case: *Throughput maximisation*
Objective: *Compute LSP placement in order to maximise utilisation of the network*
PCE type: *Stateful*



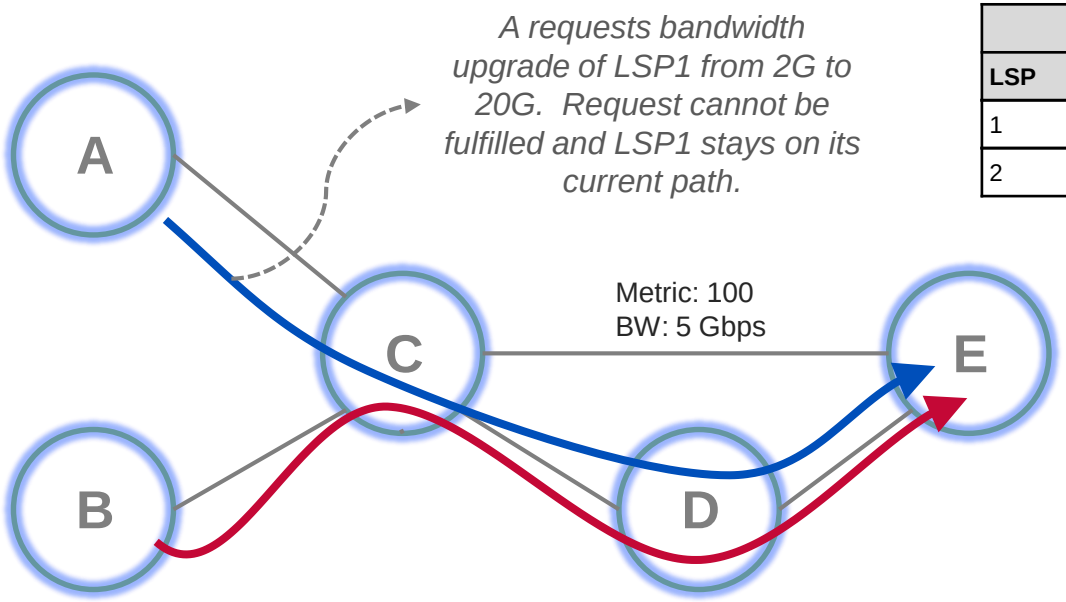
All IGP metrics are equal to 10 with bandwidth of 10Gbps unless otherwise indicated

LSP computation requests:						
Time	**LSP**	**Src**	**Dest**	**Demand**	**Routable ?**	**Path**
t1	1	A	E	5	Yes	A-C--E
t2	2	B	E	10	Yes	B-C-D-E

Outcome:
By intelligent placement of LSP1, LSP 2 could also be accommodated

Motivations: LSP placement (3a)

Use case: Deadlock
Objective: Avoid effects of deadlock when an LSP bandwidth increase request cannot be satisfied
PCE type: Stateless



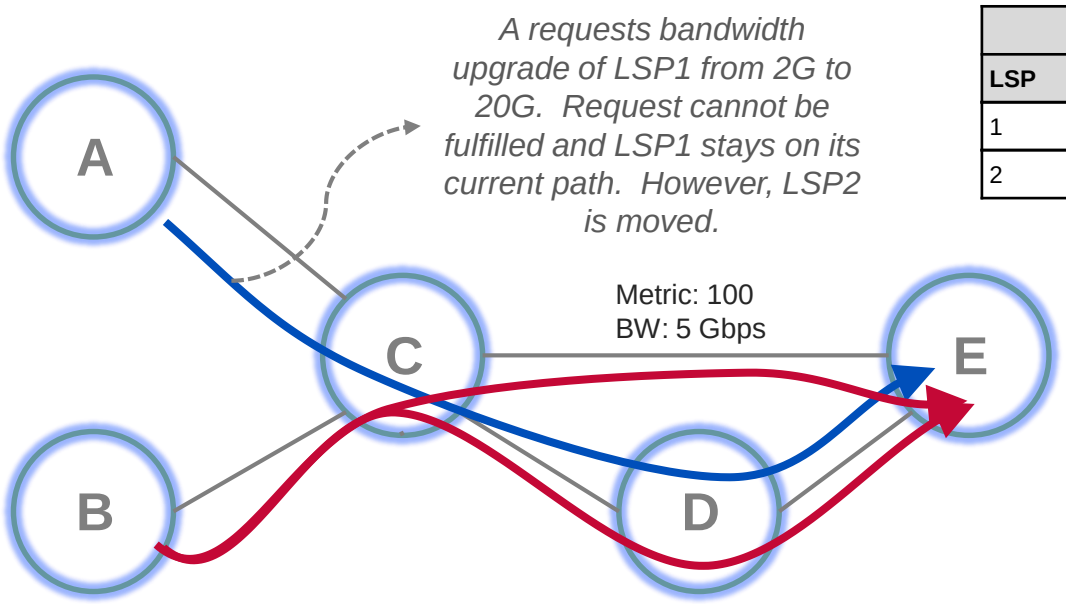
LSPs				
LSP	**Src**	**Dest**	**Demand**	**Path**
1	A	E	2	A-C-D-E
2	B	E	2	B-C-D-E

All IGP metrics are equal to 10 with bandwidth of 10Gbps unless otherwise indicated

Outcome:
Failure to satisfy bandwidth upgrade request for LSP1 degrades its performance as well as that of LSP2

Motivations: LSP placement (3b)

Use case: Deadlock
Objective: Avoid effects of deadlock when an LSP bandwidth increase request cannot be satisfied
PCE type: Stateful



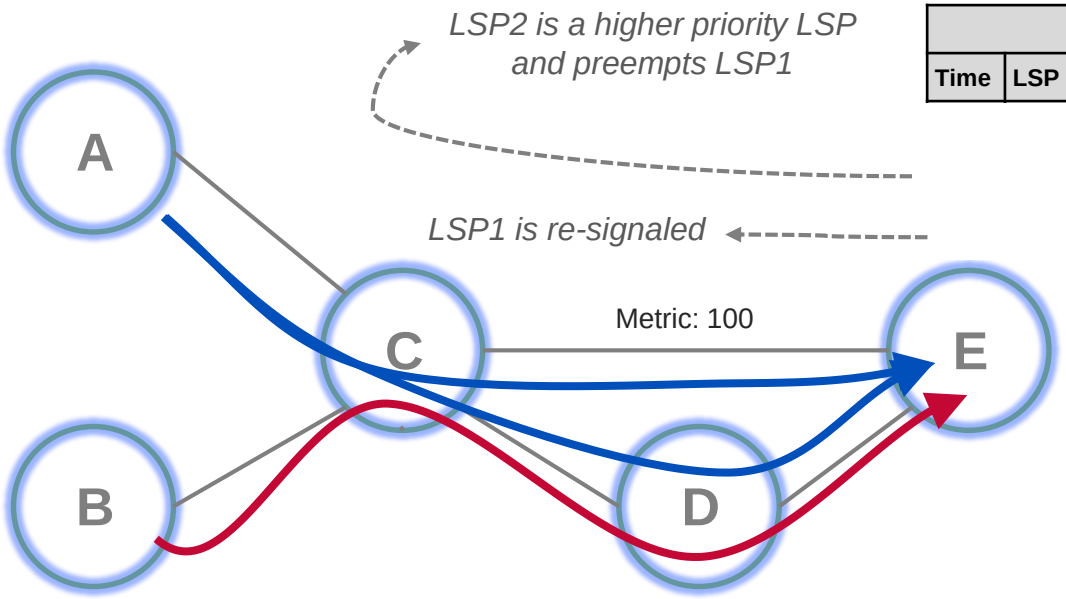
LSPs				
LSP	**Src**	**Dest**	**Demand**	**Path**
1	A	E	2	A-C-D-E
2	B	E	2	**B-C-E**

All IGP metrics are equal to 10 with bandwidth of 10Gbps unless otherwise indicated

Outcome:
Failure to satisfy bandwidth upgrade request for LSP1 degrades its performance but does not affect LSP2 as it is moved to another, diverse path.

Motivations: LSP placement (4a)

Use case: *Minimum perturbation*
Objective: *Limit the impact on the network (in terms of affected LSPs) when new LSP paths are computed*
PCE type: *Stateless*



LSP computation requests:							
Time	LSP	Src	Dest	Demand	Priority	Routable ?	Path

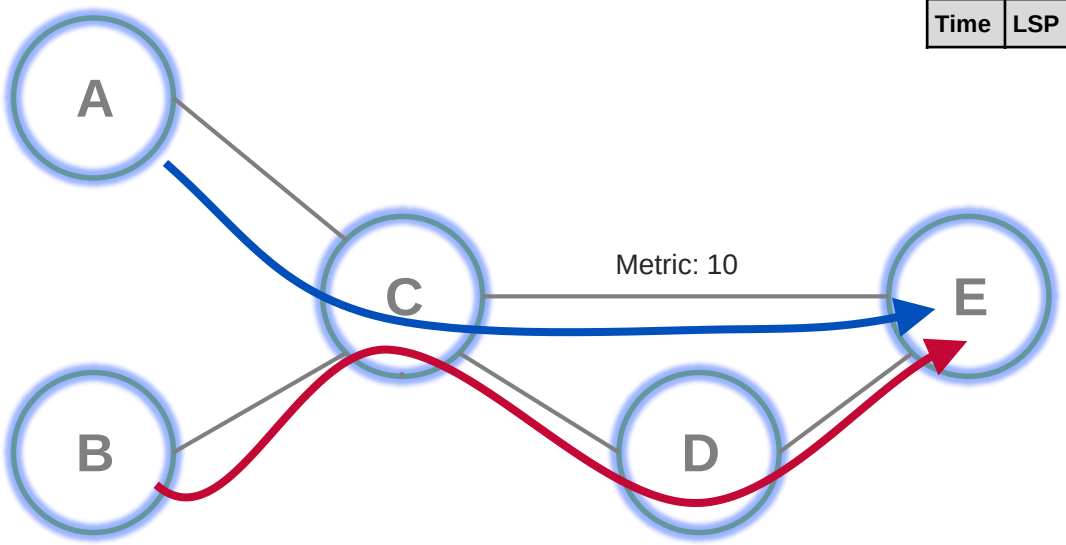
Outcome:
Unnecessary network perturbation resulted from LSP1 being pre-empted

All IGP metrics are equal to 10 with bandwidth of 10Gbps unless otherwise indicated

Motivations: LSP placement (4b)

Use case: *Minimum perturbation*
Objective: *Limit the impact on the network (in terms of affected LSPs) when new LSP paths are computed*
PCE type: *Stateful*

LSP computation requests:							
Time	LSP	Src	Dest	Demand	Priority	Routable ?	Path

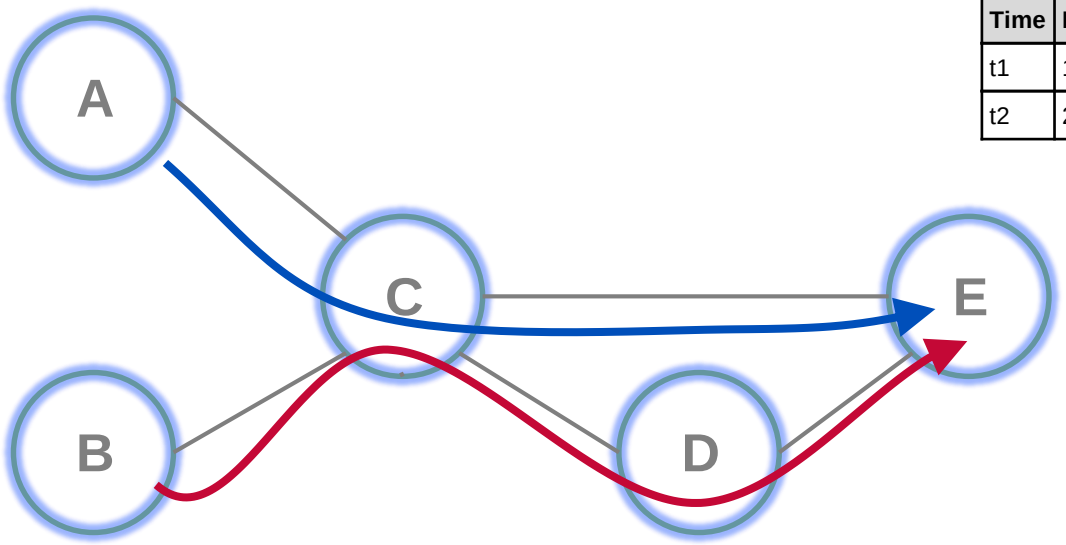


All IGP metrics are equal to 10 with bandwidth of 10Gbps unless otherwise indicated

Outcome:
Unnecessary network perturbation was avoided assuming the path computation requests were relatively close in time.

Motivations: LSP placement (5a)

Use case: *Predictability*
Objective: *To provide a better degree of predictability over how paths are routed*
PCE type: *Stateless*



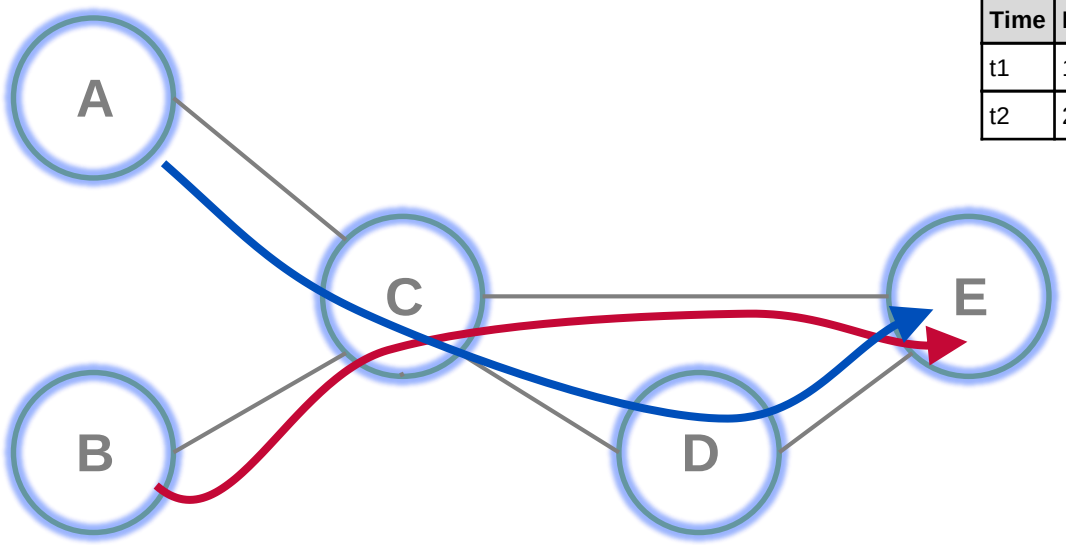
All IGP metrics are equal to 10 with bandwidth of 10Gbps unless otherwise indicated

LSP computation requests:						
Time	**LSP**	**Src**	**Dest**	**Demand**	**Routable ?**	**Path**
t1	1	A	E	7	Yes	A-C-E
t2	2	B	E	7	Yes	B-C-D-E

Outcome:
LSP from A to E gets the shorter path since it was requested first

Motivations: LSP placement (5b)

Use case: *Predictability*
Objective: *To provide a better degree of predictability over how paths are routed*
PCE type: *Stateless*



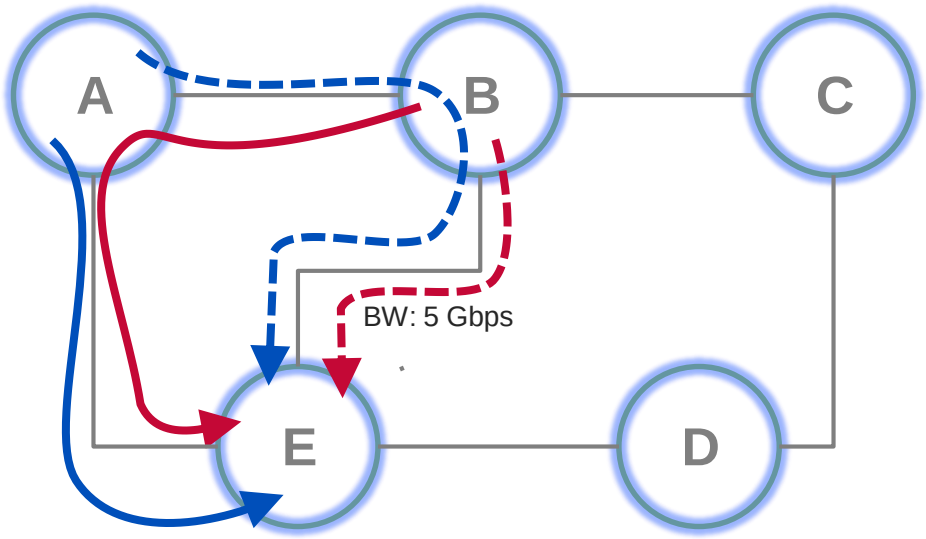
All IGP metrics are equal to 10 with bandwidth of 10Gbps unless otherwise indicated

LSP computation requests:						
Time	**LSP**	**Src**	**Dest**	**Demand**	**Routable ?**	**Path**
t1	1	B	E	7	Yes	B-C-E
t2	2	A	E	7	Yes	A-C-D-E

Outcome:
LSP from B to E gets the shorter path since it was requested first

Motivations: Recovery (6a)

Use case: *Co-ordinated protection*
Objective: *To ensure protection resources are not over-subscribed*
PCE type: *Stateless*



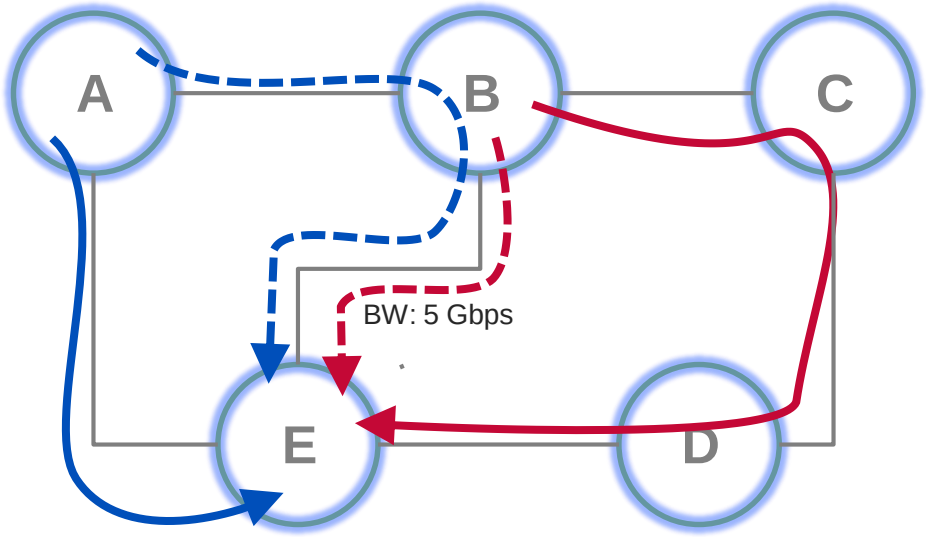
All IGP metrics are equal to 10 with bandwidth of 10Gbps unless otherwise indicated

LSP computation requests:						
Time	**LSP**	**Src**	**Dest**	**Demand**	**Routable ?**	**Path**
t1	1-W	A	E	5	Yes	A-E
	1-P	A	E		Yes	A-B-E
t2	2-W	B	E	5	Yes	B-A-E
	2-P	B	E		Yes	B-E

Outcome:
If link A-E fails, both LSPs fail and end up using the same protection resources (B-E link)

Motivations: Recovery (6b)

Use case: *Co-ordinated protection*
Objective: *To ensure protection resources are not over-subscribed*
PCE type: *Stateful*



All IGP metrics are equal to 10 with bandwidth of 10Gbps unless otherwise indicated

LSP computation requests:						
Time	**LSP**	**Src**	**Dest**	**Demand**	**Routable ?**	**Path**
t1	1-W	A	E	5	Yes	A-E
	1-P	A	E		Yes	A-B-E
t2	2-W	B	E	5	Yes	B-C-D-E
	2-P	B	E		Yes	B-E

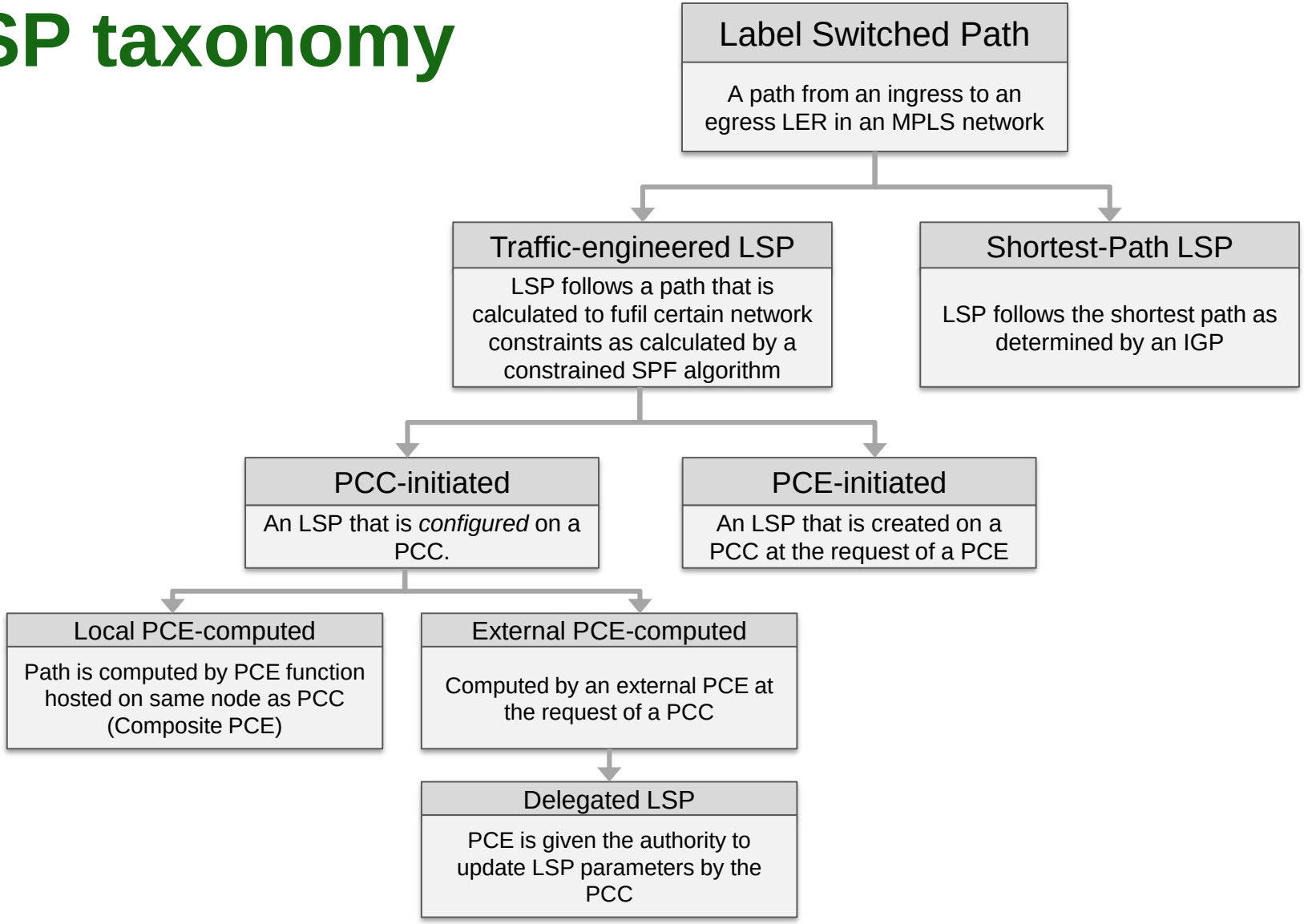
Outcome:
LSPs 1 and 2 share no SRLGs so can share protection resources (B-E link)

PCE-initiated LSPs

PCE-initiated LSPs

- Till now, we have considered PCE models where a PCE computes LSP paths at the request of PCCs i.e. all LSP-creation requests were PCC-initiated.
- In a stateful PCE scenario:
  - PCEs are synchronised with all LSPs on all PCCs and are kept informed of changes impacting any of these LSPs
  - The PCC may delegate a PCE the authority to make changes to LSPs subsequent to their creation.
- A PCE-initiated LSP is one that is created on a PCC at the request of a PCE

LSP taxonomy



Motivations

- PCC-instantiated LSPs are suited to networks where traffic flows are reasonably static.
- The PCE-initiated LSP is better suited to SDN environments requiring on-demand connectivity
  - Dynamic network optimisation algorithms that place new LSPs and update existing LSPs in order to better balance traffic in the network
  - NFV applications where VNFs are dynamically instantiated, scaled and torn down

Path Computation Element Protocol (PCEP)

Objectives

- Path Computation Element Protocol (PCEP): protocol used for communication between:
  - a PCC and a PCE, OR
  - between two PCEs
- A PCC uses the PCEP to send a path computation request for one or more TE LSPs to a PCE
- The PCE replies to the PCC with zero or more computed paths

Sessions

- Runs over TCP on port 4189

- For a stateless PCE, PCEP sessions can be either:
  - Permanent: kept permanently up, OR
  - Temporary: in which case a session is established whenever a path computation is required and torn down when completed.

- For a stateful PCE, PCEP sessions are required to be set up on a permanent basis

PCEP Messages (1)

Value	Type	Abbreviated name
1	Open	
2	Keepalive	
3	Path Computation Request	PCReq
4	Path Computation Reply	PCRep
5	Notification	PCNtf
6	Error	
7	Close	
8	Path Computation Monitoring Request	PCMonReq
9	Path Computation Monitoring Reply	PCMonRep
10	Report	PCRpt
11	Update	PCUpd
12	Initiate	PCInit

Original set of messages specified for Stateless PCE (RFC5440)

Extensions for PCE monitoring (RFC5886)

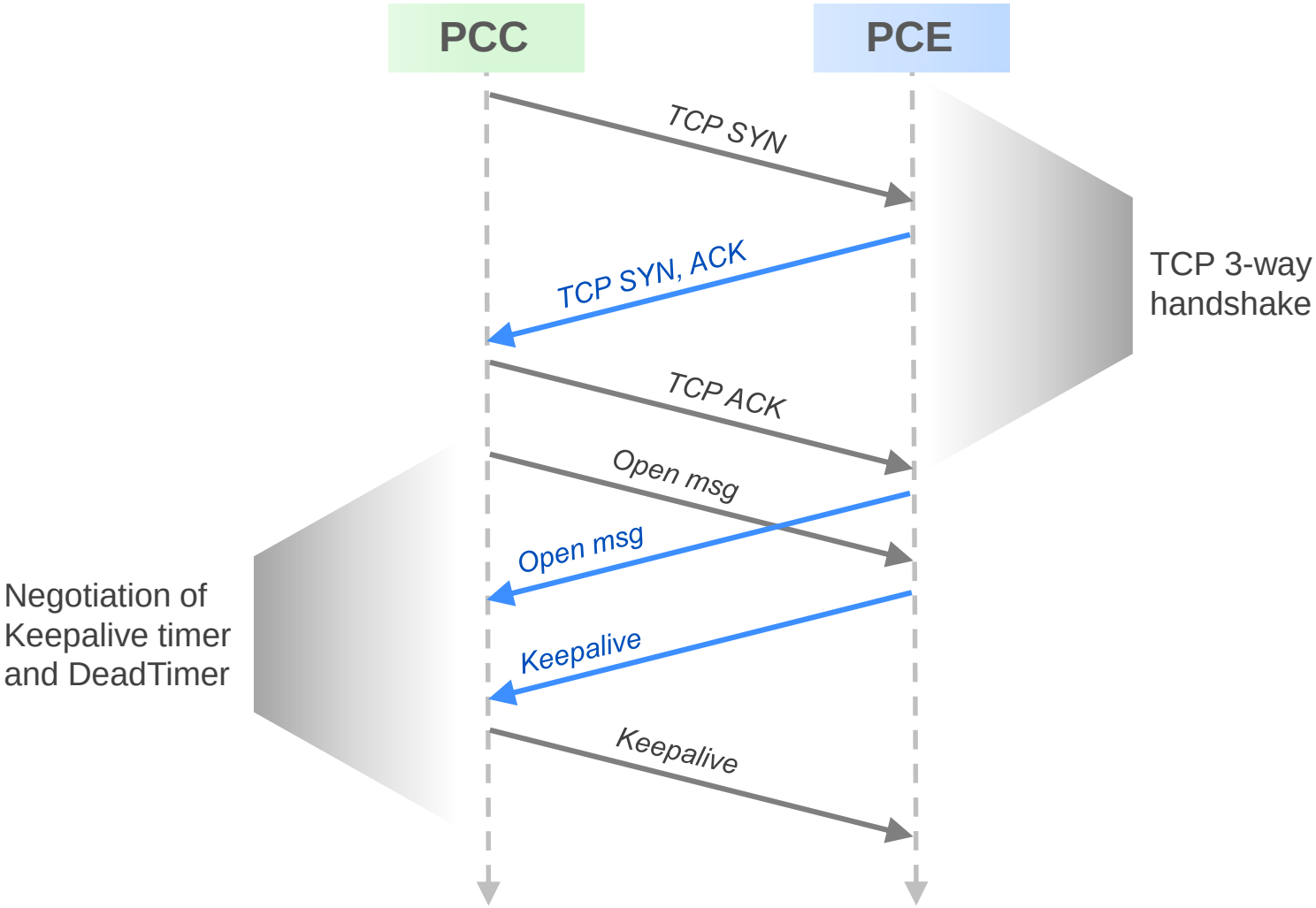
Stateful PCE extensions (draft-ietf-pce-stateful-pce)

Extensions for PCE-initiated LSPs (draft-ietf-pce-pce-initiated-lsp)

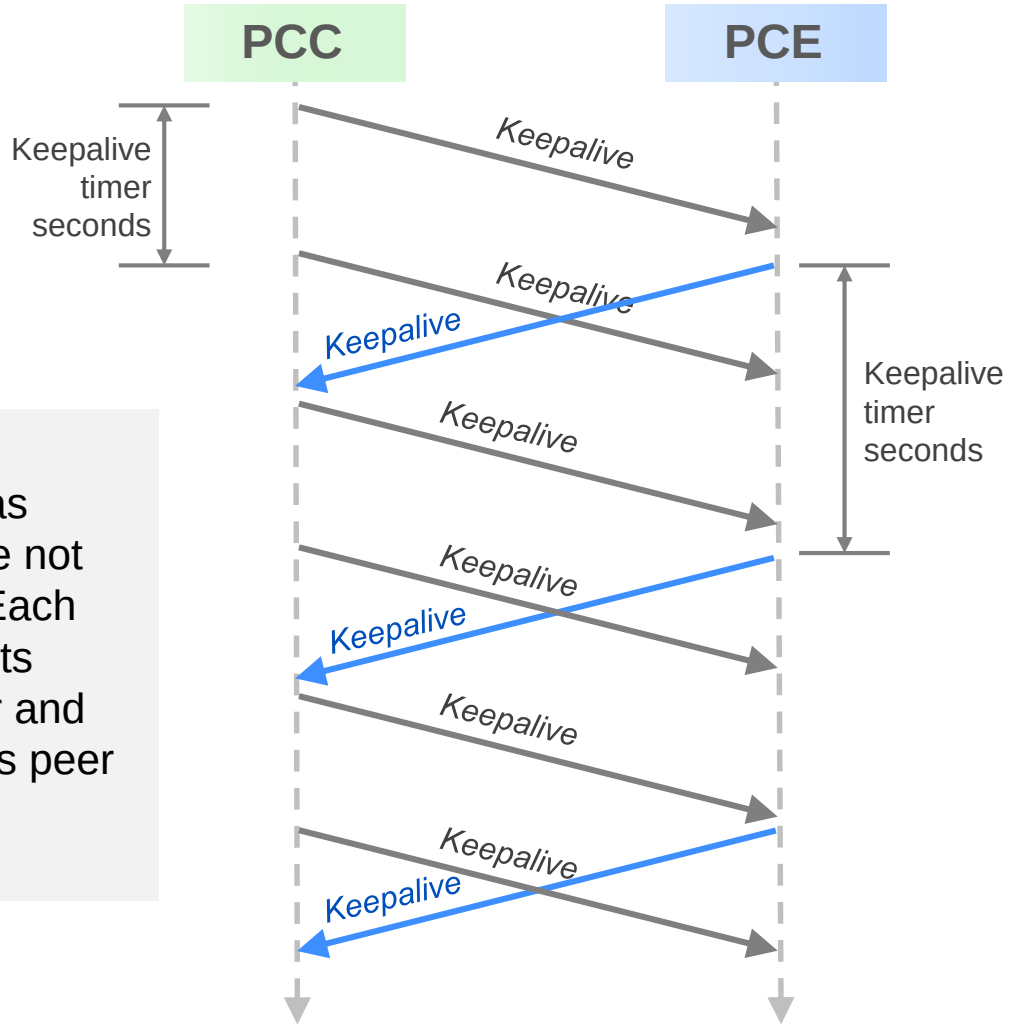
PCEP Messages (2)

Value	Type	Short name	Purpose
1	Open		Initiate PCEP session
2	Keepalive		Maintain PCEP session
3	Path Computation Request	PCReq	Sent by PCC to PCE to request a path computation
4	Path Computation Reply	PCRep	Send by PCE to PCC in response to a PCReq
5	Notification	PCNtf	Sent by either side to notify of a specific event
6	Error		Sent on occurrence of an error
7	Close		Used to close a session
8	Path Computation Monitoring Request	PCMonReq	Sent by a PCC to a PCE to get PCE state metrics
9	Path Computation Monitoring Reply	PCMonRep	Sent by a PCE to a PCE in response to a PCMonReq
10	LSP State Report	PCRpt	Sent by a PCC to a PCE to report the current state of an LSP
11	LSP Update Request	PCUpd	Sent by a PCE to a PCC to update the attributes of an LSP
12	Initiate	PCInitiate	Sent by a PCE to a PCC to trigger LSP creation or deletion

Operation: Initialisation

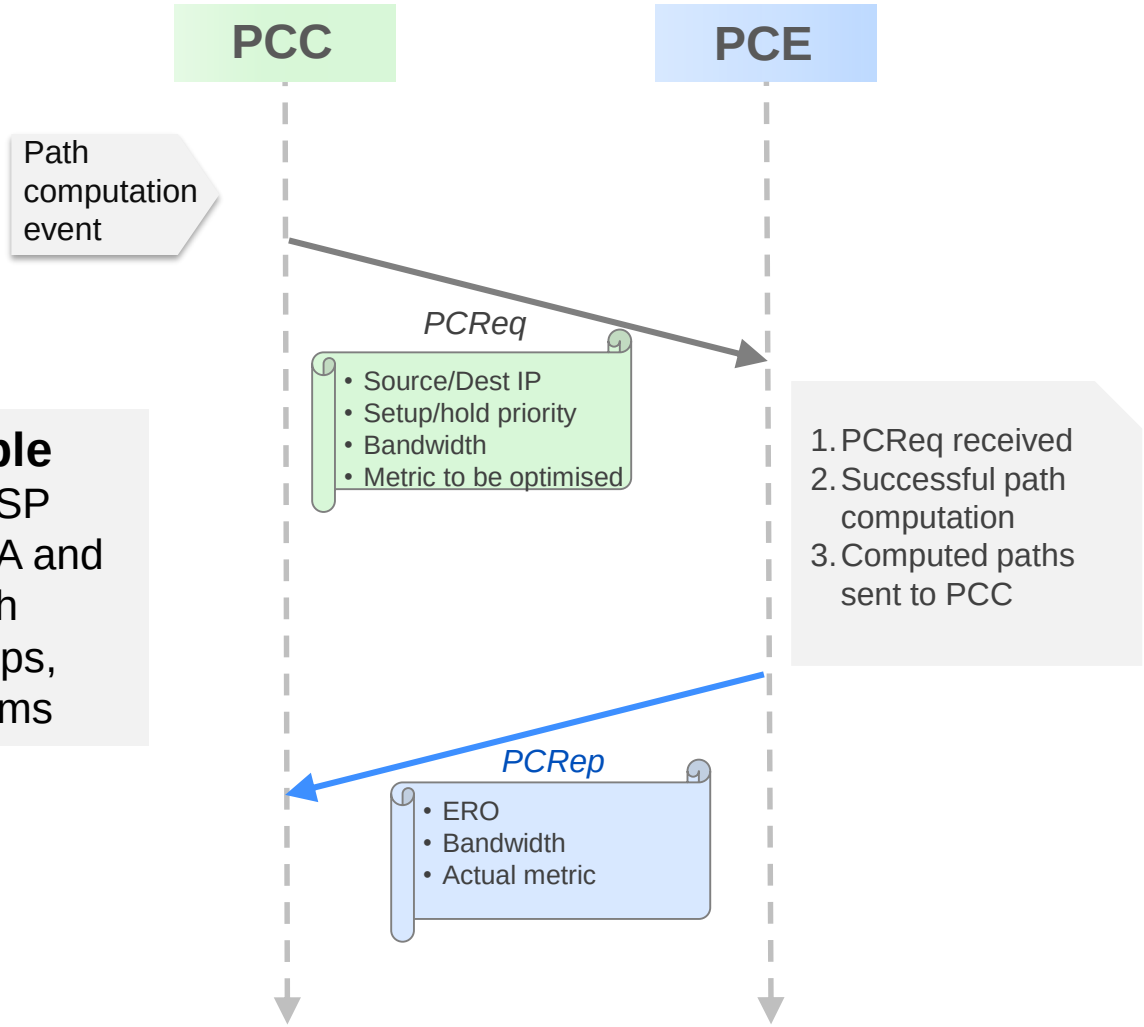


Operation: Keepalives



Keepalives
Keepalives act as beacons and are not responded to. Each side advertises its Keepaliver timer and the Deadtimer its peer should use.
Default: 30s

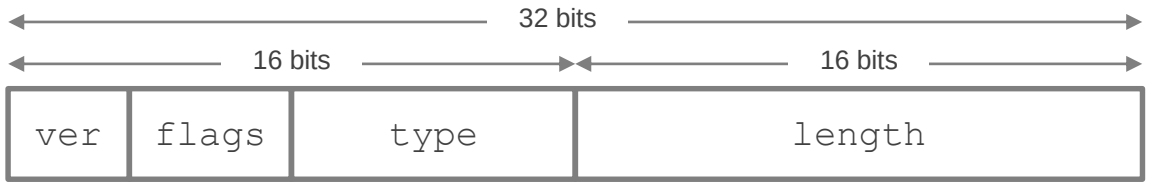
Operation: Path computation



PCReq Example
Compute a TE LSP between source A and destination B with bandwidth=B Gbps, max latency=80 ms

Header format

- All PCEP messages use the following common header followed by a set of message-specific objects

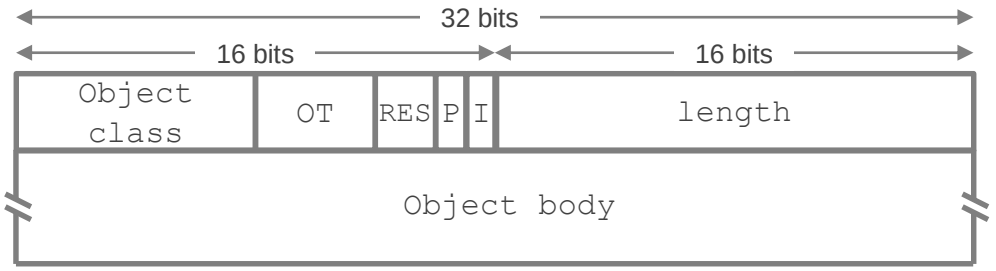



```
version:
  - current version is
    version 1

message type:
  - as per previous table
```

Object format

- All PCEP objects messages use the following common object header followed by object-specific fields



Object class:
- Identifies the PCEP object class

OT:
- Object type
- Combination of object class and OT uniquely identifies each PCEP object

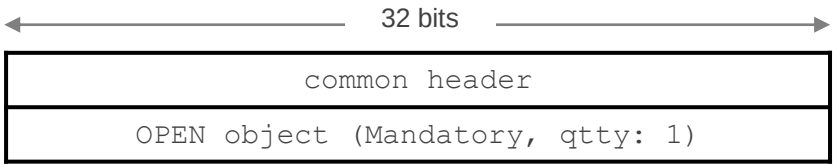
P-flag:
- Set for objects that must be considered in path computation

I-flag:
- Set for optional objects that were ignored during path computation

Message formats (1)

Open Message

Used to establish a PCEP session



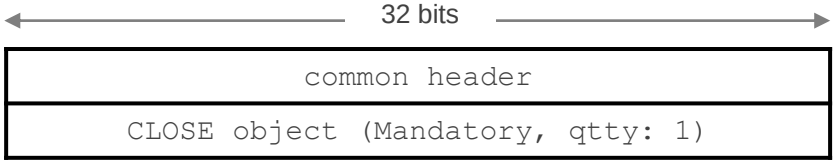
- Keepalive
- Deadtimer
- PCEP session identifier

Keepalive Message

Used to keep the PCEP session active

Close Message

Used to terminate a PCEP session

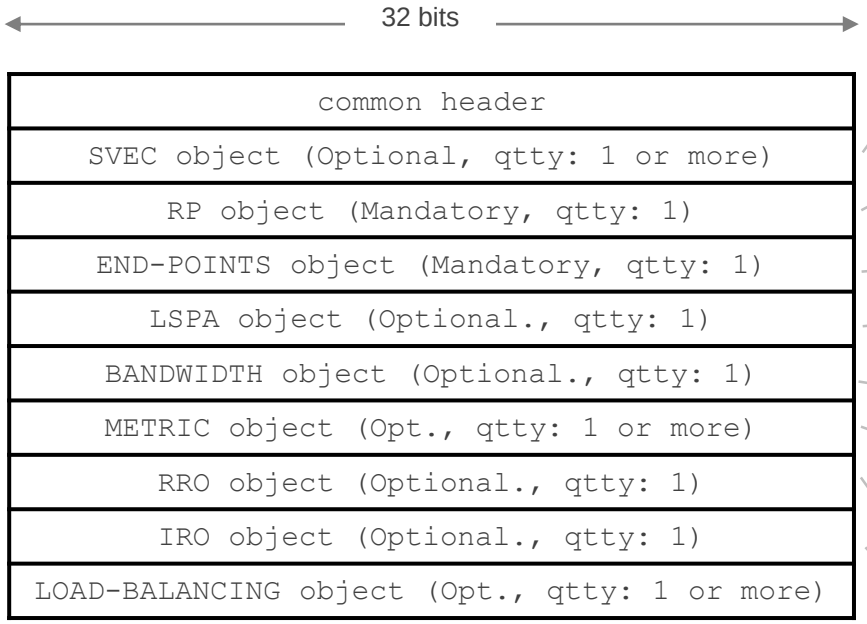


- Reason for closing the PCEP session

Message formats (2)

PCReq Message

- *Used to request a path computation*
- *May carry more than one request*

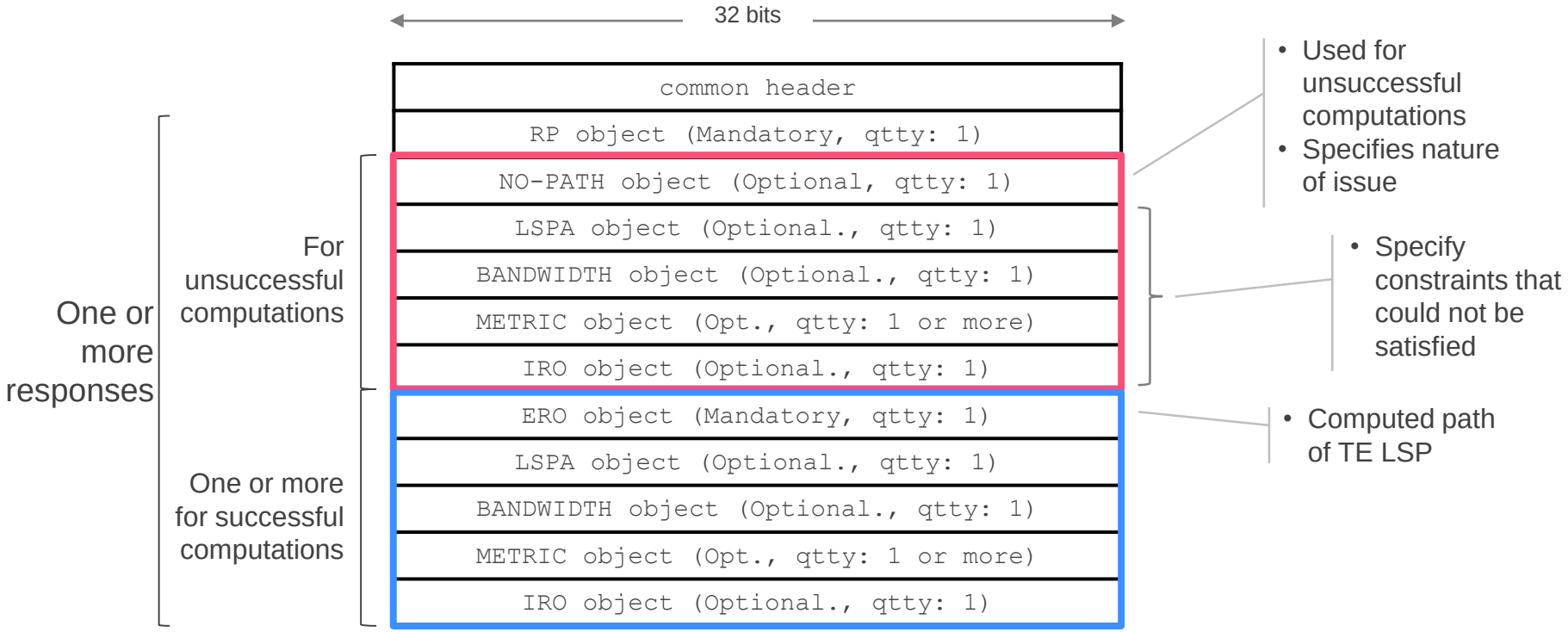


- Sychronisation Vector (SVEC) object
  - Lists a set of request IDs that must be synchronised
  - Can indicate dependency such as:
    - Link diverse
    - Node diverse
    - SRLG diverse
- Request Parameters (RP):
  - Priority
  - Re-optimisation flag
  - Request ID
- Source and destination IP address of path
- LSP attributes to be considered
  - Setup/hold priority
- Requested bandwidth
- Indicates metric to be optimised:
  - IGP metric
  - TE metric
  - Hop count
- Maximum bound on path cost
- Reported Route Object for re-optimisation
- Include Route Object for element that must be included in path
- Allows use of set of TE LSPs to satisfy a large bandwidth demand

Message formats (3)

PCRep Message

- *Sent by PCE to PCC in response to a PCReq*
- *May carry more than one reply*



PCEP extensions for stateful PCE

Objectives

- Allow a single PCC to interact with a mix of stateless and stateless PCEs using the same PCEP protocol

- Provide a mechanism for state synchronisation between the PCC and a stateful PCE

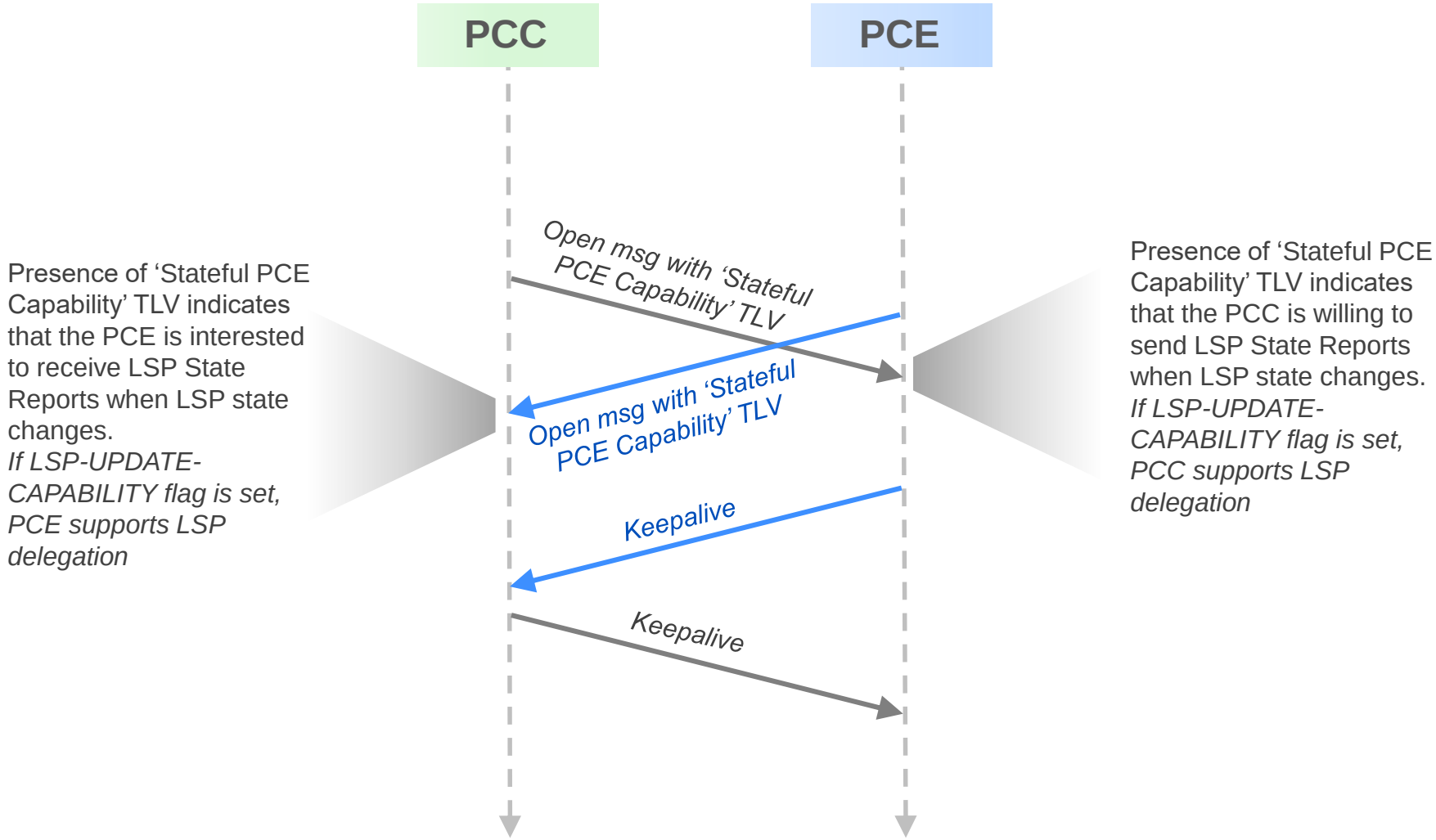
- Allow a PCC to delegate control of LSPs to an active stateful PCE

New messages

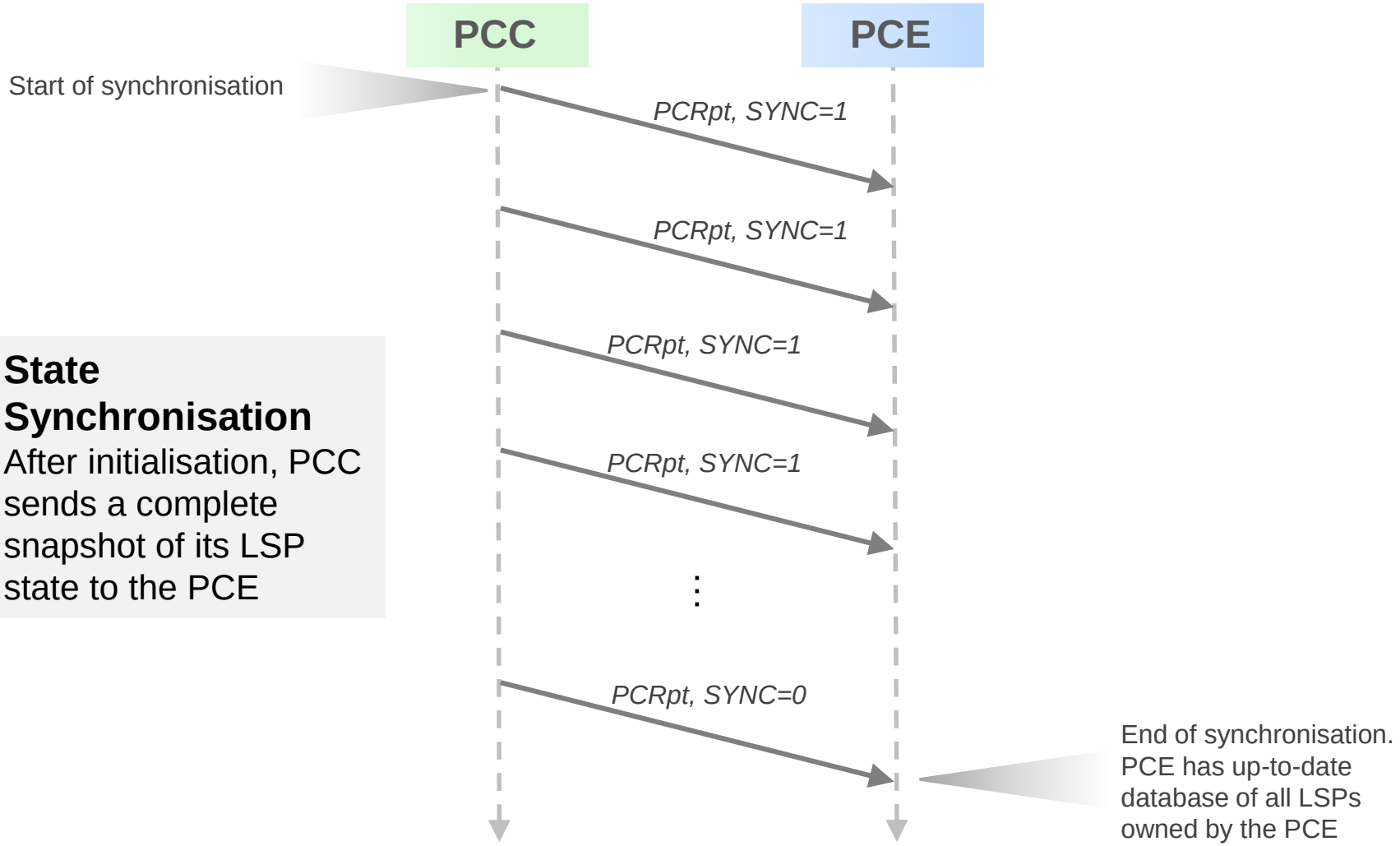
- Path Computation State Report (PCRpt)
  - sent by a PCC to a PCE to report on the status of one or more LSPs
  - includes the LSP's path, bandwidth, operational status, administrative status
  - also used for delegating control of an LSP to a PCE
  - sent:
    - whenever the state of an LSP changes
    - in response to an LSP Update Request from a PCE

- Path Computation Update Request (PCUpd)
  - sent by a PCE to a PCC to update parameters for one or more LSPs

Operation: capability advertisement

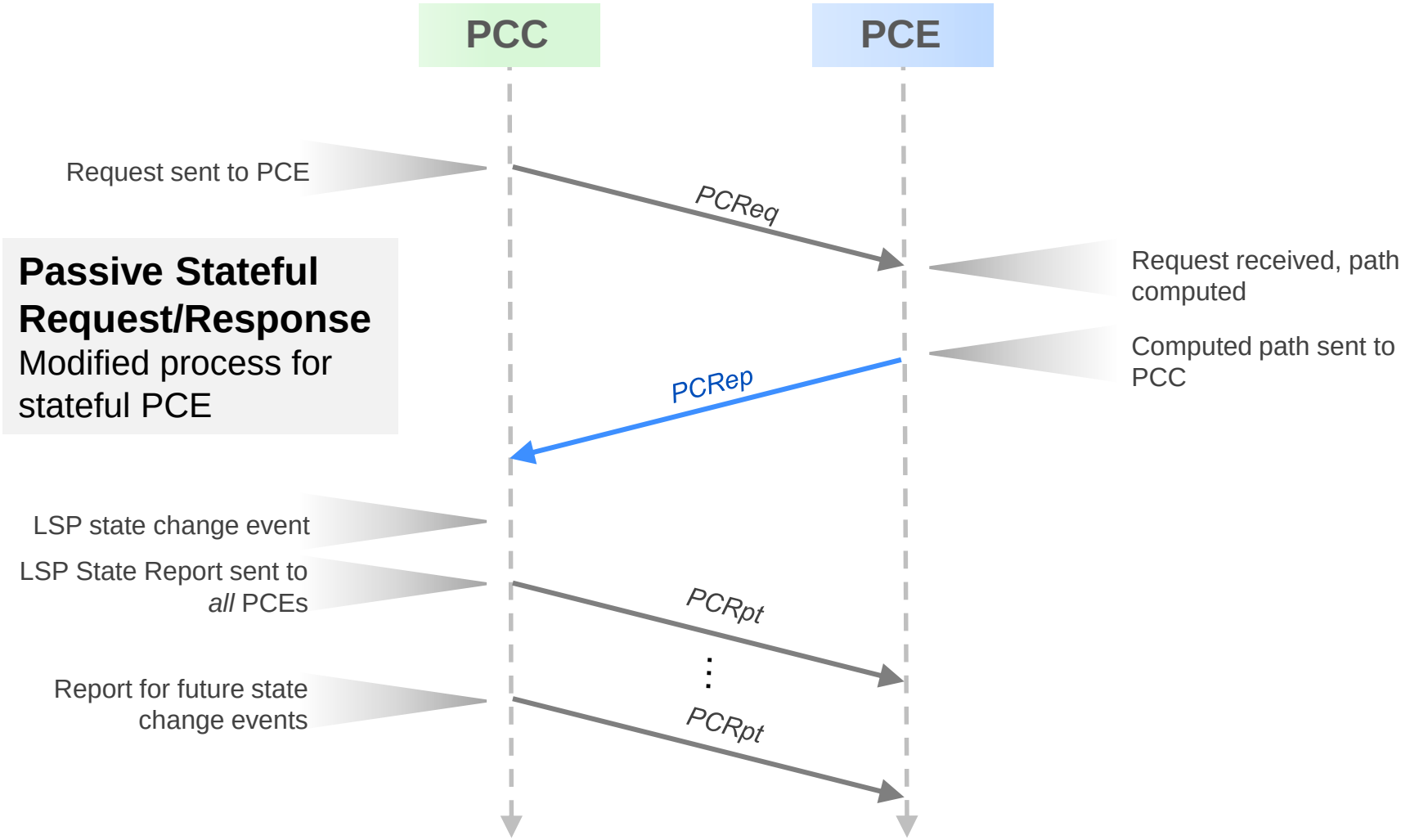


Operation: state synchronisation



State Synchronisation
After initialisation, PCC sends a complete snapshot of its LSP state to the PCE

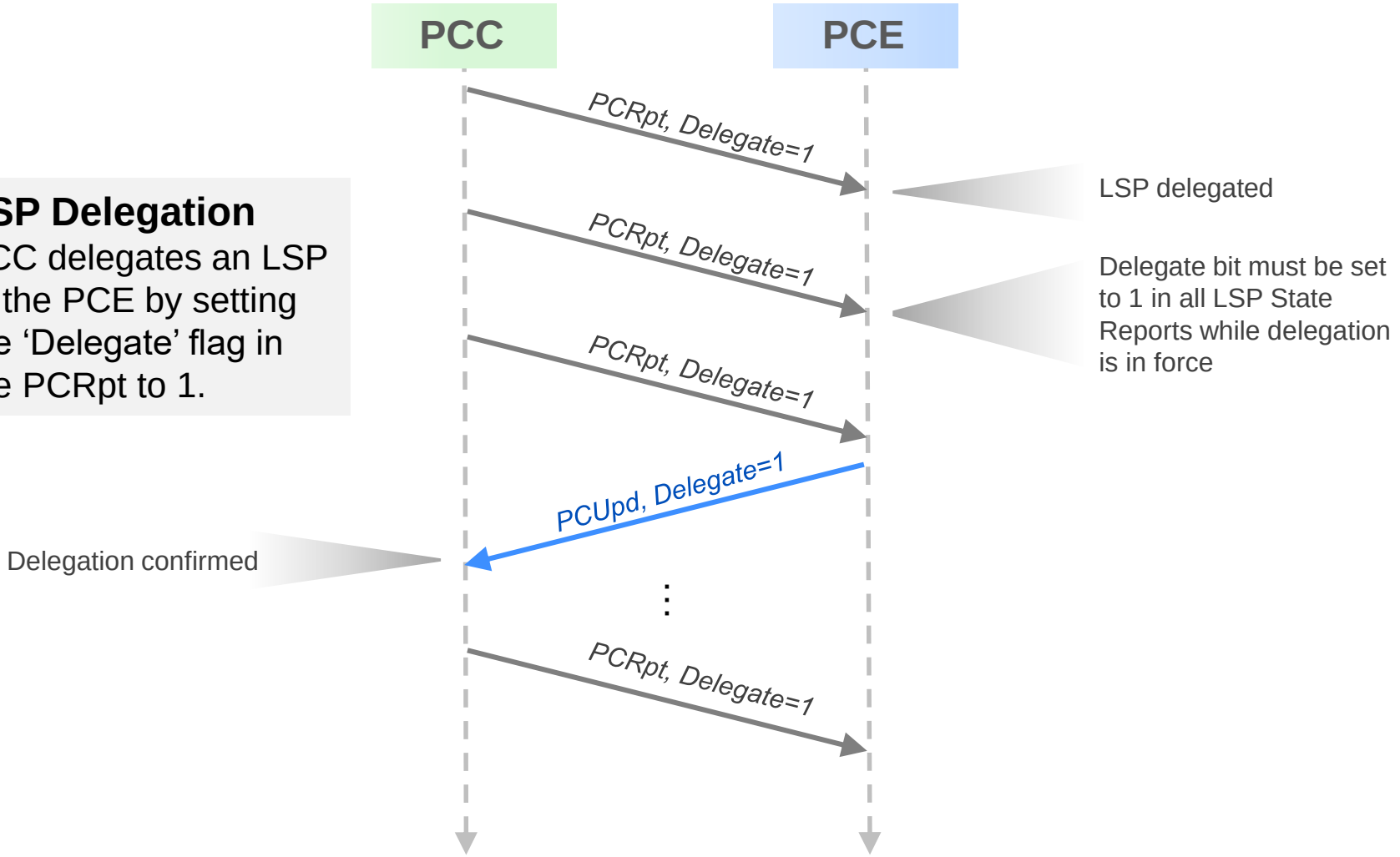
Operation: passive stateful req/rep



Passive Stateful Request/Response
Modified process for stateful PCE

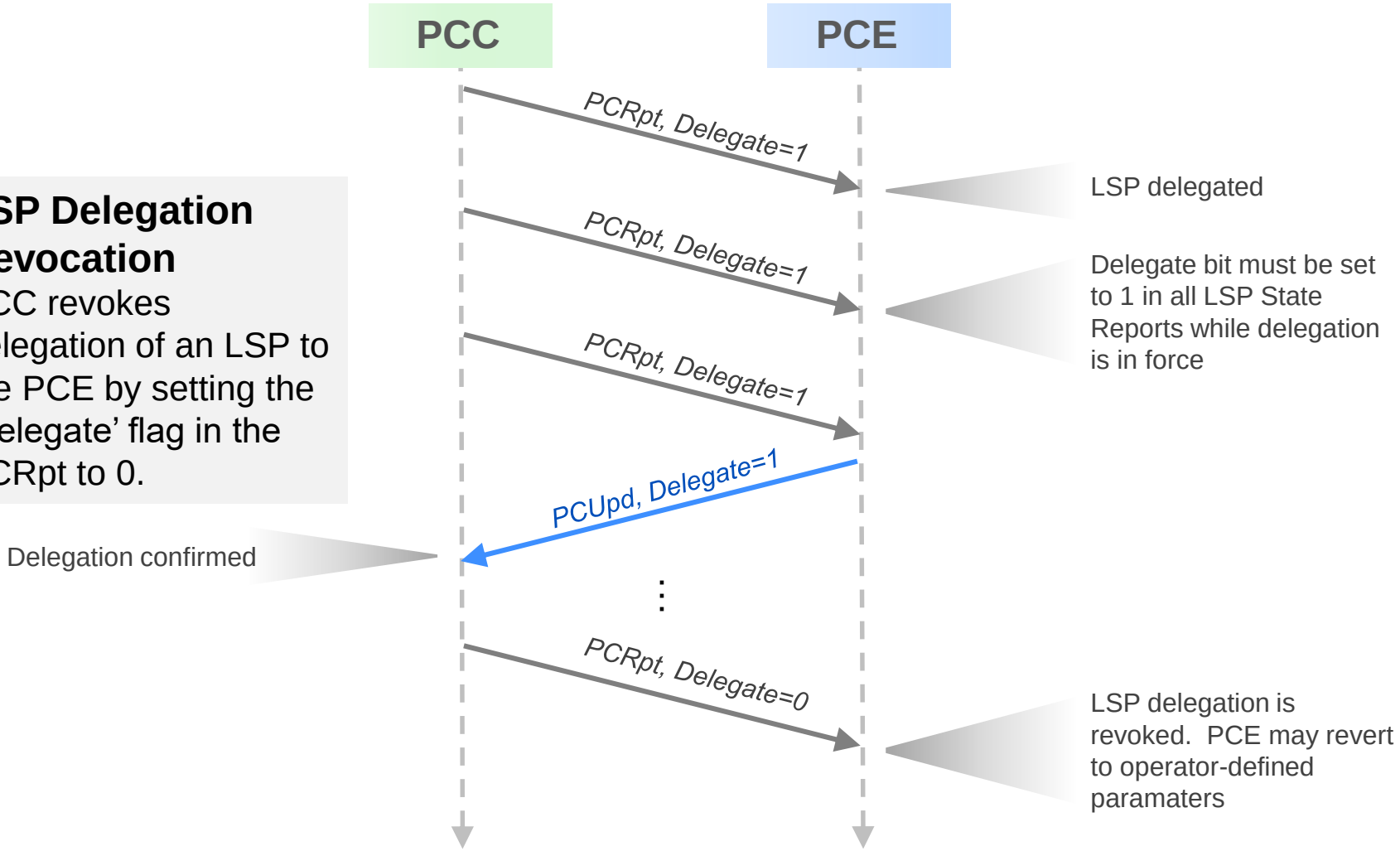
Operation: LSP delegation

LSP Delegation
PCC delegates an LSP to the PCE by setting the ‘Delegate’ flag in the PCRpt to 1.



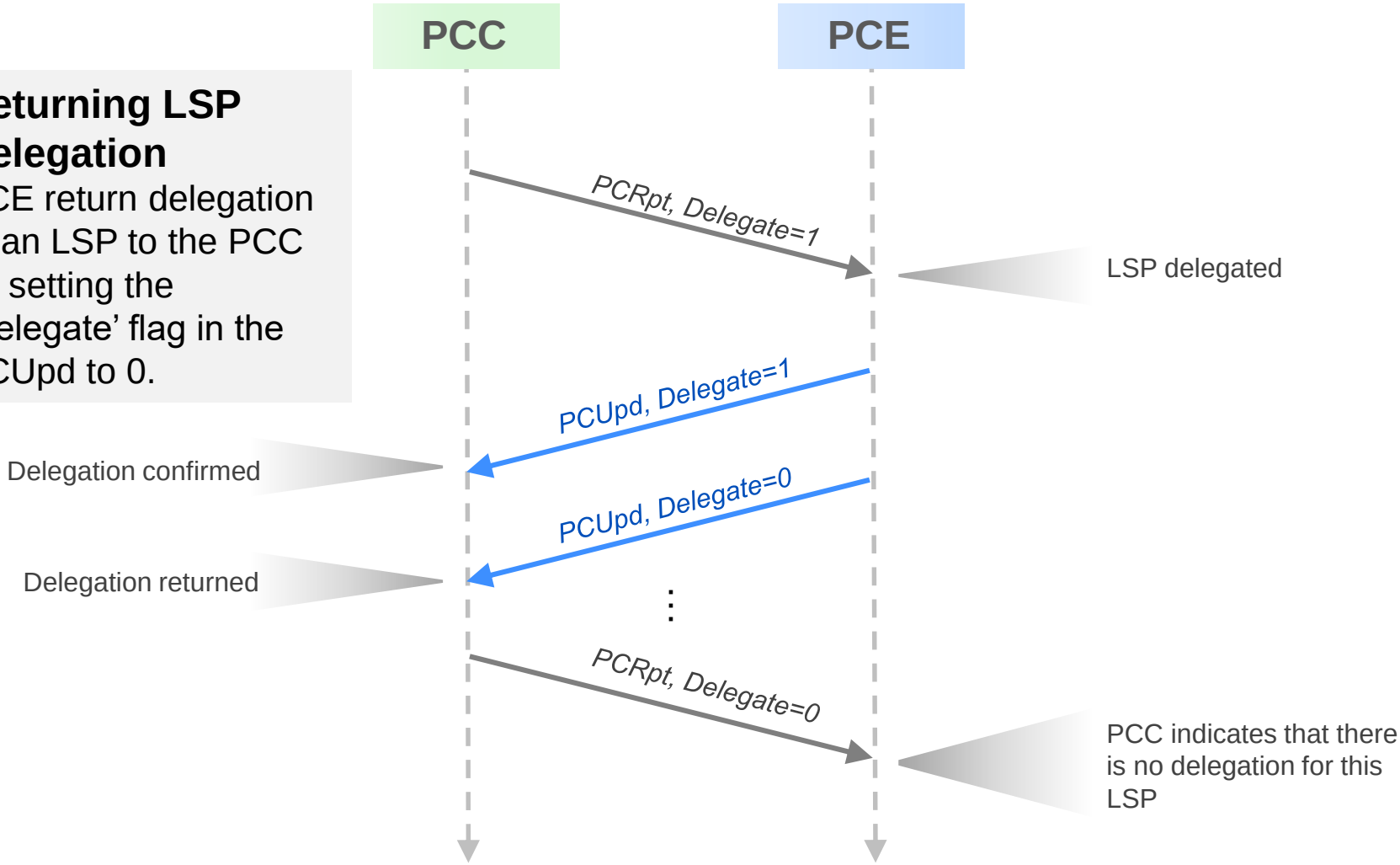
Operation: LSP delegation revocation

LSP Delegation Revocation
PCC revokes delegation of an LSP to the PCE by setting the 'Delegate' flag in the PCRpt to 0.

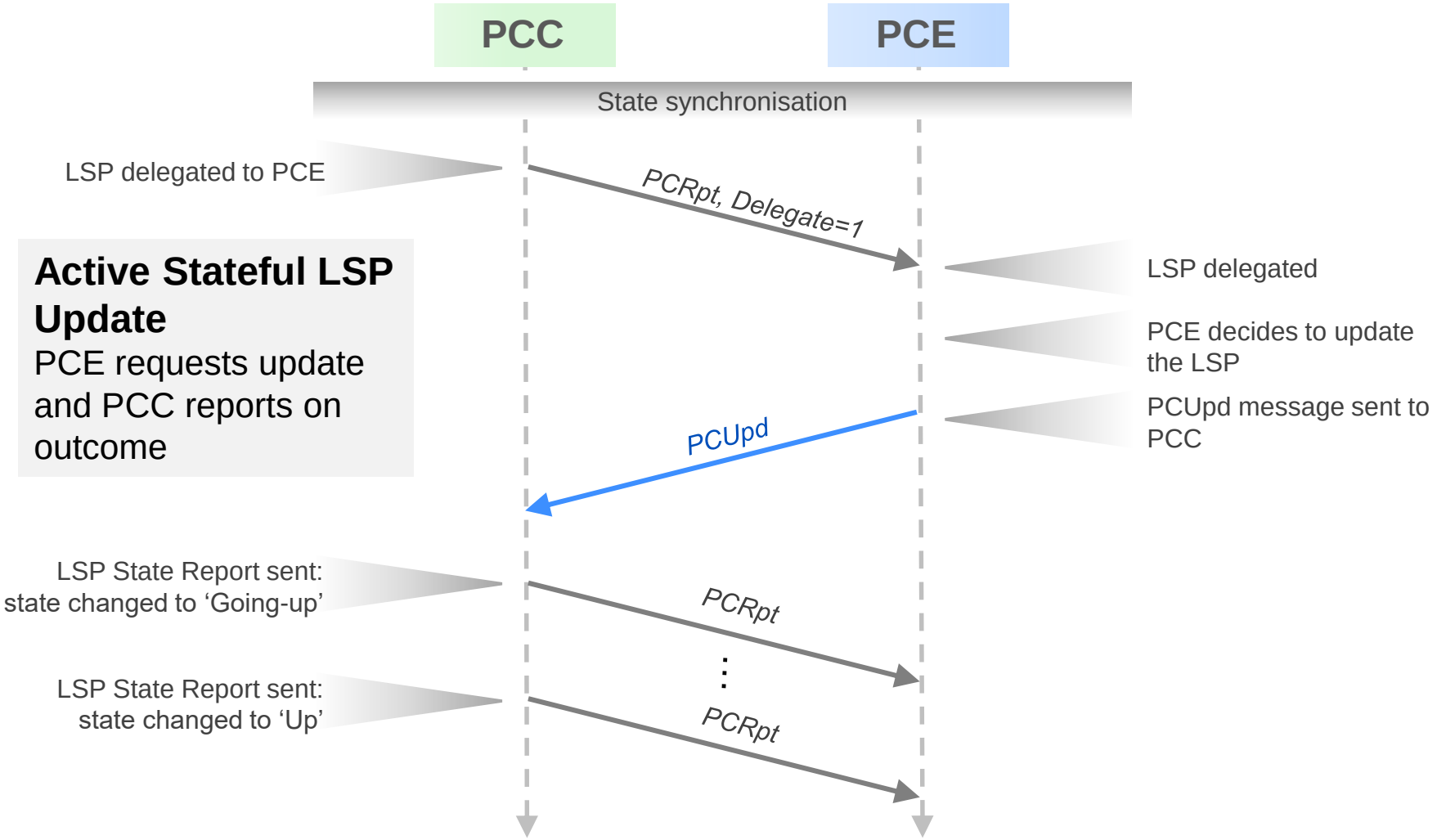


Operation: returning LSP delegation

Returning LSP Delegation
PCE return delegation of an LSP to the PCC by setting the 'Delegate' flag in the PCUpd to 0.



Operation: active stateful LSP update



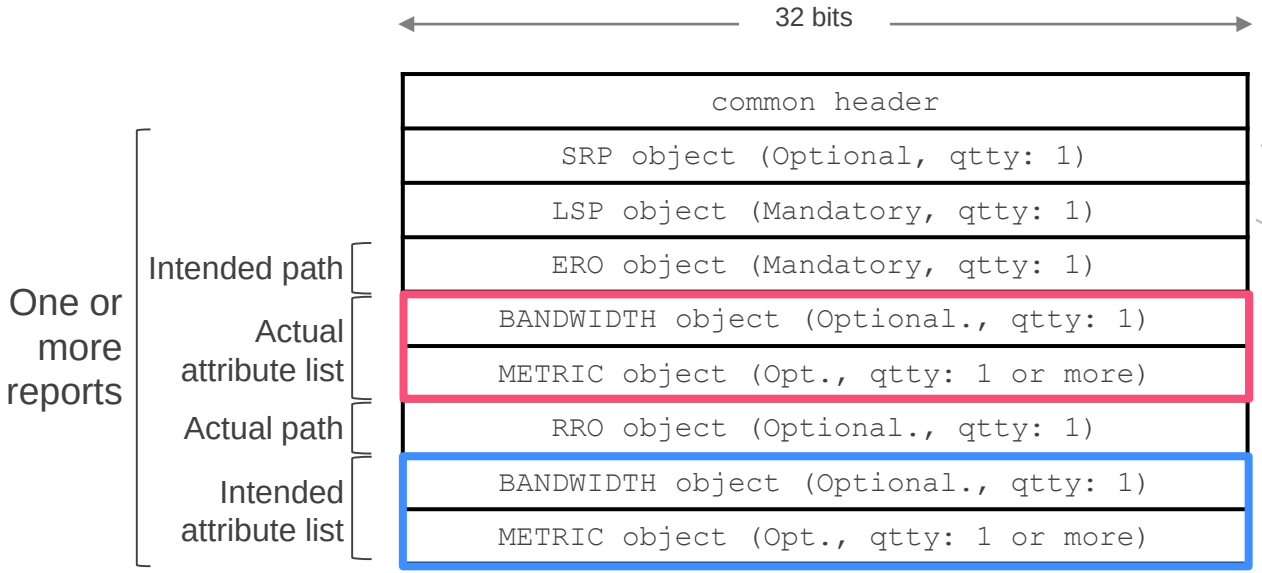
Active Stateful LSP Update

PCE requests update and PCC reports on outcome

Message formats (1)

PCRpt Message

- *Used by PCC to report state of LSPs to PCEs*
- *May carry more than one request*



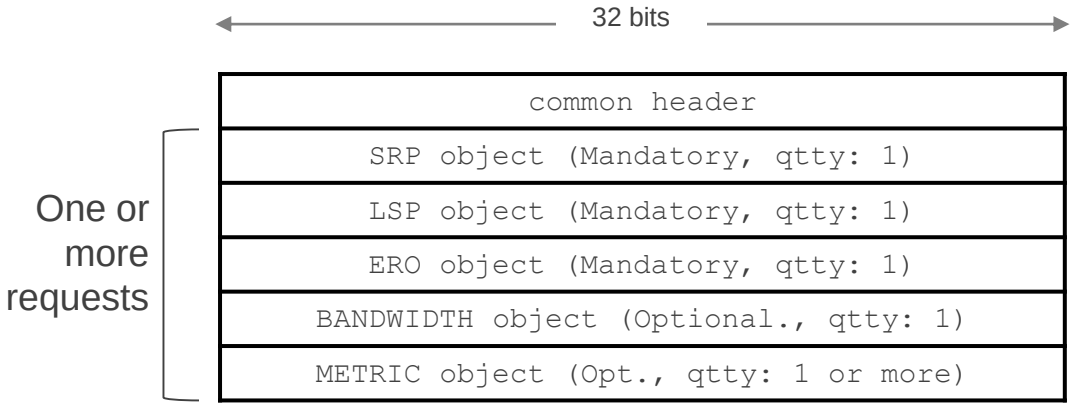
- Stateful PCE Request Parameters (SRP):
  - Required for PCRpts sent in response to PCUpd
  - Carries SRP-ID: uniquely identifies PCE update request for an LSP
  - Used to correlate between PCUpd request and corresponding State Reports

- LSP object:
  - PLSP-ID: Unique PCEP-specific LSP identifier allocated by PCC
  - Flags:
    - Delegate flag
    - Sync flag
    - Remove flag: used by PCC to notify PCE that LSP has been removed
    - Operational state: down, up, active, going-down, going-up
  - LSP Identifiers TLV (for RSVP-TE):
    - Tunnel Sender Address
    - LSP ID
    - [Extended] Tunnel ID
    - Tunnel Endpoint Address
  - Symbolic Path Name TLV

Message formats (2)

PCUpd Message

- *Used by PCE to request update of LSP parameters*
- *May carry more than one request*



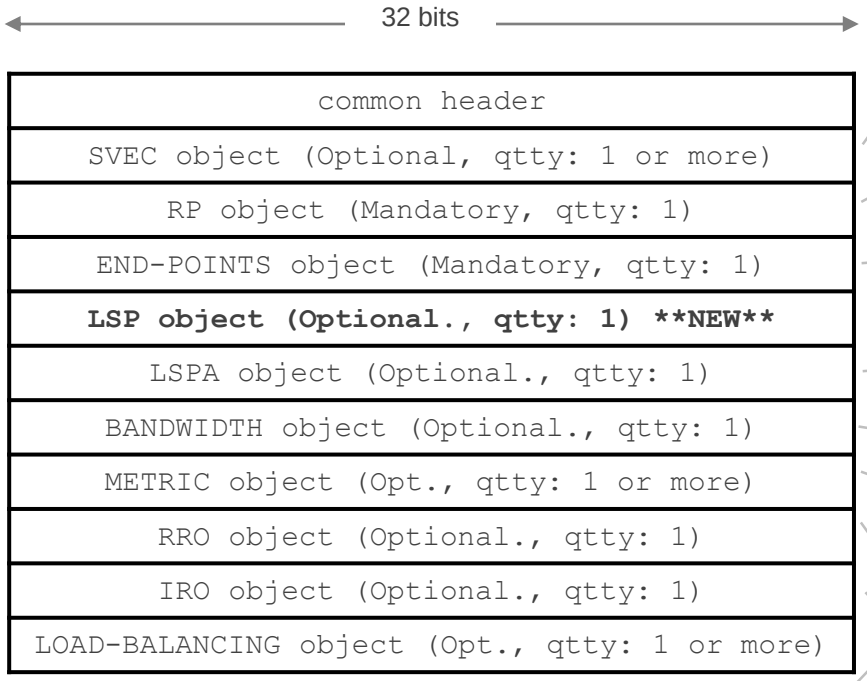
- Stateful PCE Request Parameters (SRP):
  - Carries SRP-ID: uniquely identifies PCE update request for an LSP
  - Used to correlate between PCUpd request and corresponding State Reports

- Explicit Route Object (ERO)
- Intended path of TE LSP

Message formats (3)

PCReq Message

- *Used to request a path computation*
- *May carry more than one request*



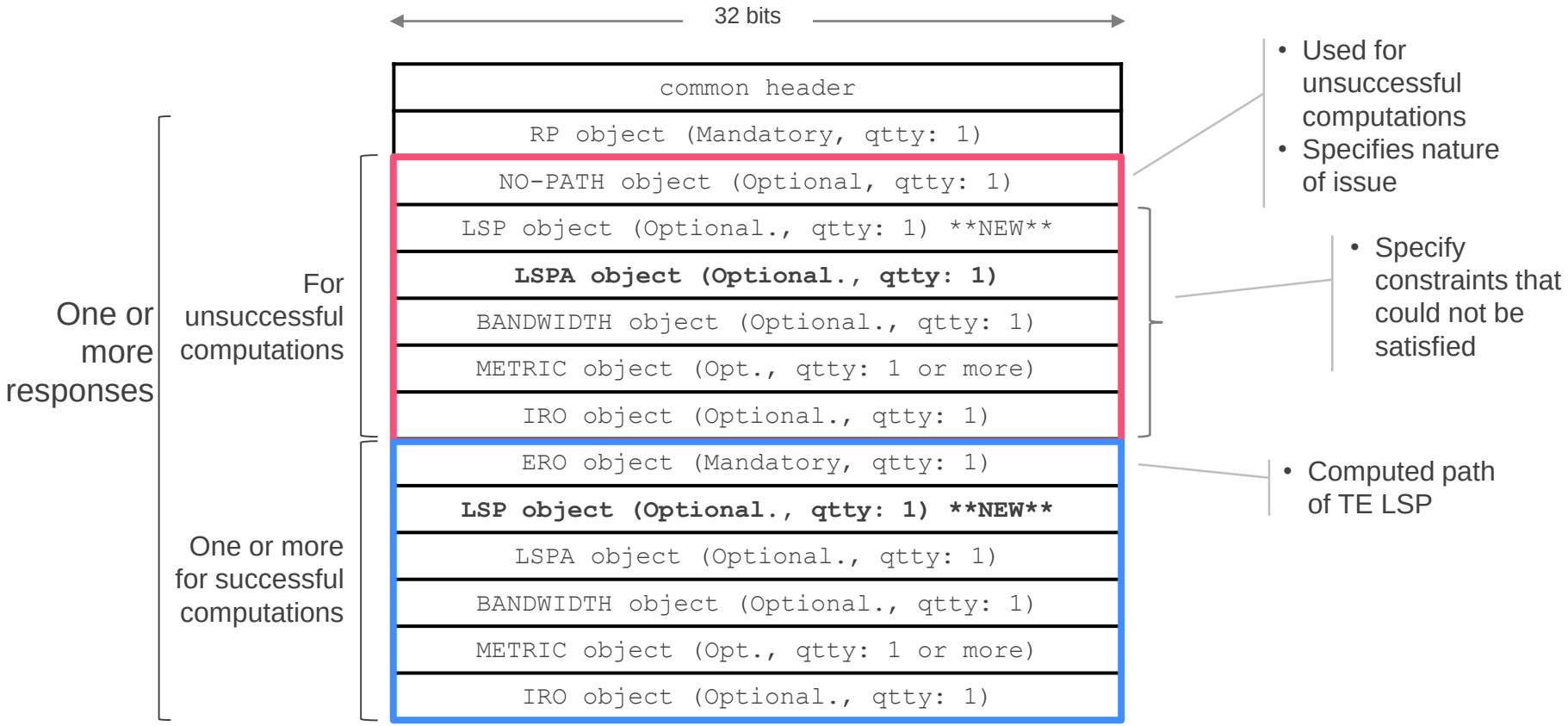
One or more requests

- Sychronisation Vector (SVEC) object
  - Lists a set of request IDs that must be synchronised
  - Can indicate dependency such as:
    - Link diverse
    - Node diverse
    - SRLG diverse
- Request Parameters (RP):
  - Priority
  - Re-optimisation flag
  - Request ID
- Source and destination IP address of path
- LSP attributes to be considered
  - Setup/hold priority
- Requested bandwidth
- Indicates metric to be optimised:
  - IGP metric
  - TE metric
  - Hop count
- Maximum bound on path cost
- Reported Route Object for re-optimisation
- Include Route Object for element that must be included in path
- Allows use of set of TE LSPs to satisfy a large bandwidth demand

Message formats (4)

PCRep Message

- *Sent by PCE to PCC in response to a PCReq*
- *May carry more than one reply*



PCEP extensions for PCE-initiated LSPs

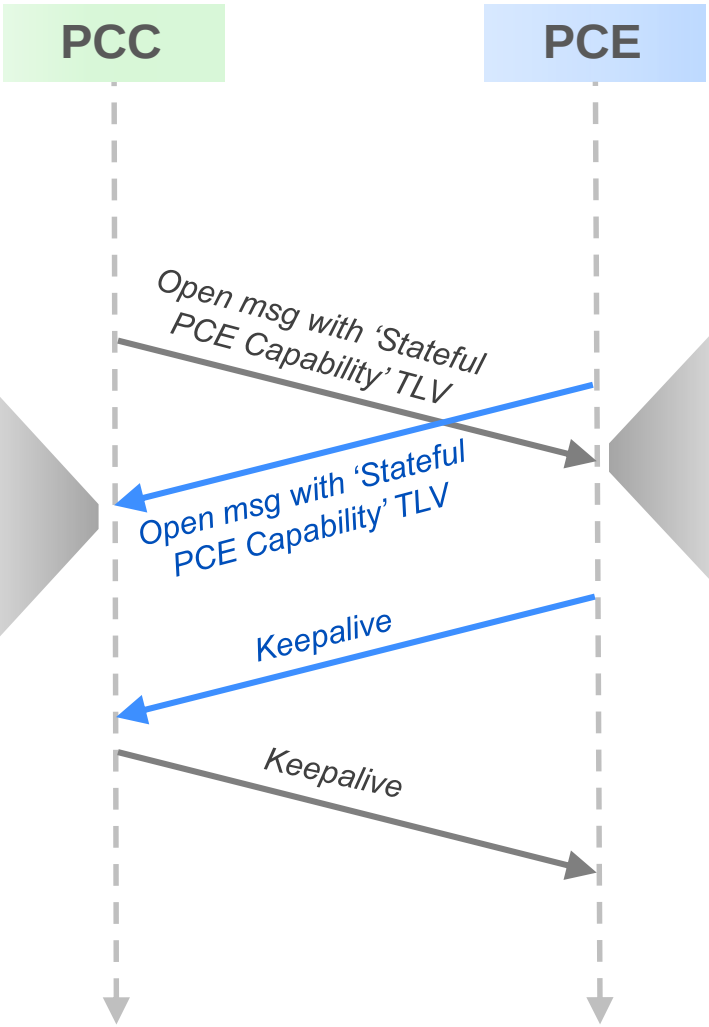
APNIC

New messages

- LSP Initiate Request (PCInitiate)
  - sent by a PCE to a PCC to request the instantiation (or deletion) of an LSP
  - the PCC:
    - creates the LSP using supplied attributes
    - generates an LSP State Report (PCRpt) for the LSP with a newly assigned PLSP-ID
    - sets the delegate bit in the LSP State Report

- The PCE is able to update parameters for the LSP by sending PCUPd messages

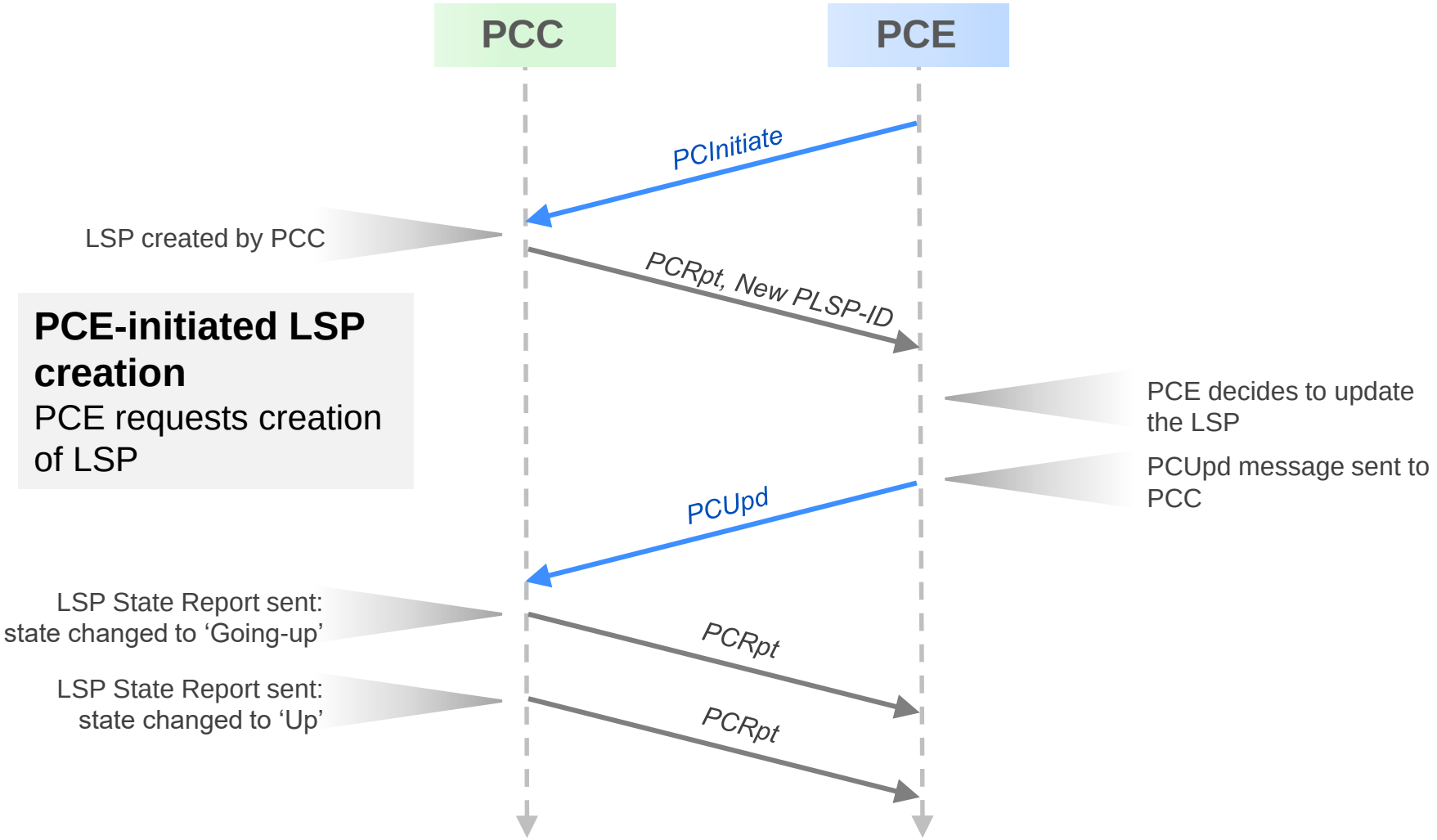
Operation: capability advertisement



Presence of 'Stateful PCE Capability' TLV indicates that the PCE is interested to receive LSP State Reports when LSP state changes.
The LSP-INSTANTIATION-CAPABILITY flag (I- flag) is set to indicate support of PCE-initiated LSPs.
The *LSP-UPDATE-CAPABILITY* flag has to be set as well.

Presence of 'Stateful PCE Capability' TLV indicates that the PCC is willing to send LSP State Reports when LSP state changes.
The LSP-INSTANTIATION-CAPABILITY flag (I- flag) is set to indicate support of PCE-initiated LSPs.
The *LSP-UPDATE-CAPABILITY* flag has to be set as well.

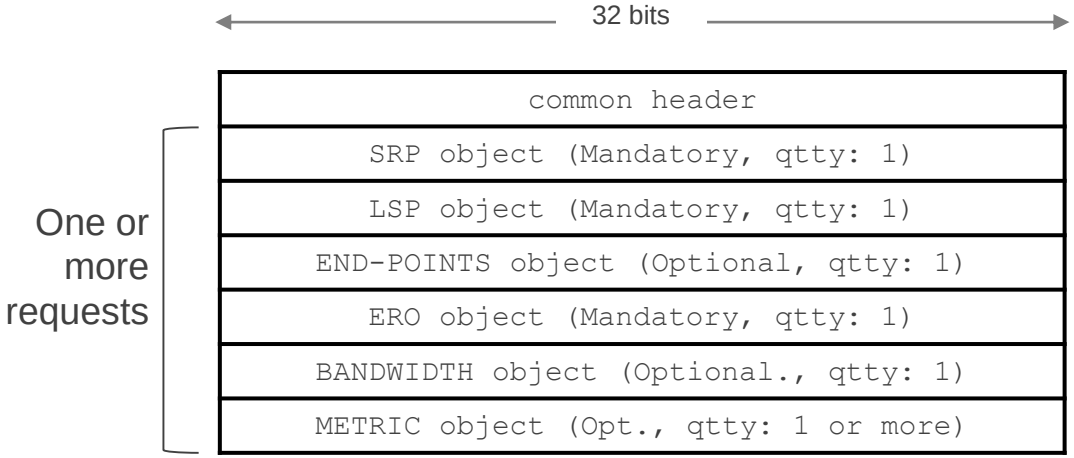
Operation: PCE-initiated LSP creation



Message formats (1)

PCInitiate Message

- *Used by PCE to request creation of an LSP by a PCC*
- *May carry more than one request*

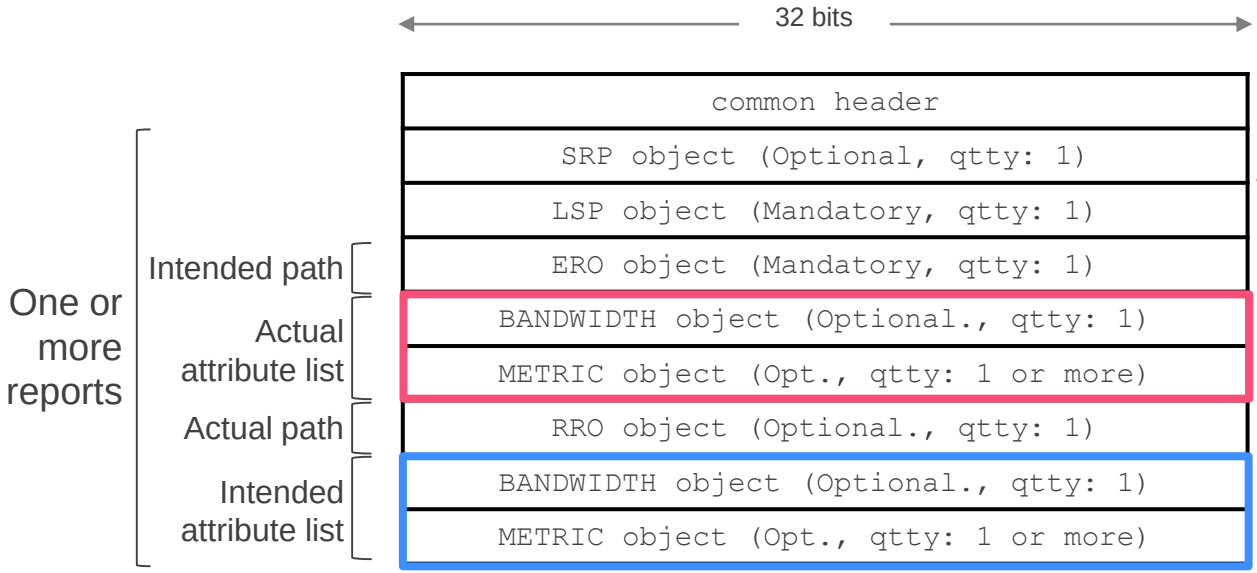


- Stateful PCE Request Parameters (SRP):
  - Carries SRP-ID: uniquely identifies PCE initiate request for an LSP
  - Used to correlate between PCInitiate request and corresponding State Reports
- LSP object:
  - PLSP-ID: set to 0
  - Flags:
    - Delegate flag
- Explicit Route Object (ERO)
- Intended path of TE LSP

Message formats (2)

PCRpt Message

- *Used by PCC to report state of LSPs to PCEs*
- *May carry more than one request*



- Stateful PCE Request Parameters (SRP):
  - Required for PCRpts sent in response to PCUpd/PCInitiate
  - Carries SRP-ID: uniquely identifies PCE update request for an LSP
  - Used to correlate between PCUpd/PCInitiate request and corresponding State Reports

- LSP object:
  - PLSP-ID: Unique PCEP-specific LSP identifier allocated by PCC
  - Flags:
    - Delegate flag
    - Sync flag
    - **Create flag: to indicate that the LSP was PCE-initiated.**
    - Remove flag: used by PCC to notify PCE that LSP has been removed
    - Operational state: down, up, active, going-down, going-up
  - LSP Identifiers TLV (for RSVP-TE):
    - Tunnel Sender Address
    - LSP ID
    - [Extended] Tunnel ID
    - Tunnel Endpoint Address
  - Symbolic Path Name TLV
  - Speaker Entity ID TLV: identifies PCE which initiated LSP

Extensions for Segment Routing

APNIC

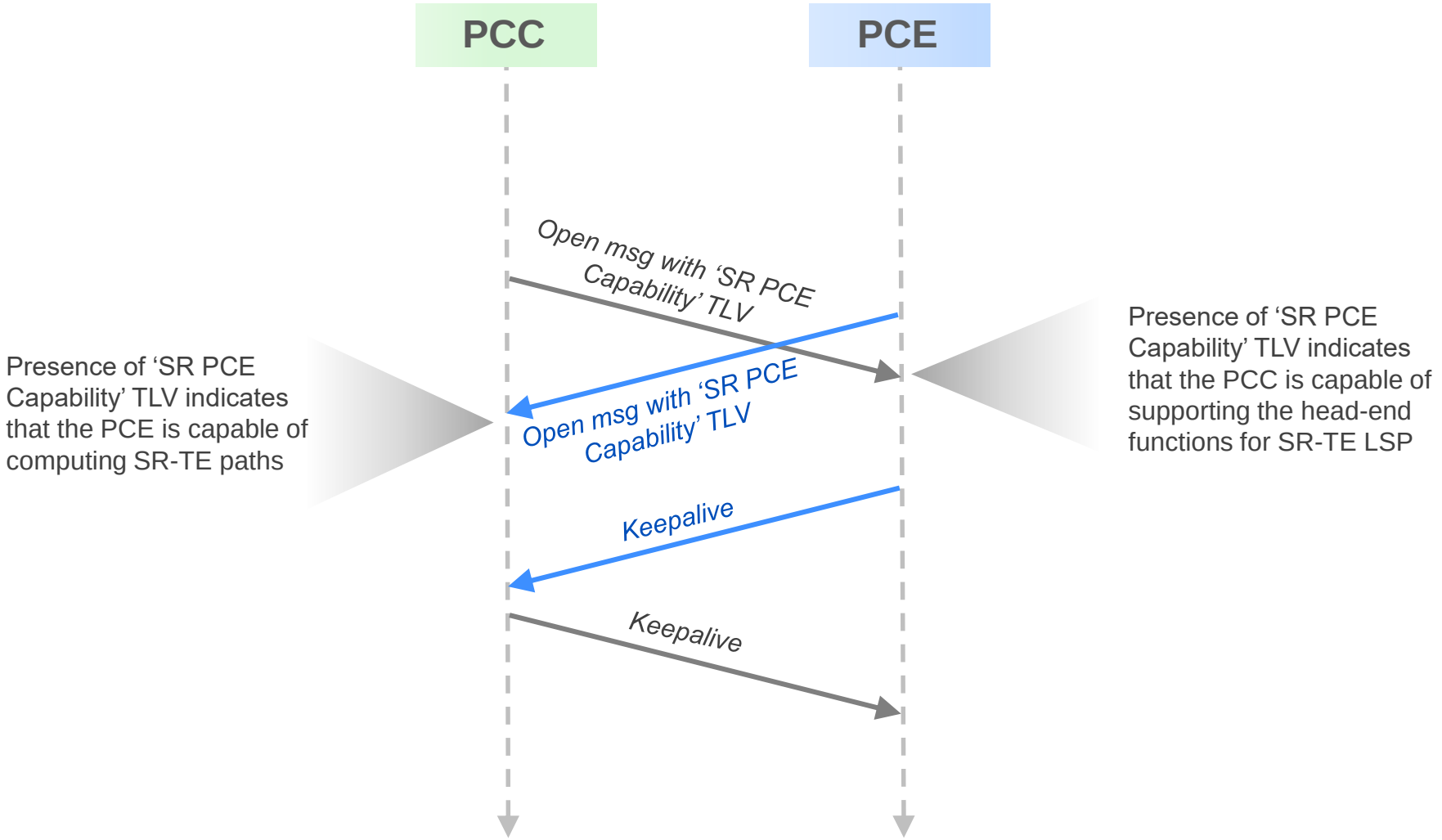
Building the Segment List

- In a Segment Routing network, the ingress LER has to impose a label stack on packets corresponding to the required TE path.
- How can the ingress LER derive this information ?
- One solution is to use an external PCE.

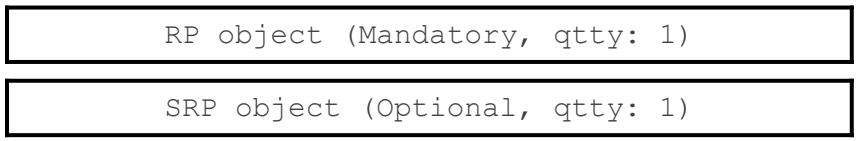
SR-TE LSP Paths

- SR-TE LSPs computed by a PCE can take one of the following forms:
  - Ordered set of IP addresses representing segments (links or nodes); in this case, the PCC would have to translate these to the appropriate MPLS labels
  - Ordered set of SIDs
  - An ordered set including both MPLS labels and IP addresses in which case the IP addresses needs to be translated to MPLS labels.

Operation: capability advertisement



Message extensions



- [Stateful PCE] Request Parameters ([S]RP):
  - Includes PATH-SETUP-TYPE TLV that indicates PST=1 (Segment Routing)

- Explicit Route Object
  - Carries new SR-ERO sub-object:
    - SID <-> NAI (Node of Adjacency Identifier)
    - Indication of loose or strict hop
    - Flags to indicate whether a complete MPLS label stack is provided or just the MPLS label
    - NAI can be an IPv4/v6 Node ID or IPv4/v4 Adjacency ID

- Reported Route Object
  - Carries new SR-RRO sub-object:
    - SID <-> NAI (Node of Adjacency Identifier)
    - Flags to indicate whether a complete MPLS label stack is provided or just the MPLS label
    - NAI can be an IPv4/v6 Node ID or IPv4/v4 Adjacency ID

Operational Considerations

PCE discovery

- PCEs may be discovered by a PCC either via:
  - Local configuration
  - IGP extensions

References

References (1)

- RFC4105: Requirements for Inter-Area MPLS Traffic Engineering
- RFC4655: A Path Computation Element (PCE)-Based Architecture
- RFC5440: Path Computation Element (PCE) Communication Protocol (PCEP)
- RFC5152: A Per-Domain Path Computation Method for Establishing Inter-Domain Traffic Engineering (TE) Label Switched Paths (LSPs)
- RFC8051: Applicability of a Stateful Path Computation Element (PCE)
- draft-ietf-pce-stateful-pce: PCEP Extensions for Stateful PCE
- draft-ietf-pce-segment-routing: PCEP Extensions for Segment Routing
- draft-ietf-pce-pce-initiated-lsp: PCEP Extensions for PCE-initiated LSP Setup in a Stateful PCE Model

References (2)

- A Quick Start to Path Computation Element (PCE) | Packet Design: www.packetdesign.com/blog/quick-start-path-computation-element-pce/
- http://www.olddog.co.uk/AdrianFarrel-TheRoleOfPCEInAnSDNWorld.pdf
- http://www.olddog.co.uk/Farrel_PCE_Tutorial.ppt
- http://www.olddog.co.uk/AdvancedPCE.pdf
- http://content.ipspace.net/get/PCEP

Thank You !

End of session