Next Steps in Path Computation Element (PCE) Architectures: From Software Defined Concepts to Standards, Interoperability and Deployment

http://www.ict-pace.net/

@PACEICT

PCE Tutorial
• Rationale and main drivers
  – Storyline and relevant standards

• Architecture overview
  – Models (centralized vs. distributed) and components
  – The PCEP protocol (basics and mechanisms)
  – Multi-domain path computation
    • Hierarchical
    • BRPC
  – The Traffic Engineering Database

• Advanced features
  – Core PCEP extensions
  – Inter-layer path computation
  – Stateful PCE
PCE tutorial

• Usage scenarios
  – PCE in optical networks
  – Global Concurrent Optimization
  – Deployment models in Software Defined Networks
  – PCE in ABNO
  – PCE for intra and inter datacenter orchestration
  – PCE role in Segment Routing
  – PCE in NFV
  – PCE in IoT
Path Computation
Motivations

• There are lots of path computation problems in networking
• Many problems can be solved off-line
  – Service planning
  – Plenty of time
  – Even failure cases can be pre-planned
• Many problems take considerable computation resources
• Increasing demand for on-line or rapid response
  – Not all problems can be solved like this
  – Network nodes (routers and switches)
    • Do not have big CPUs
    • Do not have spare memory
PCE Inception

• All of these problems point to the use of dedicated path computation resources (i.e., servers)
• But PCE was invented for a completely different reason
• Aimed to solve a very specific problem
  – Find an MPLS-TE path to a virtual PoP
  – I can see in my domain, but not into my peer’s
  – Which exit-point should I choose?
The PCE – A short history

• PCE: Path Computation Element - “An entity (component, application, or network node) that is capable of computing a network path or route based on a network graph and applying computational constraints.” from RFC 4655

• That means that a PCE is a functional component in an abstract architecture.
  – It’s purpose is to determine paths though a network
  – It operates on a topology map (the Traffic Engineering Database – TED)
    • Nodes and links == connectivity graph
    • Node constraints and link constraints == metrics and capabilities
    • Learned from the routing protocol in the network, or from the inventory database, or direct from the network nodes
  – It can be realised as a component of an existing device (NMS, router, switch) or as a dedicated server (or virtualised service)

• Benefit of identifying PCE as a separate service…
  – Offload CPU-heavy computations
    • Provide advanced and sophisticated algorithms
  – Coordinate computation across multiple paths
  – Operate on an enhance TED

• Primary initial purpose was for Traffic Engineered MPLS LSPs
  – Rapidly picked up for optical transport networks
PCE Deployment Models

- The Path Computation Client (PCC) may be co-located with the PCE or separate.
• **PCE**: Path Computation Element. An entity (component, application, or network node) that is capable of computing a network path or route based on a network graph and applying computational constraints – RFC 4655
  – This does not say it is a dedicated server
  – It can be embedded in a router
  – It can be embedded in *every* router

• For virtual PoP use case
  – PCE function in head-end LSR for local domain
  – PCE function in remote ASBR accessed through remote call
The PCE architecture originates in the IETF
- The main focus of the IETF is to specify protocols

PCEP is the request/response protocol for accessing the services of a PCE
- Session-based carried over TCP

Like PCE, PCEP had a very narrow purpose
- Simple path computation request/response for MPLS-TE LSPs

Initial proposals and early implementations
- Used RSVP-TE Path messages
  - It is "kind of obvious": that is exactly what we will signal
  - Just use the TCP session to give context to the usage
- It really worked

But was that really extensible?
- Even in the MPLS-TE context we needed multiple extensions
- RSVP has a lot of baggage

Result:
- A new container protocol and re-use of RSVP objects

PCE Protocol (PCEP)
• Like PCE, PCEP had a very narrow purpose
  – Simple path computation request/response for MPLS-TE LSPs

• Initial proposals and early implementations
  – Used RSVP-TE Path messages
  – It is “kind of obvious”: that is exactly what we will signal
  – Just use the TCP session to give context to the usage
  – It really worked

• But was that really extensible?
  – Even in the MPLS-TE context we needed multiple extensions
  – RSVP has a lot of baggage

• Result:
  – A new container protocol and re-use of RSVP objects
Packet networks have not been a roaring success for PCE
  - Initially, only Cisco implemented
  - It is implemented and deployed
  - Main use cases are
    • Dual-homed IGP areas
    • Centrally controlled TE domains

There is a huge amount of research and experimentation
  - More than 20 projects funded by the EU have PCE as a core component
  - A number of operators have in-depth experimentation

Commercial and Open Source Implementations
  - Software stacks from Metaswitch and Marben
    • But these are PCEP implementations, not full PCEs
  - Several Open Source implementations exist
  - Hardware vendors
  - Network operators

The best take-up for PCE so far is in optical networks
PCE - Evolution

• PCE evolved very quickly after it was invented
• Advanced PCEP encodings for non-packet environments
• PCEP extensions for coordinated path computations
  – Path protection
  – Network re-optimisation
• Cooperating PCEs for multi-domain applications
• Applicability to sophisticated services such as point-to-multipoint
• Hierarchical PCE for selection of paths across multiple domains
• And evolution continues today
Cooperating PCEs

- The first “interesting” problem for PCE was inter-domain TE
  - “A domain is any collection of network elements within a common sphere of address management or path computation responsibility.” RFC 4655
  - An IGP area or an Autonomous System
  - An optical island
- Nodes in one network cannot see into other networks
  - PCEs must ask each other for advice
Hierarchical PCE

- How do I select a path across multiple domains?
- Parent PCE (pPCE) has
  - An overview topology showing connectivity between domains
  - Communications with each Child PCE (cPCE)
- Parent can selectively and simultaneously invoke children to assemble an end-to-end path
Hierarchical PCE status

• Well implemented and being experimented with
  – Particularly for optical networks
    • Navigating a mesh of small domains
    • Optimising multi-layer networks (especially IP-over-optical)

• Makes sense when all domains under one administration
  – No questions of information sharing or security
  – Separation is for operational reasons
  – Partition may allow vendor specialisation

• Who owns a PCE in a multi-administrator system?
  – A Path Computation Broker?
  – Or is it all governed by policy?
Traffic Engineering Basis

• Traffic placement over the network
  – at a finer granularity than IP Interior Gateway Protocols (IGPs)

• For various motivations
  – network load balancing (more flexible than multi-path)
  – quality of service (e.g. available bandwidth)
  – protection (e.g. diversity of paths, Fast ReRoute)
  – constraint-based routing (e.g. resource inclusion/exclusion)

• In circuit-switched networks, traffic engineering is implicit
  – e.g.: SDH/OTN containers, Wavelengths Division Multiplexing
  – per circuit placement
  – bandwidth reservation implied by circuit provisioning
  – no other alternative than constraint-based routing
Traffic Engineering with MPLS

• MPLS provides a solution for packet Traffic Engineering (TE)
  – i.e. control plane extensions to implement TE features
    • IGPs advertise parameters enabling constraint-based routing in network nodes
    • tunnel head node computes paths according to requested constraints (which might be a fully-specified path)
    • signaling distributes tunnel configuration information along selected paths (notion of “TE-Label Switched Path”)

• GMPLS provides a control protocol suite for connection-oriented technologies at large
  – can control MPLS, SDH, OTN, WDM...
  – protocol suite generalized from MPLS-TE
  – control messages define how nodes exchange information
    • no need to carry them in-band/have in-band lookup
MPLS-TE / (G)MPLS

- MPLS-TE and (G)MPLS designate a Tools Box
  - QoS routing policy and resources reservation are possible
- MPLS-TE allow to share traffics on top of the network
  - Complexity is reported to Network Engineering for an efficient tunnel placement
- QoS is guarantee
  - if rate-limit and policer are associated to the TE-LSP
- Possibility to combine DiffServ and MPLS-TE
  - DS-TE allow reservation per Class Type instead of per physical link

FEC configuration combine with a rate-limit to specify which IP packet could follow the LSP-TE

Best-Effort Route

TE parameters are exchange through IS-IS-TE or OSPF-TE
Constraint-Based Routing Principles in (G)MPLS

- Topology Distribution (IGP-TE)
- TE Database
- CSPF
- Explicit Route
- Constraints
- Signaling

operator

service head node

in case of late TED update, crankback enables to refine the constraint set
Motivations for PCE usage [1]

- **CPU-Intensive Path Computation**
  - routers may have constraints on their CPUs
  - computation can be off-loaded to some other PCE(s)

- **Partial Visibility**
  - the node responsible for path computation may have limited topology visibility to the destination
  - computation can be performed by distributed and cooperating PCEs across different domains

- **Absence of the TED**
  - TED maintenance may require a lot of memory and CPU activity for routers and network nodes
  - an external PCE may be appropriate to establish and maintain the TED

- **Node Outside the Routing Domain**
  - A LER might not be part of the routing domain for administrative policies
  - A distinct PCE may help for optimal shortest path computations that couldn't be guaranteed otherwise
Motivations for PCE usage [2]

• Network Element Lacks Control Plane Capability
  – legacy optical networks typically do not have control plane capability
  – Running a PCE allows to compute paths and send cross-connection commands to each involved node

• Backup Path Computation for Bandwidth Protection
  – fast reroute protection of TE LSPs may benefit from an external PCE
  – backup TE LSPs can be computed in a coordinated way allowing bandwidth sharing across backup tunnels

• Multi-layer Networks
  – different client- and server-layer networks may be considered distinct path computation regions
  – usage of cooperating PCEs may solve path computation issues from one client-layer network region, across the server-layer network, to another client-layer network region.

• Path Selection Policy
  – Delegating path computations to an external PCE helps to apply policies in a centralized point
Limitations of Head-Node Computation

- Path computation happens on a per-domain basis
  - TE information is not advertised out of IGP areas (scalability issue)
    - only reachability is (using IGP or BGP)
  - for multi-area, multi-AS, multi-layer (typically overlay interconnection)
    - path computation is sub-optimal
    - path selection in domain 1 may lead to a dead-end in domain 2 (e.g. not enough bandwidth)
- Path computation may be a burden for some nodes
  - limited control boards (CPU, RAM…)
  - multiple constraints, point-to-multipoint computation…
  - coordinated computations (e.g. diversity, optimization)…
• The PCE is defined as:
  • an entity that is capable of computing a network path or route based on a network graph, and of applying computational constraints during the computation

• It is a network application that can be located within a network node (e.g. a router) or on dedicated server

• PCE is not conceived to replace the Internet model with its distributed intelligence

• And above all, PCE is not a centralized "all-seeing oracle in the sky"
  • It makes use of cooperative and distributed functionalities for TE information maintenance and computational ability
- Path computation is out-sourced from the signaling head node
  - client server relationship between...
    - LSP head node, i.e. “Path Computation Client” (PCC)
    - path selection function, i.e. “Path Computation Element” (PCE)

- PCEs need to have topology visibility
  - like a path computation engine inside a typical node

- No change in TE-LSP provisioning
  - using signaling mechanism from head node to tail node

PCE function
- either standalone device(s)
- or embedded in some nodes
Main components

- The PCE architecture has two main components
  - The Path Computation Client (PCC)
    - It is any client application requesting a path computation to a PCE
    - It may be located within a network node (e.g. GMPLS controller), in the NMS, or even in a PCE
    - It may use different PCEs for path computation
      - e.g., to distribute the set of requests for load balancing purposes
  - The PCE itself
    - It is the remote component that is able to provides complex path computation functions on behalf of its PCC(s)
    - It may be located in an LSR, in the NMS, or a dedicated server
    - The PCE collects and maintain the TE information in a dedicated TE Database (TED)
Path Computation Element Building Blocks

- Topology Distribution (IGP-TE)
- TE Database
- CSPF
- Explicit Route
- PCEP
- PCE
- PCC

1. Path Computation Element (PCE)
Path computation models

• Two models are defined
  • Centralized path computation
    1. all path computations for a given domain are performed by a single, centralized PCE
       – dedicated server (external PCE node)
       – designated router (a composite PCE node)
       – all PCCs in the domain send their path computation requests to the central PCE
    2. to avoid single point of failure issues, backup PCE(s) may be deployed
  • Distributed path computation
    – multiple PCEs are deployed in a given domain
    – computation of paths is shared among those PCEs
    3. single PCE path computation
       – A single PCE computes entire paths in a single IGP area without collaboration of other PCEs
       – e.g. at the ingress LSR/composite PCE node, or at an external PCE
    4. multiple PCE path computation
       – more than one PCE cooperate in the computation of a single path
       – E.g. loose hop expansion performed by transit LSRs/composite PCE nodes, or computations in multi-domain scenarios
Path computation models

1. Centralized Path Computation
2. Centralized Path Computation with backup
3. Distributed path computation without collaboration
4. Distributed path computation with collaboration
Improvements Brought by the PCE Architecture

• Enable more complex path computations
  – multiples constraints (e.g. optical impairments)
  – point to multi-point, over single or multiple domains
  – coordinated requests, global optimization

• Optimized tunnel placement in multi-area/AS/domain networks
  – in a scalable manner
  – while keeping domain topology confidentiality

• Optimized tunnel placement in multi-layer networks
  – e.g. MPLS over GMPLS-controlled WDM
  – topology usually advertised per layer (“overlay” model)
    • respect operational constraints
    • addresses the scalability of the “peer” model
New Doors Opened by the PCE Architecture

• Allows finer routing customization on a per tunnel basis
  – inclusion/exclusion of network resource
  – metric to use in the routing algorithm
  – metric bound for acceptable responses
  – objective function to specify optimization criteria

• Standardizes the protocol to access a path selection engine
  – higher influence on parameters to support
    • for vendors’ equipment capable of acting as PCE
  – enables the use of standalone path computation devices
    • could be finely tuned/implemented by operators
    • a first step of MPLS/GMPLS in Software-Defined Networks
PCEP: basics and mechanisms

• The scenario
• PCEP concept
• PCEP messages
• PCEP steps
PCE is an element whose function is to compute a path or a set of paths between 2 nodes of the network. A node which requests a path computation is named Path Computation Client (PCC).

The path computation is based on information such as network topology and network resources. These information can be provided by routing protocols or other means and will be further discussed.

We can assume that network information is always available and updated for PCE to appropriately run.
The scenario (II)
PCEP is a protocol which aims to enable the communication between PCEs and PCCs. It defines a set of messages which are exchanged between peers over TCP, satisfying the requirements for reliable communication and simplifying PCEP implementation.

Communication between two PCEs is also important and will be discussed later in this tutorial.
Each PCEP message (excluding the Keepalive message) carries a set of information organized as objects, which are message specific.

<table>
<thead>
<tr>
<th>Message</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open</td>
<td>Start a PCEP session</td>
</tr>
<tr>
<td>Keepalive</td>
<td>Maintain a PCEP session active</td>
</tr>
<tr>
<td>PCReq</td>
<td>Path Computation Request sent from the PCC to the PCE</td>
</tr>
<tr>
<td>PCRep</td>
<td>Path Computation Reply sent from the PCE to the PCC</td>
</tr>
<tr>
<td>PCNtf</td>
<td>Notification message used to inform the occurrence of predetermined events</td>
</tr>
<tr>
<td>PCErr</td>
<td>Error message</td>
</tr>
<tr>
<td>Close</td>
<td>End of a PCEP session</td>
</tr>
</tbody>
</table>
PCEP Steps: Overall view

- Session establishment
- Path Computation Request
- PCE Reaction
  - Path Computation Reply
  - Illustrative Example
  - Notification Message
  - Error Message
- Finishing the PCEP session
  - Session duration
PCEP Steps: Session establishment (I)

- The first step is the establishment of a TCP connection between the PCC and PCE. The initialization is triggered by the PCC which needs to reach some other network node.

- The second step is the establishment of a PCEP session over the TCP connection exchanging Open messages, which includes various session parameters like Keepalive message frequency, Deadtimer, and others.

- Each time the Keepalive timer expires, a Keepalive message is sent and the timer is restarted. Nevertheless, other messages may also restart this timer, avoiding unnecessary traffic of Keepalive messages. The Deadtimer is also restarted when a message is received.
PCEP Steps: Session establishment (II)

TCP 3-way handshake
Parameters negociation
Each PCEP peer may have different Keepalive time

PCEP Steps: Session establishment (II)

PCC

PCE

SYN

SYN/ACK

ACK

Open msg

Open msg

Keepalive

Keepalive
PCEP Steps: Path Computation Request

- A PCC uses a PCReq message to request a path computation to the PCE. It must include at least the source and destination.

- It may include path constraints like requested bandwidth, points of failure that should be excluded from the path being calculated, etc.

- PCC may also assign the request priority and if load-balancing is allowed, including maximum number of paths and minimum path bandwidth in a path group.
PCEP Steps: PCE reaction

- Once PCE receives a PCEReq from the PCC it can issue any of the following messages:
  - PCE Reply
  - PCE Notification Message
  - PCE Error Message

- PCE can also do nothing
  - In that case the PCC may issue a Notification Message to the PCE cancelling the previous request
PCEP Steps: Path Computation Reply

- A PCE uses a PCRep message to send the path computation result to the PCC. It may carry one or more computed paths if the path computation succeeded, or a negative reply if a path could not be determined.

- When sending a negative reply, the PCE may specify the reasons why the path could not be determined and include advice about which constraints could be relaxed to be more likely to succeed in a future request.
**PCEP Steps:**
**Illustrative example**

PCC sends a PCReq message:
- **source:** A, **destination:** E
- **Bandwidth = 30mbps**, **avoid B**

PCE computes the request. No path available with required constraints.

PCE sends a PCReq message:
- **source:** A, **destination:** E
- **Bandwidth = 20mbps**, **avoid B**

PCC sends a PCReq message:
- **source:** A, **destination:** E
- **Bandwidth = 20mbps**, **avoid B**

PCE computes the request. Success!
- Path: A-D-E

PCE sends a PCRep message:
- **Path:** A-D-E

Communication established between A and E
PCEP Steps: Notification message

- The PCNtf message may be sent by the PCE or by the PCC:
  - PCE may notify the PCC that it is overloaded, that some requests will not be satisfied, etc. The PCC may decide to redirect requests to other PCE.
  - PCC may notify some particular events to the PCE, like cancellation of pending requests.
PCEP Steps:
Error message

- PCErr message must be sent when a protocol error condition is met, like:
  - when a message without a mandatory object is received.
  - unexpected message
  - policy violation
PCEP Steps: Finishing a PCEP session

- Any PCEP peer may decide to finish the session, send a Close message and close the TCP connection.
- When receiving a Close message:
  - PCC clears all pending requests sent to PCE
  - PCE clears all pending requests received from PCC
- A PCEP session is immediately finished if a TCP connection failure occurs.
When a PCC sends path calculation requests on a high frequency, it may keep the session alive to avoid additional processing delays.

On the other hand, if the path calculation request is a rare event, the session may be opened and closed for each request.
The TED

- Traffic Engineering Database (TED) is an essential internal component of a PCE
  - provides the updated snapshot of the controlled network and its resources
  - PCE algorithms resort to TED as primary information source input
What does TED store?

• The topology of the controlled network
  – Nodes
  – Links
  – Nodes/Links connectivity

• The available resources and attributes
  – Available Link Bandwidth: 500 Mb/s
  – Link Metrics (e.g., costs): 1
TED: which network?

• The network controlled by PCE
  – **Single domain**: the whole domain network
  – **Multi-layer**: the x-layer topology (e.g., IP topology)
  – **Multi-area**: the network area and the ASBRs
  – **Multi-domain**: the owned network domain, including border nodes and inter-domain links
Hierarchical TED (H-TED)

• In HPCE architecture only
• Parent PCE resort to a special H-TED
  – All inter-domain links with available bw
  – Network domains and border node connectivity
  – Abstract intra-domain representation
    • Abstract node, link, star, etc.
TED update

- Different mechanisms may be used
  - Passive OSPF-TE or IS-IS-TE peering
    - TE info updated by means of Opaque Link State Advertisement or Link State PDU
  - Management-based (e.g., SNMP)
  - PCEP Notifications

**OSPF Opaque LSA**

```
0 1 2 3 4 5 6 7 8 9 0 1
+---------------------------------------------------------------+
| LS age | Options | 0, 10 or 11 |
+---------------------------------------------------------------+
| Opaque Type | Opaque ID |
+---------------------------------------------------------------+
| Advertising Router |
+---------------------------------------------------------------+
| LS Sequence Number |
+---------------------------------------------------------------+
| LS checksum | Length |
+---------------------------------------------------------------+
```

**SNMP**
Topology Synchronization: OSPF-TE or IS-IS-TE

- Usual link-states IGPs with traffic engineering extensions
  - the standard IP routing protocols (OSPF and IS-IS)...
  - ... extended to provide TE data base synchronization

- States of links in the area
  - metrics, Shared Risk Link Groups, admin. groups/resource classes
  - reservable bandwidth, unreserved bandwidth
  - GMPLS switching capability (packet, L2, TDM, lambda, fiber)
  - unnumbered links (Node ID + port ID) may be used

- Optionally some node information
  - TE router ID
  - TE capabilities

- Optionally some inter-domain TE-links connected to the area
  - no IGP adjacency
  - area-scoped flooding of configured parameters
Topology Synchronization of Multiple IGP Areas / Layers

- Adjacencies with IGP areas enable PCE to import topologies
  - if not forwarding-capable PCE (like some IGP monitoring tools)
  - a PCE may run several IGP instances/protocols
  - adjacencies may run over links or over tunnels
PCE Discovery: OSPF-TE or IS-IS-TE

• Enables to dynamically advertise PCE parameters and aliveness
  – In complement to TE flooding in IGP
• Solution based on IGP-TE extensions
  – protocols already there for topology synchronization
  – enables to convey per PCE information
    • address
    • scope: intra-area, inter-area, inter-AS, inter-layer
    • domains (areas or ASes) of topology visibility
    • neighbor domains (reachable area or AS borders)
    • capabilities: bidirectional, diverse, synchronized, P2MP…
Advanced features

- Core PCEP extensions
- Inter-layer path computation
- Stateful PCE
Backward Recursive Path Computation (BRPC)

- Procedure to optimize multi-domain path selection
  - a virtual shortest path tree is built while replies progress to PCC
  - the head end PCE can select the shortest path over that tree
  - AS path could be specified as IRO Object for inter-domain
A PCE can subcontract a segment computation to another PCE
- PCEP exchange between PCEs similar to a PCC-PCE dialogue
- A PCE may act both as a server and as a client (PCC)

Topology confidentiality between domains
- Key-based mechanism in PCEP and RSVP-TE

No mechanism to compute AS Path is provided
- Generally AS Path follow BGP AS Path which is generally inefficient as it corresponds to the Best-Effort route
Hierarchical Traffic Engineering

- Hierarchical Traffic Engineering abstract AS topology and flood it
  - Re-use IGP-TE functions on a dedicated Area / Region to export abstracted model
  - Abstract model ensure Confidentiality and Scalability
- New functions have been added to the PCE
  - IGP-H-TE (OSPF-TE for ETICS prototype) to learn AS topology
  - H-TED and H-CSPF to compute AS path
  - Run standard BRPC with neighbours PCE drive by the AS-Path
- All run in a new router named AS Virtual Router (ASVR)
1. Explicit route exclusions
2. Path confidentiality
3. Objective functions
4. DiffServ support
5. PCEP extensions for GMPLS
6. Point-to-Multipoint TE LSPs
Explicit route exclusions

What does it mean a route exclusion in path computation?

- Sometimes, a Path Computation Client (PCC) needs to specify abstract nodes, resources, and Shared Risk Link Groups (SRLGs) that are to be explicitly excluded from the path computation.

When do we need to specify a route exclusion?

- In inter-domain Label Switched Paths (LSPs), disjoints paths may be computed by cooperation between PCEs (computing separated segments)
- When a network operator wants a path to avoid specified nodes for administrative reasons.
There are two types of route exclusion described in [RFC 4874]:

1. Exclusion of certain abstract nodes or resources from the whole path. This set of abstract nodes is referred to as the **Exclude Route List**.

2. Exclusion of certain abstract nodes or resources between a specific pair of abstract nodes present in an explicit path. Such specific exclusions are referred to as an **Explicit Route Exclusion**.

New objects for route exclusions are defined in [RFC 5521]:

- New **Exclude Route Object (XRO)** is defined to convey the **Exclude Route List**
- Adding to the existing **Include Route Object (IRO)**:
  - The **Explicit Exclusion Route subobject (EXRS)** to convey **Explicit Route Exclusions**.
Explicit route exclusions

**XRO object in PCReq message**
- Provides a list of network resources that have to be excluded from the path computation.
- Flags associated with each list member instructs the PCE as to:
  1. Whether the network resources must be excluded.
  2. Whether the PCE should make best efforts to exclude the resources.

**XRO object in PCRep message**
- Using NO-PATH object to indicate the set of elements of the original XRO that prevented the PCE from finding a path.
- For a successful path computation when the PCE wishes to provide a set of exclusions to be signaled during LSP setup using the extensions to RSVP-TE.
Explicit route exclusions

XRO Body format

```
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+--------------------------------------------------+
| Reserved                                         | Flags |
+--------------------------------------------------+

// (Subobjects) //
```

| Reserved | MUST be set to zero on transmission and SHOULD be ignored on receipt. |
| Flags    | F (Fail - 1 bit) → When set, the requesting PCC requires the computation of a new path for an existing TE LSP that has failed. |
| Subobjects | The XRO is made up of one or more subobject(s). |
Explicit route exclusions

Explicit Exclusion Route subobject (EXRS)

- Defines network elements that must not or should not be used on the path between two abstract nodes or resources explicitly indicated.

EXRS subobject format

<table>
<thead>
<tr>
<th>L</th>
<th>Type</th>
<th>Length</th>
<th>Reserved</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- MUST be set to zero on transmission and MUST be ignored on receipt.
- MUST be set to zero on transmission and SHOULD be ignored on receipt.
Core PCEP extensions

1. Explicit route exclusions
2. Path confidentiality
3. Objective functions
4. DiffServ support
5. PCEP extensions for GMPLS
6. Point-to-Multipoint TE LSPs
Path confidentiality

Why do we need confidentiality in PCEP?

- An important element of inter-domain TE is that TE information is not shared between domains for scalability and confidentiality reasons.
- A single PCE is unlikely to be able to compute a full inter-domain path.

In which scenario we apply path confidentiality?

When PCE-based path computation technique is required with cooperation between PCEs located in different domains (such as the case between Autonomous Systems (ASes) belonging to different Service Providers).
New object for path confidentiality is defined in [RFC 5520]:

- New Path-Key Subobject (PKS) is defined.

\[
\text{PKS} = \text{Path-Key} + \text{PCE-ID}
\]

**Path-Key**

- Identifier used to represent the Confidential Path Segment (CPS) within the context of the PCE identified by the PCE-ID.

**PCE-ID**

- Identifies the PCE that can decode the Path-Key using an identifier that is unique within the domain that the PCE serves.

- Has to be mapped to a reachable IPv4 or IPv6 addr of the PCE by the first node of the CPS and a PCE MAY use one of its reachable IP addresses as its PCE-ID.

Two subobjects are defined to allow IPv4 and IPv6 addresses to be carried:

- PKS with 32-Bit PCE ID and PKS with 128-Bit PCE ID
### PKS with 32-Bit PCE ID

<table>
<thead>
<tr>
<th>L</th>
<th>The L bit SHOULD NOT be set, so that the subobject represents a strict hop in the explicit route.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type</strong></td>
<td>Subobject Type for a Path-Key with 32-bit PCE ID (64).</td>
</tr>
<tr>
<td><strong>Length</strong></td>
<td>The Length contains the total length of the subobject in bytes, including the Type and Length fields. The Length is always 8.</td>
</tr>
<tr>
<td><strong>PCE ID</strong></td>
<td>A 32-bit identifier of the PCE that can decode this path-key. The identifier MUST be unique within the scope of the domain that the CPS crosses, and MUST be understood by the LSR that will act as PCC for the expansion of the PKS.</td>
</tr>
</tbody>
</table>
Path confidentiality

PKS with 128-Bit PCE ID

<table>
<thead>
<tr>
<th>L</th>
<th>The L bit SHOULD NOT be set, so that the subobject represents a strict hop in the explicit route.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Subobject Type for a Path-Key with 128-bit PCE ID (65).</td>
</tr>
<tr>
<td>Length</td>
<td>The Length contains the total length of the subobject in bytes, including the Type and Length fields. The Length is always 20.</td>
</tr>
<tr>
<td>PCE ID</td>
<td>As above but with a identifier of 128-bit.</td>
</tr>
</tbody>
</table>
Path confidentiality

Mode of operation

1. PCC Ingress sends a path computation request to PCE-1.

2. PCE-1 is unable to compute an end-to-end path and invokes PCE-2.

3. PCE-2 computes a path segment from ASBR-2 to the egress as \{ASBR-2, C, D, Egress\}.

4. In order to **protect the confidentiality** of the topology, PCE-2 may send PCE-1 a path segment expressed as \{ASBR-2, PKS, Egress\} where the **PKS is a Path-Key Subobject**.

5. PCE-1 supply the full path to the PCC as \{Ingress, A, B, ASBR-1, ASBR-2, PKS, Egress\}.

6. Signaling proceeds in the first AS as normal, but when the Path message reaches ASBR-2, the information in the **PKS** is used to request PCE-2 for a path segment, and PCE-2 will return the segment \{ASBR-2, C, D, Egress\} allowing signaling to continue to set up the LSP.
Core PCEP extensions

1. Explicit route exclusions
2. Path confidentiality
3. Objective functions
4. DiffServ support
5. PCEP extensions for GMPLS
6. Point-to-Multipoint TE LSPs
What does it mean an objective function in path computation?

The computation of one or a set of Traffic Engineering Label Switched Paths (TE LSPs) in MPLS or GMPLS networks is subject to a set of one or more specific optimization criteria, referred to as objective functions (e.g., minimum cost path, widest path, etc.).

Why do we need to use objective functions in PCEP?

- In the PCE architecture, a PCC may want a path to be computed for one or more TE LSPs according to a specific objective function.
- The PCC needs to instruct the PCE to use the correct objective function.
- It is useful for the PCC to be able to automatically discover the set of objective functions supported by each PCE.
Supported objective functions

An objective function is used by the PCE during the path computation in order to select the "best" candidate paths.

There are three objective functions that apply to the computation of a single path:

- Code 1: Minimum Cost Path (MCP)
- Code 2: Minimum Load Path (MLP)
- Code 3: Maximum residual Bandwidth Path (MBP)

There are three objective functions that apply to a set of path computation requests the computation of which is synchronized:

- Code 4: Minimize aggregate Bandwidth Consumption (MBC)
- Code 5: Minimize the Load of the most loaded Link (MLL)
- Code 6: Minimize the Cumulative Cost of a set of paths (MCC)
New Metric Types

Three metric types are specified in [RFC5440] for the METRIC object:

- Type 1: IGP metric
- Type 2: TE metric
- Type 3: hop count

Four new metrics for objective functions extension are defined in [RFC5541]:

- Type 4: *Aggregate bandwidth consumption.*
- Type 5: *Load of the most loaded link.*
- Type 6: *Cumulative IGP cost.*
- Type 7: *Cumulative TE cost.*

These metrics may be used in a:

- PCReq: to indicate a bound or to request the computation of a metric.
- PCRep: to indicate a computed metric.
Objective functions

PCEP OF-List (Objective Function list) TLV

- MAY be carried within an OPEN object sent by a PCE in an Open message to a PCEP peer so as to indicate the list of supported objective functions.

<table>
<thead>
<tr>
<th>0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>OF Code #1</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>padding</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>N * 2 (where N is the number of objective functions)</td>
</tr>
<tr>
<td>Value</td>
<td>List of 2-byte objective function code points, identifying the objective functions supported by the sender of the Open message.</td>
</tr>
<tr>
<td>OF Code (2 bytes)</td>
<td>Objective Function code point identifier.</td>
</tr>
</tbody>
</table>
Objective functions

Objective Function object (OF)

OF object can be carried within:

- **PCReq message**: to indicate the desired/required objective function to be applied by the PCE during path computation.
- **PCRep message**: to indicate the objective function that was used by the PCE during path computation.

### OF Code (2 bytes)

The identifier of the objective function.

### Reserved (2 bytes)

This field **MUST** be set to zero on transmission and **MUST** be ignored on receipt.

### Optional TLVs

Optional TLVs may be defined in the future so as to encode objective function parameters.
Core PCEP extensions

1. Explicit route exclusions
2. Path confidentiality
3. Objective functions
4. DiffServ support
5. PCEP extensions for GMPLS
6. Point-to-Multipoint TE LSPs
Is Diffserv supported by PCEP?

- Diffserv-aware MPLS Traffic Engineering (DS-TE) in PCEP is able to enforce different bandwidth constraints for different classes of traffic.
- It describes mechanisms to achieve per-class traffic engineering, rather than on an aggregate basis across all classes by enforcing Bandwidth Constraints (BCs) on different classes.

How it is achieved?

- A new PCEP object called CLASSTYPE is defined, which carries the Class-Type of the TE LSP in the path computation request.
- A PCE uses the Class-Type to identify the bandwidth constraint of the TE LSP.
CLASSTYPE Object

- Used to specify the Class-Type of a Traffic Engineering Label Switched Path (TE LSP).
- If the TE LSP for which the path is to be computed belongs to Class 0, the PCReq MUST NOT contain the CLASSTYPE object (backward compatibility).

The CLASSTYPE object contains a 32-bit word PCEP common object header defined in [RFC5440].

CT: 3-bit field that indicates the Class-Type. Values allowed are 1, 2, ..., 7. The value of 0 is Reserved.
Reserved: 29-bit reserved field. It MUST be set to zero on transmission and MUST be ignored on receipt.
Diffserv Support

Determination of TE Class and PCReq message format with CT Object

- Up to 8 TE-classes
- The CLASSTYPE value carried in the CLASSTYPE object and the setup priority in the LSP Attribute (LSPA) object are used to determine the TE-class corresponding to the path computation request.
- If the LSPA object is absent, the setup priority is assumed to be 0.
- The CLASSTYPE object be inserted after the END-POINT objects.

```
<PCReq Message>::= <Common Header>
    [<SVEC-list>]
    <request-list>
where:
    <svec-list>::=<SVEC>[<svec-list>]
    <request-list>::=<request>[<request-list>]
    <request>::= <RP>
        <END-POINTS>
            [<CLASSTYPE>]
            [<LSPA>]
            [<BANDWIDTH>]
            [<metric-list>]
            [<RRO>]
            [<IRO>]
            [<LOAD-BALANCING>]
where:
    <metric-list>::=<METRIC>[<metric-list>]
```
Error codes for CLASSTYPE Object

Error types for Diffserv (defined in [RFC 5440])

- **Error-Type = "Unknown Object"**: when the PCE does not recognize the CLASSTYPE object.

- **Error-Type = "Diffserv-aware TE error"**.
  - Error values (defined in [RFC 5455])
    - **Error-value=1: Unsupported Class-Type ⇒** when a PCE does not support the particular Class-Type.
    - **Error-value=2: Invalid Class-Type ⇒** when a PCE determines that the Class-Type value is not valid.
    - **Error-value=3: Class-Type and setup priority do not form a configured TE-class**
Core PCEP extensions

1. Explicit route exclusions
2. Path confidentiality
3. Objective functions
4. DiffServ support
5. PCEP extensions for GMPLS
6. Point-to-Multipoint TE LSPs
PCEP extensions for GMPLS

Are the GMPLS networks supported by PCEP?

- The initial effort focused on solving the path computation problem within a domain or over different domains in MPLS networks.
- Service providers have also requirements for path computation in GMPLS-controlled networks (e.g. based on WDM, TDM, or Ethernet).

What are the PCEP requirements for GMPLS?

- Switched data flow by the LSP: switching Type (for instance L2SC or TDM), switching Encoding (e.g., Ethernet, SONET/SDH), signal Type (e.g. in case of TDM/LSC switching capability).
- Data flow specific traffic parameters, which are technology specific (e.g. SDH/SONET or G.709 OTN)
- Support for asymmetric bandwidth requests and unnumbered interface identifiers
- Label information and technology specific label(s) such as wavelength labels.
- Ability to indicate the requested granularity for the path ERO: node, link or label.
New PCEP extensions for GMPLS are:

1. GMPLS capability advertisement
   - New optional GMPLS-CAPABILITY TLV for use in the OPEN object to negotiate the GMPLS capability.

```
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+---------------------------------------------+
<table>
<thead>
<tr>
<th>Type=14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
</tr>
<tr>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>Flags</td>
</tr>
</tbody>
</table>
+---------------------------------------------+
```

2. RP object extension
   - In order to indicate the used routing granularity in the response, a new 2-bit flag in the RP object is added.

```
0 : reserved
1 : node
2 : link
3 : label
```
New PCEP extensions for GMPLS are:

3. Traffic parameters encoding, BANDWIDTH object extensions

- A new object type is introduced for the **BANDWIDTH object**
  (Generalized-Bandwidth, Generalized Bandwidth of existing TE-LSP)

**Bandwidth spec length**: indicates the length of the BANDWIDTH object type 3 and 4.

**Reverse bandwidth spec length**: indicates the length of the reverse bandwidth spec field.

**Bw Type field**: determines which type of bandwidth is represented by the object.

The encoding of the field **generalized bandwidth** is the same as in RSVP-TE.
New PCEP extensions for GMPLS are:

4. Traffic parameters encoding, LOAD-BALANCING

- A new object type is introduced for the **LOAD-BALANCING object** (Generalized bandwidth).

**Bandwidth spec length**: total length of the min bandwidth specification.

**Reverse bandwidth spec length**: total length of the reverse min bandwidth specification.

**BW Spec Type**: bandwidth specification type.

**Max-LSP**: maximum number of TE LSPs in the set.

**Min Bandwidth spec**: minimum bandwidth spec of each element of the set of TE LSPs.

**Min Reverse Bandwidth spec**: minimum reverse bandwidth spec of each element of the set of TE LSPs.
New PCEP extensions for GMPLS are:

5. **END-POINTS Object extensions**
   - A new object type is introduced for the **END-POINTS object** (GENERALIZED-ENDPOINT).

6. **IRO and XRO extensions**
   - A new **TLV type for label** is allowed in IRO and XRO objects.

```
<table>
<thead>
<tr>
<th>X</th>
<th>Type=3</th>
<th>Length</th>
<th>U</th>
<th>Reserved</th>
<th>C-Type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Label</td>
</tr>
</tbody>
</table>
```

**X**: the X-bit indicates whether the exclusion is mandatory or desired.

**Type**: the Type of the XRO Label subobject is 3.

**Length**: the total length of the subobject in bytes (including the Type and Length fields).

**U**: See [RFC3471]

**C-Type**: The C-Type of the included Label Object as defined in [RFC3471].

**Label**: See [RFC3471].
New PCEP extensions for GMPLS are:

8. LSPA extensions
   - A new **PROTECTION-ATTRIBUTE TLV** is added to the LSPA object.

```
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Type | Length |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| S | P | N | O | Reserved | LSP Flags | Reserved | Link Flags |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| I | R | Reserved | Seg.Flags | Reserved |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The content is as defined in [RFC4872], [RFC4873].

**LSP Flags** or **Link flags** field can be used by implementation for routing policy input.

9. NO-PATH Object Extension
   - A new **NO-PATH-VECTOR TLV** added to the NO-PATH object
1. Explicit route exclusions
2. Path confidentiality
3. Objective functions
4. DiffServ support
5. PCEP extensions for GMPLS
6. Point-to-Multipoint TE LSPs
Does the legacy PCEP provide support for the computation of P2MP TE LSPs?

- PCEP is initially designed as a communication protocol between PCCs and PCEs for point-to-point (P2P) path computations.
- The RFC 5671 describes the motivation for using PCE to compute paths for P2MP TE LSPs and the RFC 6006 describes the mechanisms to handle requests and responses for the computation of paths for P2MP TE LSPs.

Which is the motivation to use PCE for P2MP TE LSPs in MPLS and GMPLS?

- Path computation for P2MP TE LSPs presents a high computation complexity for the LSRs that can interfere with the control and management plane operation to maintain LSPs.
- The PCE architecture offers a way to offload such path computations from LSRs.
New PCEP extensions for P2MP TE LSPs:

Open Message Extension

- New optional TLV for PCEP OPEN object to indicate the PCE's capability to perform P2MP path computations [RFC6006].
- IANA has allocated value 6 from the "PCEP TLV Type Indicators" sub-registry. The description is "P2MP capable", and the length value is 2 bytes.

Efficient passing of P2MP LSPs between the PCE and PCC

- A new PCEP object class and type are requested for Secondary Explicit Route Object (SERO) and the Secondary Record Route Object (SRRO).
  - SERO and SRRO are used to report the route of an existing TE LSP for which a reoptimization is desired. The format and content of the SERO and SRRO are defined in [RFC4875].
New PCEP extensions for P2MP TE LSPs:

**Request and Reply Messages**

- Modification of the PCReq and PCRep message to support P2MP, preserving compatibility with the [RFC 5440], adding the following flags to the RP Object:
  - The F-bit is added to the flag bits to indicate to the receiver that the request is part of a fragmented request, or is not a fragmented request.
  - The N-bit is added in the flag bits field to signal the receiver of the message that the request/reply is for P2MP or is not for P2MP.
  - The E-bit is added in the flag bits field to signal the receiver of the message that the route is in the compressed format or is not in the compressed format.

**RP and END-POINTS objects extension**

- RP extension in order to a PCC can signal a P2MP path computation request to the PCE, receiving the PCEP request.
- END-points extension to improve the efficiency of the message exchange between PCC and PCE in the case of P2MP path computation.
New PCEP extensions for P2MP TE LSPs:

### P2MP Objective Functions

- Six objective functions have been defined in [RFC5541] for P2P path computation.
- The [RFC6006] adds two additional objective functions namely:
  - SPT (Shortest Path Tree) with code 7:
    - Minimize the maximum source-to-leaf cost with respect to a specific metric.
  - MCT (Minimum Cost Tree) that apply to P2MP path computation with code 8:
    - Minimize the total cost of the tree, that is the sum of the costs of tree links, with respect to a specific metric.

### New Metric Object Types

- There are three types defined for the <METRIC> object in [RFC5440]
- The [RFC6006] adds three additional types for the METRIC object:
  - P2MP IGP metric: T=8
  - P2MP TE metric: T=9
  - P2MP hop count metric: T=10
PCEP extensions also provide support for:

**P2MP TE Path Reoptimization Request**
- Specified by the use of the R-bit within the RP object.

**Adding and removing leaves to/from the P2MP Tree**
- With P2MP request using END-POINTS objects.

**Discovering Branch nodes**
- By the use of Branch Node Object providing a list of nodes that can be used or not as branch nodes.

**Synchronization of P2MP TE Path Computation Requests**
- Needed for the use of backups for P2MP LSPs.

PCEP extensions also provide support for:

**Request and Response Fragmentation***
- Needed in certain P2PM scenarios with hundreds or thousands of leaves, where the message may need to be fragmented into multiple messages.

**UNREACH-DESTINATION Object**
- This object allows the PCE to optionally specify the list of unreachable destinations.

**P2MP PCEP-ERROR Objects and Types***
- New error values associated with policy violation, capability errors or message fragmentation errors.

**PCEP NO-PATH Indicator****
- To communicate the reasons for not being able to find P2MP path computation.

Stateful PCE - Motivation

• The lack of global LSP state information may result in sub-optimal PCE algorithms. For example:
  – Non-linear effects of optical fibers may cause the provisioning of a new optical connection to degrade (QoS) of in-service connections.
    • Impairment-aware RMSA algorithms should compute new paths that ensure an acceptable QoS of the existing ones. To this end, impairment-aware RWA or RSA algorithms must also know the existing LSPs in order to re-compute the considered QoS parameters (e.g., Q factor).
  – Minimal perturbation problem → route a demand along the path that requires the lowest number of preemptions
    • Without knowledge of LSPs, preempting low-priority LSPs based on the minimum number of links may not result in the smallest number of LSPs being disrupted

• Information on the active LSP resource usage
  – Conveyed in the request (e.g. GCO [5557])
    • Not practical for a large number of connections, overhead, latency
  – Available “somewhere else” or
  – Managed in a Database (LSPDB), local to the PCE
Stateful PCE - Definition

• A stateful PCE allows for efficient path computation considering both
  – the network state (TED)
  – the LSP state (LSPDB) (i.e., set of computed paths and reserved resources in use in the network).

• A synchronization mechanism is required
  – How does the PCE construct the LSPDB?
  – PCEP protocol extensions so the PCC component of a node can “report” to the PCE

  • PCRpt message
Passive Stateful PCE

- Path computation services (new connections, re-optimization, restoration)
- LSP Setup/Reroute/Release
- Network state dissemination
- Neighbor discovery and Link property correlation

NMS

LSP Setup /Reroute/ Release Request

TED Synch

LSPDB Synch
Passive Stateful PCE

- NMS autonomously control (when and which) existing LSPs (i.e., re-optimize) and/or new LSPs (i.e., setup and release)
- PCE computes paths for new or existing LSP based on the TED and LSPDB
Active PCE

- The lack of PCE based control of path reservations (e.g., modification or rerouting) may also result in sub-optimal performance
  - The PCE is unable to reroute a connection to free resources usable for a new connection.

- An *active* PCE is able to recommend re-routing or instantiation of LSPs
  - May be stateless or stateful, but likely to hold state.
  - PCEP protocol extensions so the PCE can “Update” an existing LSP or “instantiate” a new one.
    - *PCUpd, PCInitiate messages*
Active Stateful P^CE

Path computation services (new connections, re-optimization, restoration)

LSP Setup/Reroute/Release Request

LSPDB Synch
LSP Control Delegation
LSP Reroute Request

LSP Setup/Reroute/Release
Network state dissemination
Neighbor discovery and Link property correlation
NMS autonomously control (when and which) existing LSPs (i.e., re-optimize) and/or new LSPs (i.e., setup and release).

- PCE computes paths for new or existing LSP based on the TED and LSPDB.

- PCE can autonomously reroute existing LSP during the path computation.
PCE with instantiation

- LSP setup/reroute and release request services
- LSPDB Synch
- LSP Control Delegation
- LSP Reroute Request
- LSP Setup/Release Request

Adaptive NMS

TED Synch

• LSP Setup/Reroute/Release
• Network state dissemination
• Neighbor discovery and Link property correlation
Active PCE with instant.
Usage scenarios

- PCE in optical networks
- Global Concurrent Optimization
- Deployment models in Software Defined Networks
- PCE in ABNO
- PCE for intra and inter datacenter orchestration
- PCE role in Segment Routing
- PCE in NFV
- PCE in IoT
PCE in Optical Networks

- Wavelength Switched Optical Networks (WSON) path computation
  - Lightpaths: wavelength connection
  - Routing and Wavelength Assignment (RWA)
  - Additional constraint: Wavelength Continuity

- In distributed GMPLS, provisioning may be affected by inefficiencies
• Functional TE separation between R (OSPF-TE) and WA (RSVP-TE)
  – Independent processes
  – Different TE databases, independently updated
  – Joint RWA not feasible -> suboptimal

• R<->WA temporal separation
  – R: performed at source node
  – WA: distributed, selection at destination node
  – Blocking probability due to database disalignments and connection collisions
WSON PCE architectures

- **Combined RWA** (joint, optimal, complex)
- **R+WA** (two independent computations)
- **R+distributed WA** (GMPLS-like, less efficient)
Optical impairments

- Additional constraint for path computation
- Quality of Transmission (QoT) needs validation
- Complex Impairment modeling:
  - Attenuation, ASE, PMD, CD, SPM, XPM...
  - Multi bitrate signals require modeling
- Besides RWA, lightpaths requires a pre-validation (either static or dynamic) ensuring acceptable QoT
Impairment-aware PCE

- Combined IV & RWA (optimal <-> complex)
- IV-candidates + RWA (pre-validated candidate paths, suboptimal)
- R + dWA + dIV (simple, suboptimal, post provisioning validation -> poor effectiveness)
• SSON: Spectrum switched Optical Networks
• 12.5GHz-multiple spectrum slots (ITU flexible grid)
  – Label expressed as \((n,m)\) couple
    • \(n\) = central frequency
    • \(m\) = **number of frequency slots**
• Variable number of contiguous slots, depending on the occupied frequency spectrum
• PCE in charge of Routing and Spectrum Assignment (RSA)
• Elastic transponders enable configurable optical transmission with different tx parameters (e.g., modulation format)

• PCE path computation
  – Routing
  – Spectrum Assignment
    • Assign n (m is input)
    • Assign (n,m) (bitrate is input) (different values of m may be eligible)
  – Additional output
    • Single /multiple carrier signal
    • Signal modulation format (e.g., QPSK, 16QAM)
    • FEC type, FEC code rate
  – Impairment validation
    • Alternative (modulation format, FEC) values guarantee QoT with different optical reach
Multi-Layer PCE and per Layer PCEs

• Same approach as multi-domain applications
  – either the PCEs know the topologies of both layers
  – or (client) packet layer PCEs forward PCEP requests to (server) optical layer PCEs

• PCE is adapted to any GMPLS-controlled technology
  – all connection-oriented data-planes: MPLS, SDH, OTN, WDM…
• Cooperation between IP/MPLS PCE and GMPLS PCE
  – When no path satisfy constraints between N1 and N4 at IP/MPLS layer, PCE trigs computation to GMPLS PCE
  – GMPLS PCE compute a new path between N1 and N4
Multi-Layer computation (2)

- 2\textsuperscript{nd} scenario will just add new capacity between congested links
  - Used less optical resources than 1\textsuperscript{st} scenario
  - Reduce global blocking factor
• In 3rd scenario check regularly link occupancy
  – Detect congested links and upgrade them before IP/MPLS need them
  – Improve availability speed due to the time necessary to fire up optical (from sec. to minutes)
Multi-Layer computation (4)

- Used only one PCE with 2 TEDs
  - One TED for IP/MPLS one TED for GMPLS network topology
  - Allow full disjoint path computation at both level
• PCEP Update, Report, Initiate msg
  – PCE can directly modify (or create) lightpaths and their attributes

• Enables advanced control plane operation
  – Shared Protection
  – Elastic operations
    • Increase Bitrate -> Expand reserved spectrum
  – Defragmentation
    • Compact reserved spectrum
    • Different techniques: make-before-break, hitless shifting, etc
  – Dynamic QoT-aware reconfiguration
    • Modulation format, FEC change
  – Transponder Slice-ability
1. PCReq (LSP 3: Elastic 100G -> 200 G)
2. PCUpd (LSP 2: Shift)
3. LSP 2 shift (100 GHz)
4. PCRpt (LSP 2: Shift OK)
5. PCRep (LSP 3: Elastic increase)
6. LSP 3 elastic increase
7. PCNtf (LSP 3: Elastic increase OK)
Greenfield Optimization (GO)

• Special case of GCO application when there are no TE LSPs already set up in the network.

• In this scenario, the ability to perform concurrent computation is desirable, or required, to utilize network resources in an optimal manner and avoid blocking.

• Greenfield optimization can be applied in:
  
  o Single-Layer Traffic Engineering: Greenfield optimization for single-layer TE can be applied to optical transport networks such as SDH/SONET, Ethernet Transport, WDM, etc.
  
  o Multi-Layer Traffic Engineering: The network resources and topology (of both the client and server layers) can be considered simultaneously in setting up a set of TE LSPs that traverse the layer boundary
Reoptimization of Existing Networks

• In some cases the need for reoptimization or reconfiguration may arise, due to a sub-optimal use of the resources.

• Two methods for reoptimization:
  
  ○ **Sequential reoptimization**: computes the reoptimized path of one TE LSP at a time without giving any other consideration.
    - This method could result in sub-optimal use of network resources.
  
  ○ **Concurrent reoptimization**: computes paths of the TE LSPs as a set, providing real benefits (e.g. preventing race condition).
    - However, a centralized system will typically suffer from a slower response time than a distributed system.
    - Two typical application scenarios for GCO are:
      - Reconfiguration of the Virtual Network Topology (VNT)
      - Traffic Migration
Global Concurrent Optimization

PCEP Requirements

- **GCO path computation indicator**
  - To ensure consistency in applying global objectives and global constraints in all path computations.

- **A Global Objective Function (GOF) field**
  - Function to which all individual path computation requests are subjected in order to find a globally optimal solution.

- **A Global constraints (GC) field**
  - To specify the list of global constraints to which all the requested path computations should be subjected.
    - Max link util, Min link util, Overbooking factor, Max hops, Exclusion of links/nodes

- **A Global Concurrent Vector (GCV) field**
  - To specify all the individual PCReq that are subject to concurrent path computation and subject to the global objective function and all of the global constraints.

- **GCO related Error Indicators**

- **Disruption minimization mechanism in Bulk Re-optimization/Migration**
Global Concurrent Optimization

PCEP Protocol Extensions for GCO

Global Objective Function (GOF) Specification

- Three global objective functions defined in [RFC5541] are used in the context of GCO:
  - Minimize aggregated Bandwidth Consumption (MBC)
  - Minimize the load of the Most Loaded Link (MLL)
  - Minimize the Cumulative Cost of a set of paths

GCO Indicator

- A new flag in the SVEC object

Request for the order of the TE LSP

- To minimize disruption in case of bulk provisioning:
  - A new flag (D flag) in the RP object
- To support the determination of whether make-before-break optimization is required:
  - A new flag (M flag) is defined in the RP object.
Global Concurrent Optimization

PCEP Protocol Extensions for GCO

The Order Response

- New optional TLV defined in the RP object carried in the PCRep Message in response to a reoptimization request.
- Indicates the order in which the old TE LSP must be removed and the new TE LSP must be setup.

```
<table>
<thead>
<tr>
<th>Type</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Delete Order</td>
<td>Setup Order</td>
</tr>
</tbody>
</table>
```

Type: 5
Length: Variable
**Delete Order:** Integer that indicates the order in which the old TE LSP should be removed.
**Setup Order:** Integer that indicates the order in which the new TE LSP should be setup.
Global Concurrent Optimization

PCEP Protocol Extensions for GCO

Global Constraints (GC) Object

- Used in a PCReq message to specify the necessary global constraints that should be applied for a global concurrent path optimization request.

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

MH (Max Hop): Integer that indicates the maximum hop count for all the TE LSPs.

MU (Max Utilization Percentage): Integer that indicates the upper-bound utilization percentage by which all links should be bound.

mU (minimum Utilization Percentage): Integer that indicates the lower-bound utilization percentage by which all links should be bound.

OB (Over Booking factor Percentage): Integer that indicates the overbooking percentage that allows the reserved bandwidth to be overbooked on each link beyond its physical capacity limit.
Global Concurrent Optimization

PCEP Protocol Extensions for GCO

Error Indicator

• To indicate errors associated with the global concurrent path optimization request a new Error-Type = 14 value is defined.
• New Error-Type = 15; Error-value=1: if memory overflow.
• New Error-Type=15; Error-value=2: if GCO is not capable.
• Error-Type =5; Error-Value = 3: to indicate “GCO not allowed under policy.”
• Error-Type=5; Error-value=5: if is not compliant with administrative privileges

NO-PATH Indicator

• To communicate the reasons for not being able to find GC path computation.
• Two new bit flags are defined in the NO-PATH-VECTOR TLV:
  o Bit 6: Indicates that no migration path was found.
  o Bit 7: Indicates no feasible solution was found that meets all the constraints associated with global concurrent path optimization.
• Software Defined Networking is emerging as an extensible and programmable open way of operating networks
  – Its main concept is the decoupling of forwarding and control functions
  – It centralizes network intelligence and state information, while providing to the upper layers an abstracted and vendor independent view of the network

• SDN allows network providers to build more scalable, agile and easily manageable networks
  – provides a software abstraction of the physical network that allows the network itself to be programmable
  – ... and closely tied to emerging applications and services needs

• SDN provides open interfaces that enable the development of software that can control the connectivity provided by a set of network resources and the flow of network traffic through them
  – such primitive functions (APIs) may be abstracted into arbitrary network services at the so-called Application to Controller Interface (A-CPI)

• SDN can also abstract heterogeneous technologies employed in the data plane and represent them in a unified way
  – this way it becomes an integrated control plane supporting heterogeneous technologies employed in different layers (e.g. optical layer, Ethernet, and IP layer)
SDN components

Ref. ONF SDN architecture 1.0
PCE framing into SDN

• PCE architecture natively decouples the path computation from the forwarding plane

• ...and the PCEP is already an open standard protocol for communication with well-defined dedicated computation engines/elements
  – this opens wide opportunities for integration of PCE with diverse control plane architectures beyond its original MPLS/GMPLS scope

• SDN can indeed highly benefit from PCE
  – PCE can offload path computations to dedicated engines/elements and assist SDN controllers for their base services
  – SDN can leverage on mechanisms and procedures for cooperation among diverse PCEs in multi-domain and multi-layer scenarios
  – integration of PCE in SDN allows well-defined and ready-to-use routing algorithms to be adopted
  – network operators migrating towards SDN can avoid to waste solid expertise and know-how in the PCE area
Mainly two options are available to integrate PCE within SDN, depending on PCE capabilities available:

1. **Stateless or passive stateful PCE**
   - can be an external application of the SDN controller (i.e. in the application plane) accessible through PCEP
   - topology/LSP information retrieved by the PCE through a dedicated set of SDN controller northbound APIs

2. **A stateful PCE with LSP initiation capabilities**
   - can be itself considered a kind of (partial) SDN controller
     - Implements functions for provisioning, modification and deletion of LSPs
   - integration with a full SDN controller may happen by means of internal interfaces with shared topology/LSP information
Application-Based Network Operation (ABNO)

• Application-Based Network Operations
  – A PCE-based Architecture for Application-based Network Operations
  – draft-farrkingel-pce-abno-architecture

• Network Controller Framework
  – Avoiding single technology domain “controller” architecture
  – Reuse well-defined components and protocols
    • Discovery of network resources and topology management.
    • Routing and path computation
    • Multi-layer coordination and interworking
    • Policy Control
    • OAM and performance monitoring

• Support a variety of southbound protocols
  – Leveraging existing technologies, support new ones

• Integrate with defined and developing standards, across SDOs
ABNO
Functional Components

- "Standardized" components
- Policy Management
- Network Topology
  - LSP-DB
  - TED
  - Inventory Management
- Path Computation and Traffic Engineering
  - PCE, PCC
  - Stateful & Stateless
  - Online & Offline
  - P2P, P2MP, MP2MP
- Multi-layer Coordination
  - Virtual Network Topology Manager
- Network Programming and Signalling
  - ForCES
  - OpenFlow
  - Interface to the Routing System
  - PCEP
  - RSVP-TE

Figure 1: Generic ABNO Architecture
Compare ABNO with SDN Architecture

- A richer function-set based on the same concepts
- Enables the use of OpenFlow and other protocols
- There are implementation/deployment choices to be made

Minimum required for SDN controller of infrastructure

What is required for commercial deployment of SDN control platforms for real networks
ETICS core system: an SDN example

ETICS User Interface

SLA Offers (Push)
- Builders
- Controller
- Publish

Rules & Bill DB

Business & Policy
- AAA
- Billing
- Policy Rules

Policy Rules

SLA DB

SLA Manager
- Service Composition
- Validation
- Termination
- Service Assurance

Monitoring

Service Instance Manager

Routing & Topology

Network Service Provider

Traffic Engineering

Network Capabilities (Pull)
- IC Routing Controller
- IC Routing protocol

PCE

NC DB

ETICS User Interface

E1/E2-B

E6/E7

E1/E2-S

E1/E2-C

E1/E2-B

ETICS core system: an SDN example
PCE uses with SDN

• PCE can ease the orchestration of path computation and service provisioning in complex inter-domain transport scenarios
  – e.g. when legacy GMPLS/PCE domains are interfaced to SDN OpenFlow domains at the periphery
    • a parent PCE with stateful and active capabilities may coordinate end-to-end services
  – this may also apply to orchestration of intra and inter datacenter connectivity

• Integration of PCE and SDN architectures is covered in the IETF ABNO framework
  – ABNO provides an SDN based framework for on-demand and application-specific provisioning of network resources in a wide range of network applications and network technologies from packet to optical
  – PCE within ABNO provides application-aware path computations and policy enforcements for the set of services supported in ABNO
What is Segment Routing?

- Segment Routing depends on a set of segments that are advertised through the correspondent IGP routing protocol which are later combined together to create the desired path to the destination.
  - No Label Distribution Protocol (LDP)
  - No Resource Reservation Protocol with Traffic Engineering extensions (RSVP-TE)

- Two types of segments are defined:
  - **Node SID**: Path to a node (Global and unique)
  - **Adjacency SID**: Local service at a certain node (Non unique)

- Three types of operations are defined: push, continue (swap) and next (pop) → As in MPLS!!

- Use of the existing MPLS data-plane: No new operations are introduced.
SR implications in signaling

• The interaction with the PCE does not change for SR:
  – Passive PCE.
    • PCReq/PCResp sequence from the PCC.
  – Active PCE.
    • PCInitiate/PCNotify from the PCE.

• However, the signaling is different:
  – Without SR:
    • Edge Router propagate RSVP-TE signaling after the Path Response or the Initiate Message.
  – With SR:
    • Edge Router locally instantiates the labels for the path.
SR extensions for PCEP

• **PCEP must be extended**
  – The edge router of the SR path incorporates in all packets a set of tags containing all the hops that the packet is going to traverse.
  – This information must be received from the PCE.

• **Four extensions are required:**
  – **SR Capability negotiation:** When a PCEP session between different PCEP entities is brought up, both parties exchange information that announces their ability to support SR.
  – **Path Set-up type:** As a PCC can signal through multiple protocols the LSP, it must indicate whether SR or RSVP is used.
  – **Explicit Route Object:** SR tags in PCEP are codified as SIDs in the ERO present in the PCInitiate and PCReply messages.
  – **Record Route Object:** SR-RRO sub object is defined to record the SR path.
• The PCC indicates that it is capable of supporting the head-end functions for SR-TE LSP adding the SR-PCE-CAPABILITY TLV in the OPEN message destined to a PCE.

• The MSD is the maximum number of SIDs that a PCC is capable of imposing on a packet.

• The SR Capability TLV is meaningful only in the OPEN message sent from a PCC to a PCE.
Path Set-up Type

• When multiple path (like SR and RSVP-TE) setup methods are deployed in a network, a given PCEP session may have to simultaneously support more than one path setup types.

• The PATH-SETUP-TYPE must be included in the RP/SRP Object within the PCReq or PCIInitiate message for SR paths.

• The PATH-SETUP-TYPE is defined as:

To use SR, the Path Set-up Type (PST) in PATH-SETUP-TYPE is set to 1.
Explicit Route Object

- An SR-TE path consists of one or more SID(s) where each SID represents the node or adjacency. This identifier is referred to as the 'Node or Adjacency Identifier' (NAI).

- To signal the SR path, the ERO must contain a SR-ERO subobject with SIDs or NAI.

- The fields in the SR-ERO Subobject are:
  - The 'L' Flag indicates whether this is a loose-hop in the LSP.
  - Type is the type of the SR-ERO subobject and Length contains the total length of the subobject in octets.
  - SID Type (ST) indicates the type of information associated with the SID contained in the object body.
  - NAI contains the NAI associated with the SID.
• There are four NAIs:
  – IPv4 Node ID
  – IPv6 Node ID
  – IPv4 Adjacency
  – IPv6 Adjacency
  – Unnumbered Adjacency with IPv4 Node IDs
• A PCC can record SR-TE LSP and report the LSP to a PCE via RRO. A SR-RRO subobject must be included:
Network Functions Virtualization (NFV) consolidates many network equipment types onto industry standard high volume servers, switches and storage, which could be located in datacenters, network nodes and in the end user premises.

NFV is applicable to any data plane packet processing and control plane function in fixed and mobile network infrastructures.

Basic terms:
- Virtual Network Function
- VNF Manager
- Orchestrator

NFV be applicable to Transport Networks Infrastructure
- Virtualize classical control processes, such as PCE ➔ vPCE
vPCE motivation

• Problem:
  – A single PCE may incur on scalability issues: the increase of computational load may worsen Path Computation response time.
vPCE approach

• We virtualize PCE Network Function
  – Several vPCE can be deployed on demand.
  – Perceived as a “unique” PCE for the users.
    • E.g. single entry point: pce.cttc.es (DNS)

  – NFV orchestrator responsible for monitoring vPCE computation load and deploy them in case of computation load increases.

• Flexible and scalable solution for providing Transport PCE NFV.
NFV PCE architecture
PCE in IoT

• Objectives
• The Rationale
• Methodology
• Context-Aware PCE
• Illustrative use cases
PCE in IoT: Objectives

• Positioning the PCE in an Internet of Things (IoT) scenario

• Describing the synergy between PCE, ID/LOC Split Architectures (ILSAs)

• Introducing the novel concept of Context-Aware PCE
• The communication model widely deployed in the current Internet architecture follows a host/location-oriented design, sweeping along several design limitations, undoubtedly restricting its deployment in an IoT scenario.

• Most of the issues related to the current host-oriented communication model boil down to the double functionality of addresses (i.e., the double functionality problem.)
• ILSA schemes deal with the double functionality problem by assigning an independent set of addresses for identification and location functions.

• ILSA schemes have received a great acceptance in network research. As a matter of fact, ILSA schemes are already available as commercial solutions.
PCE in IoT: Methodology

- The endpoint (address) of a Path Computation Request (PCReq) is coupled to an ID.

- The mapping between an ID to a locator(s) is done according to a given context information.

- Enhance current Path Computation Element Protocol (PCEP) in order to support both context-information and IDs based endpoints.

- Enable interaction between a PCE and an ILSA scheme
• CONTEXT-AWARE PCE is a PCE that resolves a PCReq according to a given context, e.g., traffic volume, trending topic, etc. The endpoints of a PCReq are obtained by interaction with an ILSA scheme.

• The paradigm of a context-aware PCE stems from the fact that the relevant is the “What” (“the service to be provided”) rather than the “Where” (“connection to a location or host”).

• In the use-cases presented in next slide a PCReq end-point is coupled to an ID. This ID is an virtual object that is not physically attached to a device, e.g., Wavelength Router (WR) or IP/MPLS router). An ID might represent a collection of devices with a common feature, e.g., data repositories, within a Data Center.
Two cross-layer connections associated to LSP A-B and one to LSP-A-C. PCC's policy is to allocate more network resources to the current Trending-Topic associated to LSP A-B.

ID Data-Repo is mapped to a locator belonging to the domain related to the given context (trending topic in this case).

Enhanced PCEP object
Source: Locator A.
Destination: ID Data-Repo
Connection-Context: Trending-Topic
Context-Parameter: Football
Bandwidth: …
Metric: …

High volume of queries related to keywords: football match, Barcelona vs. Madrid

PCReq for LSP A-Data-Repo

Time 18:00
Football match over, queries related to keywords: restaurant, dinner, Barcelona

Time 20:00

Football match over, queries related to keywords: restaurant, dinner, Barcelona
✓ It is not **unrealistic** to assume a periodic and highly dynamic behavior regarding trending-topics and internet traffic.

✓ The figure shown in this slide depicts the Queries distribution (daily) between June-September (2014) in Barcelona (Relative Values)*. Certain search topics such as movie theaters exhibit a drastic increase during weekends. The opposite occurs with search topics such as banks) This drives us to encourage the practice of dynamical allocation of network resources dedicated to a connection.

* www.google.com/trends/

The x axis is the accumulative for three days of a month, i.e., Monday=Data of three Mondays for June-September.
• There is an increasing deployment of smart devices coining the so-called **Smart Cities** – a concept encompassed by the IoT.

• Upon the composition of a **smart-service**, physical connectivity to the data sources must be established.

• Traditionally, both service composition and path computation have been two processes **independently** from each other.

• Combine Apps devoted to service composition and a context-aware PCE scheme into a collaborative ecosystem, where a context-aware PCE computes a path based on both the **Context-Aware Graph and Transport Network Graph**
Context-Aware Graph

<table>
<thead>
<tr>
<th>IDs</th>
<th>LOCATORS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Movie-Theaters DB</td>
<td>A</td>
</tr>
<tr>
<td>GIS Provider 1</td>
<td>E</td>
</tr>
<tr>
<td>GIS Provider 2</td>
<td>C</td>
</tr>
<tr>
<td>Taxis DB</td>
<td>D</td>
</tr>
</tbody>
</table>

**Context-Aware Graph**

**Transport Network Graph**

**Selected Path**: Movies-Theater DB, GIS Provider 1, Taxis-DB

**Path 1 cost (3)**: Movies-Theater DB, GIS Provider 1, Taxis-DB

**Path 2 cost (3)**: Movies-Theater DB, GIS Provider 2, Taxis-DB
Operational constraints

• Robustness of SDN or PCE based network controller
  – Must be reliable i.e. run safer 99,999% of the time
  – Not frozen the network
    • Many planned work on devices occurs in a carrier network

• Scalability and flexibility
  – Must scale on large network i.e. > 500 routers
  – Must handle large variety of commercial devices and technologies

• Security protection against attacks
  – PCEP over TLS is an on-going draft
  – FireWall could also protect PCE
  – How to open North bound interface to 3rd Party?
    • AAA is our friend

• Read RFC7149
  – “Software-Defined Networking: A Perspective from within a Service Provider Environment”
Operational constraints

• Reliability of network control
  – Path computation and LSP-TE enforcement must be completed by de-configuration of devices
    • Avoid dummy connectivity
    • Track un-synchronize state
  – Handle grace full restart
    • Not only after a crash but for all planed work

• Measurement and Monitoring tools
  – Operational entities needs operational and maintenance tools to monitor the network
  – Measurement will complete the Engineering of the network in particular to automatically adjust the size of LSP-TE

• Planning tools to smoothly handle growing of the traffic
  – Network controller could be help full to manage traffic matrix by automatically enforce LSP-TE

• Do Not promise that SDN will do all
  – Be Pragmatic and Design Carefully ➔ Do less, but efficiently
  – Add new features when they are mature
  – Carrier are not debugger
Standardization Status in IETF

• Orange is both...
  – co-creator of PCE architecture and PCEP (Jean-Louis Le Roux)
  – co-chair of the PCE working group (Julien Meuric)
• Available standard track RFCs
  – 5088, 5089: OSPF and IS-IS extensions for PCEP discovery
  – 5440: PCEP
  – 5441: Backward Recursive Path Computation Procedure (BRPC)
  – 5455: PCEP object for diffserv-aware class-type
  – 5520: Path key for inter-domain confidentiality
  – 5521: PCEP extension for route exclusion (XRO)
  – 5541: Encoding of objective functions in PCEP
  – 5557: Global concurrent optimization
  – 5886: Monitoring toolset
  – 6006: PCEP extension for P2MP paths
  – 6007: Synchronization vector list for dependent computations
  – 6805: Determination of Sequence of domain in MPLS or GMPLS
  – 7150: Vendor-specific constraints in PCEP
Work in Progress in Standardization

- Main PCE working group drafts
  - PCEP extensions for GMPLS
  - PCEP extensions for inter-layer
  - MIBs: PCEP, PCE discovery
  - PCE State full
  - Inter-domain P2MP procedure

- Some promising individual drafts
  - PCEP extensions for GMPLS-controlled WDM

- Beyond IETF: the Optical Internetworking Forum (OIF)
  - new project recently started: “use of PCE within ASON context”
  - to be used for routing across multiple optical domains
  - expectation: applicability without protocol extensions
What PCE Is Not...

• NOT a network management system knowing everything
  – LSP provisioning relies on signaling
  – does not (have to) know the states of LSPs, only of resources
  – only knows the topology(ies) fed to it
• NOT a full centralization of path computation
  – not more centralized than a local engine inside the head node
  – all nodes might be PCEs
  – the communication protocol may for instance be used only for inter-domain or multi-layer path computations
• NOT an architectural component, but a function
  – which can be integrated in network devices, servers…
  – the communication protocol allows to open node internal path computation engine to remote network devices
    • MPLS-TE/GMPLS node + comm. protocol => PCE node
Current Limitation of PCE Standards

• In inter-AS scenarios, the AS chain may be computed from...
  – BGP routes
    • but no TE information beyond the 1st AS (unlikely to change)
    • the AS path may be non-optimal from a QoS perspective
  – BRPC
    • BRPC requests may be sent to all connected domains
    • but does not scale beyond a limited number of domains
  – hierarchical PCE knowing domain interconnection topology
    • turns PCE chain into hub&spoke (“parent&child”) dialogue

• However, from the standardization point of view
  – PCE deployments for multi-AS routing are not so common
  – the number of PCE-enabled traffic-engineered ASes which are currently interconnected together is low

• The major issue remains the TED and LSP-DB synchronization
  – Without a perfect knowledge of the underlying network the PCE is blind
Perspectives

• Main efforts must go to the Topology acquisition
  – Inter-domain and multi-layer
  – LSP-DB synchronization

• State Full PCE is an on going work
  – Very promising for SDN Controller
  – But many major issues must be solve

• SDN could be the main target for PCE deployment
  – Up to know there is no real PCE deployment within carrier
  – PCE remains a research subject with few real commercial offers
  – Tool box must be enhanced to complete the architecture

• Segment Routing will be another main target for PCE
  – Help in computing multi-hop segment
  – Mandatory to continue TE with Segment Routing network
Explicit route exclusions


Path confidentiality


Objective functions


Global Concurrent Optimization


