Inter-Domain Connectivity Setup by OpenDaylight

Dr. Olivier Dugeon, Orange Labs

olivier.dugeon@orange.com

Orangeexpert
future networks

orange™

Outline

Context & Use Cases

Backward Recursive Path Computation a.k.a RFC 5441

Inter-domain connectivity challenges

Operational constraints

Stitching Label principle

Stateful PCE as SDN Controller

Stitching Label in action

Experimental Prototype

Next Steps

Context & Use Cases

MEF Lifecycle Service Orchestration (LSO)

Proposed a multi-domain standard API to setup inter-domain Ethernet connectivity
Proof of concept done through experimentation with different operators

TMF standardised Open API

Multi-domain is also take into account

Inter Data Centre connectivity

Customers request complex services that should be deployed on different DC and thus connected
E.g. Extranet between Main Corporate and its subcontractor / subsidiary

VPN across multi-domain

Customers request VPN between their different Offices across different country

Factory 4.0

Emerging new market that requires multi-domain connectivity with high QoS

Per AS/Domain PCE

A PCE can subcontract a segment computation to another PCE

PCEP exchange between PCEs similar to a PCC-PCE dialogue

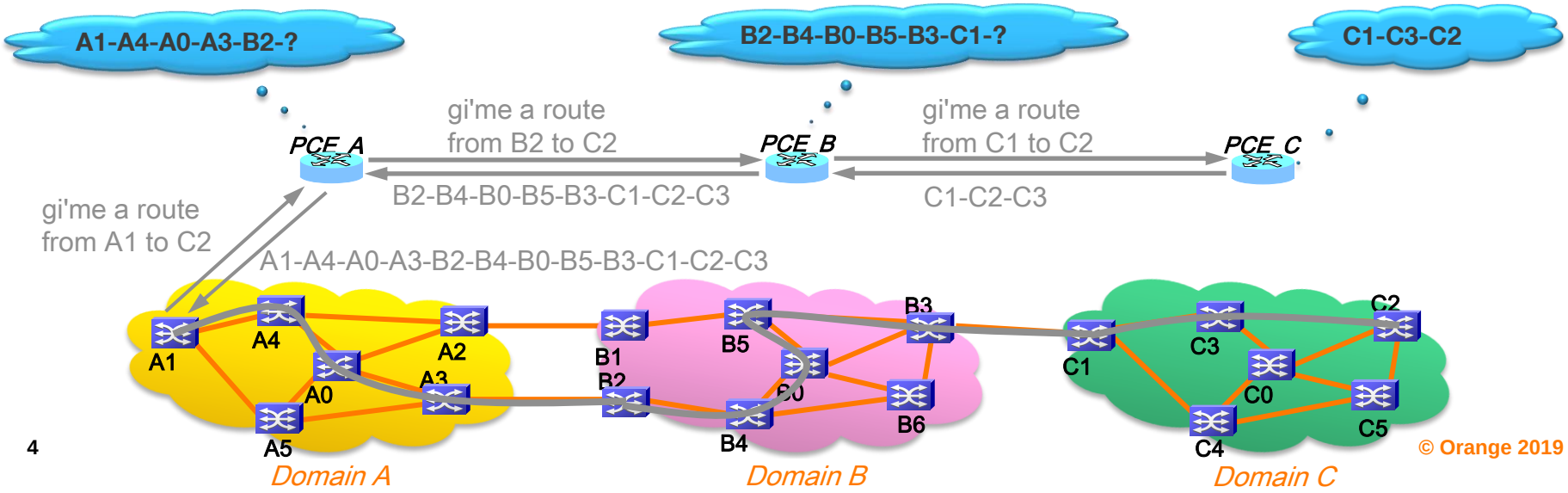
a PCE may act both as a server and as a client (PCC)

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 default BGP route



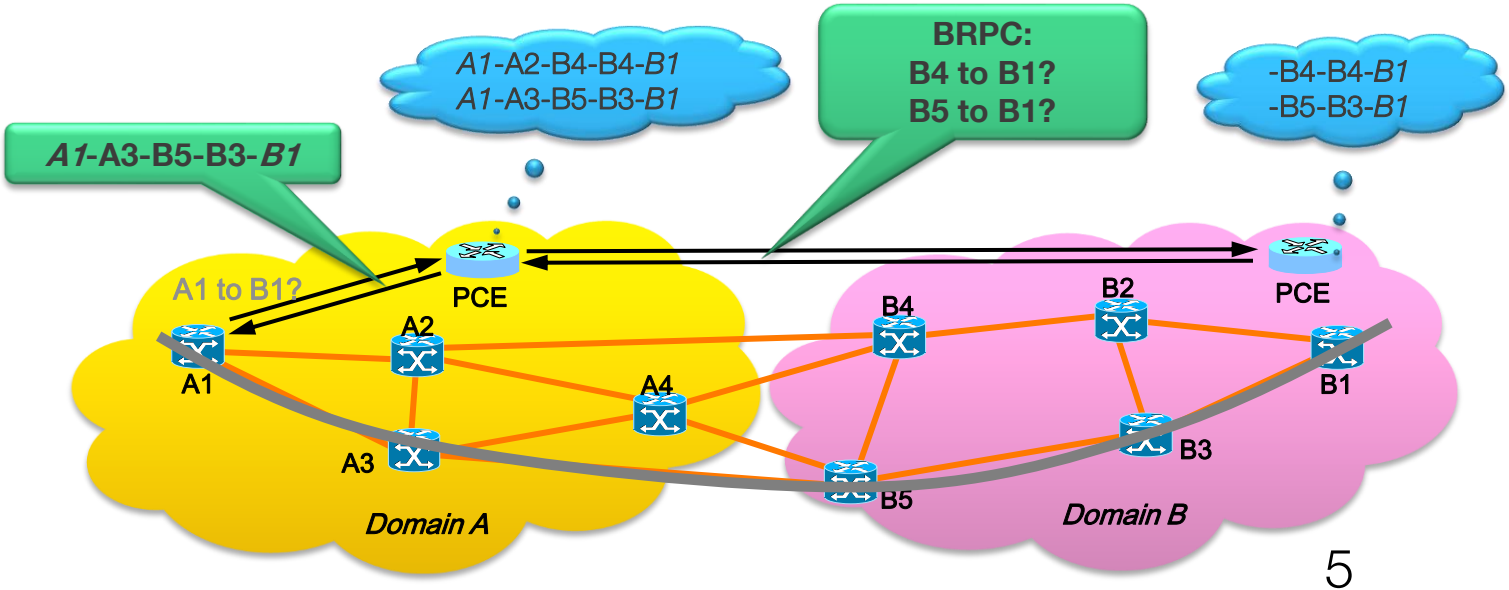
RFC 5441: Backward Recursive Path Computation (BRPC)

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



Path Key Confidentiality

Domain A might not want to expose the details of its internal topology

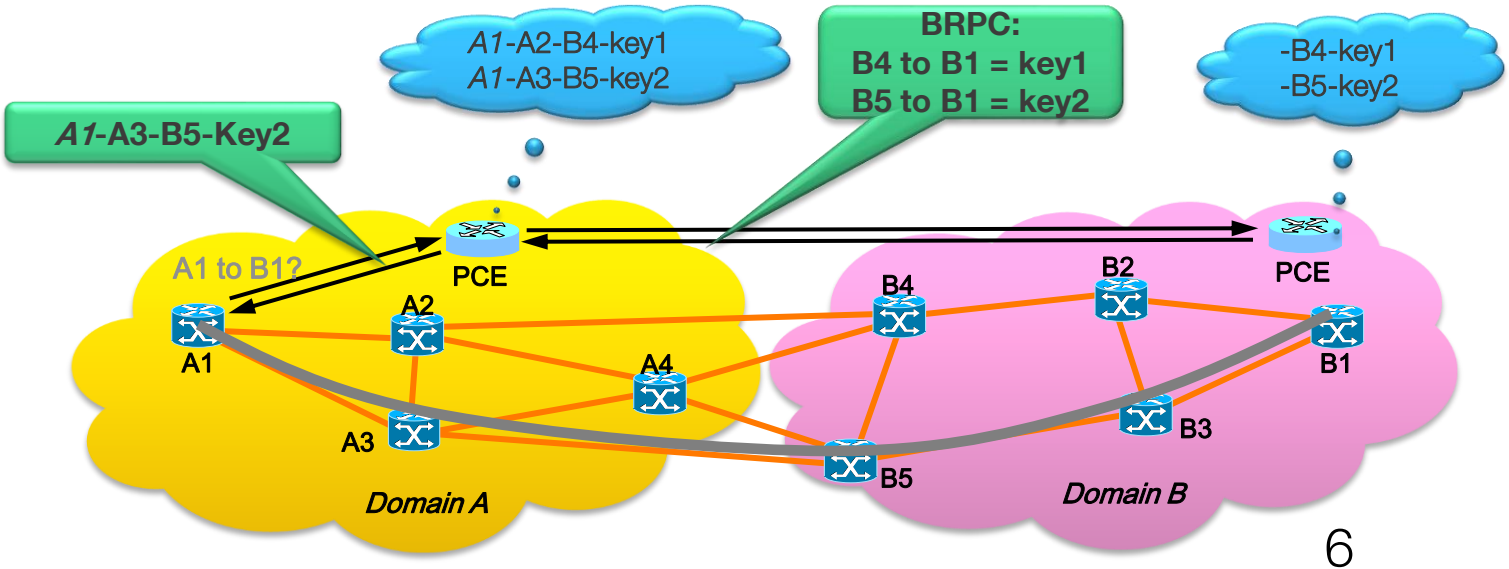
BRPC is enhanced with Path Keys

PCE-B returns {B4, key1, 4} and {B5, key2, 2}

PCE-A selects A1,A3,B5key2 (cost = 4)

Signalling reaches B5

B5 consults PCE-B with key2 for the path B5, B3, B1



6

Pseudo-Node Abstraction Model: A BRPC Helper

Abstract AS topology with one pseudo-node (SDN controller) and pseudo-links to BGP routers

Exchange information with BGP LS (RFC 7752 + RFC 8571)

Only Border Router with other Operator are announced

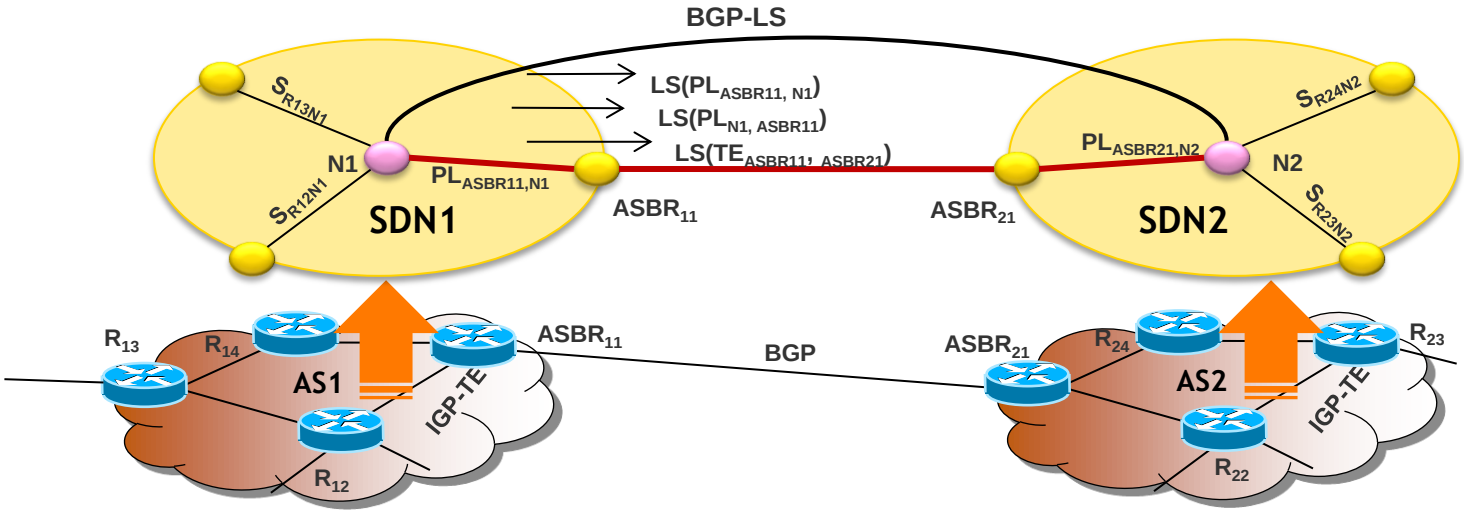
- **This reduce the number of announcement**

Create a new Hierarchical TED (H-TED)

Run Path Computation Algorithms on H-TED

Allow composition of SDN1 ↔ SDN2 link TE metrics

- **E.g. TE (N1, N2) = $PL_{N1,\ ASBR11}$ + $TE_{ASBR11,\ ASBR21}$ + $PL_{ASBR21,\ N2}$**
- **$PL_{ASBR21,\ N2}$ is learnt from SDN2**
- **Allow adjustment of TE link metrics**



Inter-Domain Connectivity Challenges

Control and setup tunnels at the inter-domain is a big challenge

Make differentiation between operators in their service offers

A good justification to increase revenue

A clear demands from Providers, especially for End-to-End and Cloud interconnectivity

Overprovisioning is no longer a viable solution

Peering points are becoming the bottlenecks

Needs sometimes to control where the connectivity is routed to meet local regulations (geopolitical routing)

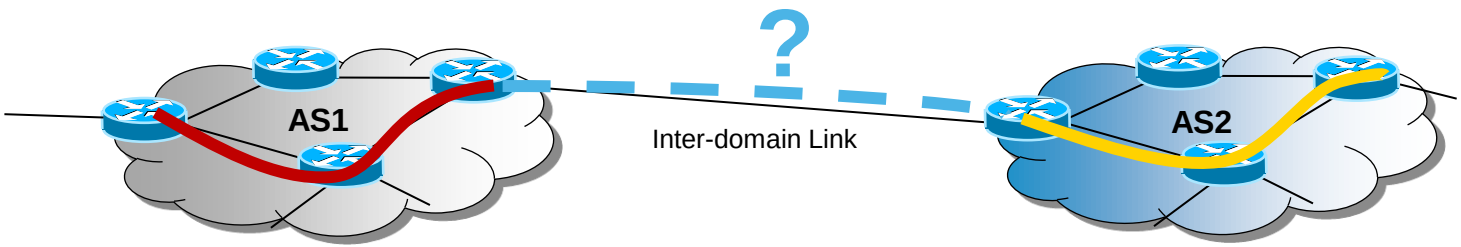
Several requirements must be addressed to setup inter-domain services

Let each operator manage independently their local tunnel (e.g., using RSVP-TE or Segment Routing)

Solve scalability issue to avoid too many RSVP-TE refresh messages

Solve technical issue to avoid large label stacks for Segment Routing

Enforce route selection at the peering point



Operational Constraints

Provide a solution for inter-domain connectivity with QoS

Operational constraints and lack of Business Agreement forbid usage of standard Traffic Engineering tooling avoiding inter-domain QoS connectivity deployment

Operators need to manage their part of multi-domain connectivity independently

Otherwise, network operations could be frozen

Commercial and technical upgrade of the connectivity offer must remain

Contiguous tunnel is not recommended

Security issues

Risk to impose constrains to the following network in the AS chain

Stitching or nesting per-AS tunnels is preferred

Allow independent tunnel configuration in each domain

Tunnel hierarchy solve scalability issues and allow smoother operation

How to exchange labels at inter-domain to connect per-AS tunnels ?

RSVP-TE is not authorized between BGP routers, mostly for security reasons

Same per-AS constraint with Segment Routing

Stitching Label principle

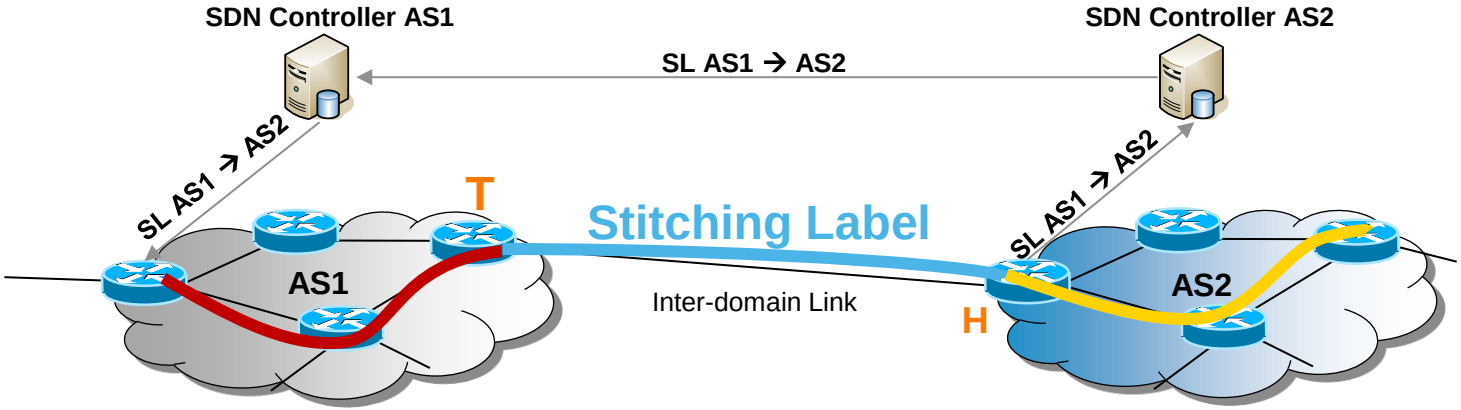
Idea: Exchange a dedicated label between per-domain centralized controllers

To stitch or nest two tunnels (RSVP-TE LSP or Segment Path)

From Destination to Source, i.e. Backward (to follow the typical MPLS downstream allocation)

Steps of the operation

- **Head of tunnel (H) in the downstream domain determines the Stitching Label with its upstream domain**
- **H send this Stitching Label to its corresponding SDN controller**
- **The controller sends the Stitching Label backward to its neighbour controller**
- **The upstream SDN controller pushes the Stitching Label as part of tunnel information to its head router**
- **The Stitching Label is used by the tail of tunnel (T) to stitch or nest its tunnel to the next**



Stateful PCE as SDN Controller

Take benefit of recent stateful enhancements to PCEP

- **PCInitiate message from PCE to head node to setup the per-AS tunnel as usual**
- **PCInitiate message between PCEs to trigger multi-domain service and request Stitching Label**
- **PCReport message to send Stitching Label and maintain the synchronisation between the PCEs**
- **PCUpdate message could be use from PCE to PCE (and per-AS head node) to modify the end-to-end tunnel**

Smooth exchange of label between per-domain PCEs using PCEP

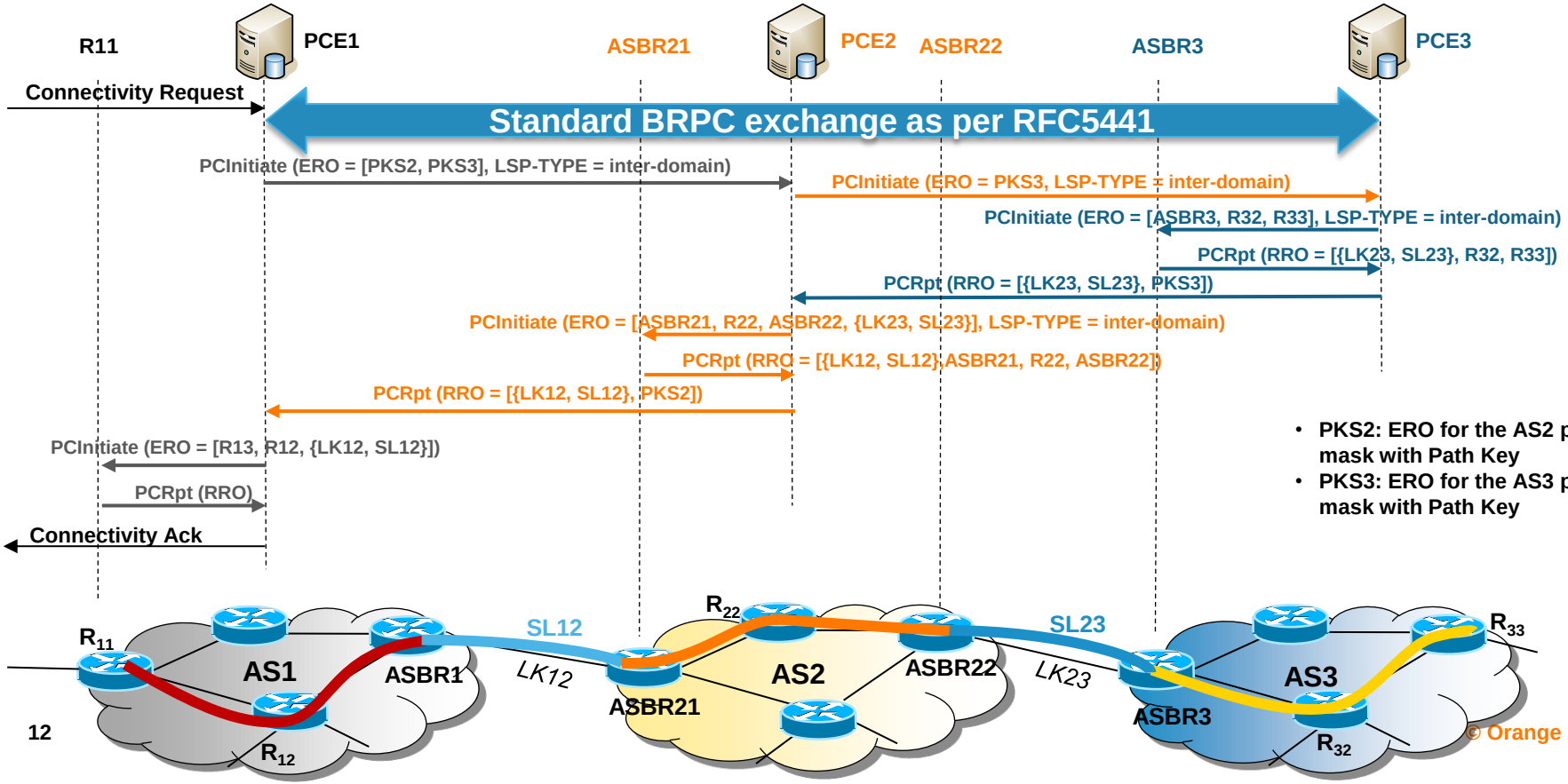
- **Done through a dedicated 'Stitching Label'**
- **Conveyed in ERO and RRO as label sub-object (RFC 3473/4003)**
- **Introduced new LSP-TYPE code points**
  - Feature defined in draft-ietf-pce-lsp-setup-type (soon RFC)
  - For PCE to PCE tunnel setup: to trigger a multi-domain service using Stitching Label (PCInitiate)
  - For PCE to PCC tunnel setup: to push tunnel information (PCInitiate) and…
    - include a downstream Stitching Label [except from destination domain]
    - request an upstream Stitching Label, to be included in the response (PCReport) [except from source domain]

Proposed draft to PCE WG: https://datatracker.ietf.org/doc/draft-dugeon-pce-stateful-interdomain

- **Waiting for WG adoption**

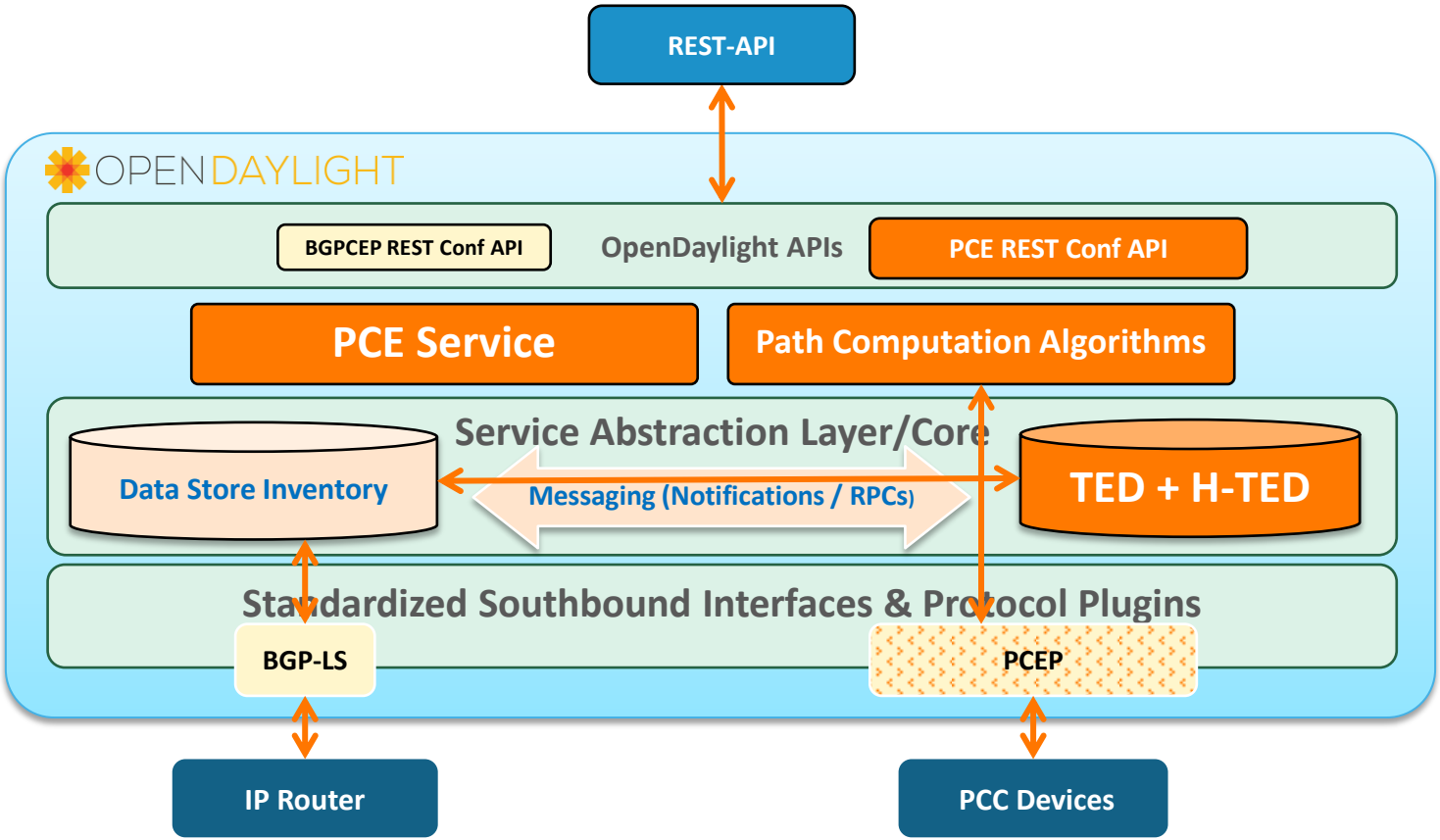
Stitching Label in action

- SL12: Stitching Label used by ASBR21 to identify the traffic coming from ASBR1 that stich the 2 tunnels
- SL23: Stitching Label used by ASBR3 to identify the traffic coming from ASBR22 that stich the 2 tunnels



- PKS2: ERO for the AS2 part mask with Path Key
- PKS3: ERO for the AS3 part mask with Path Key

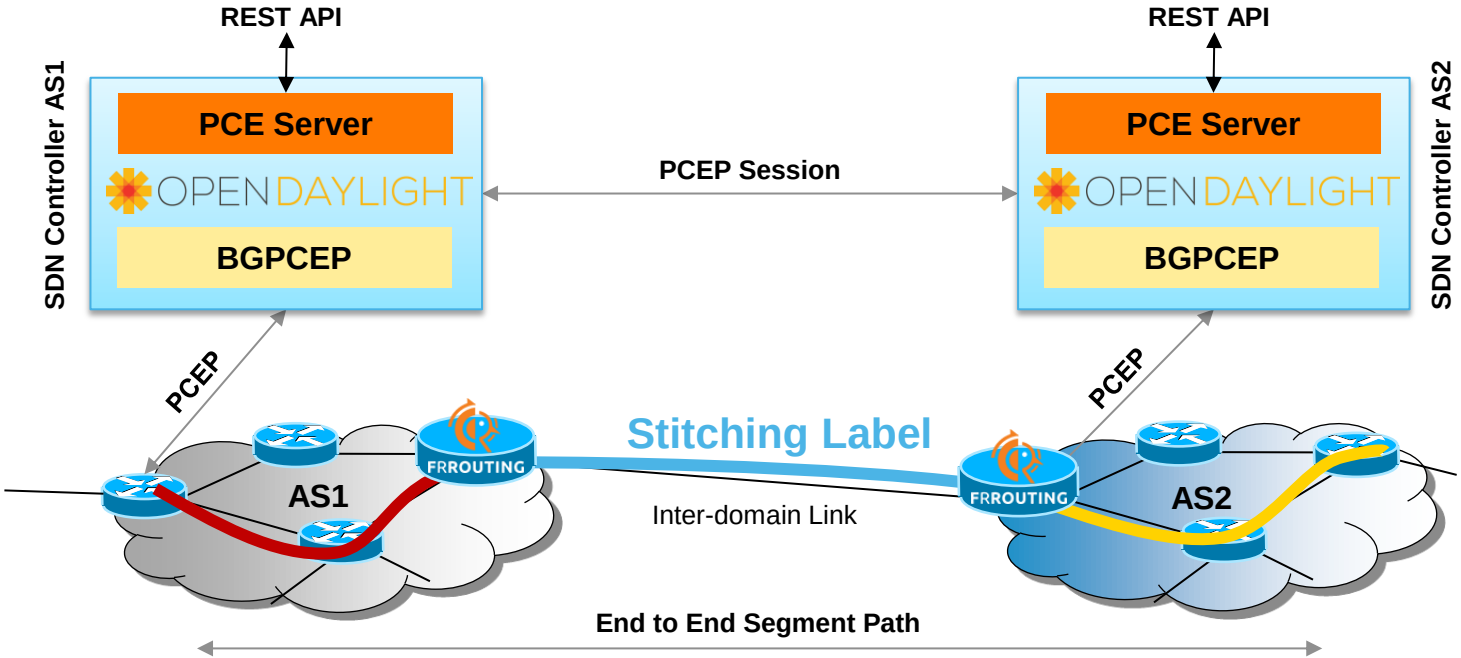
PCE architecture



Experimental Prototype

Tunnels are set up as Segment Paths using Segment Routing

- OpenDaylight has been enhanced to behave as a full PCE server (algorithm + BRPC + LSP-TYPE)
- Border Router uses latest FreeRangeRouting with OSPF Segment-Routing
- Python PCC client + FRRouting CLI enforces Segment Path and Stitching Label

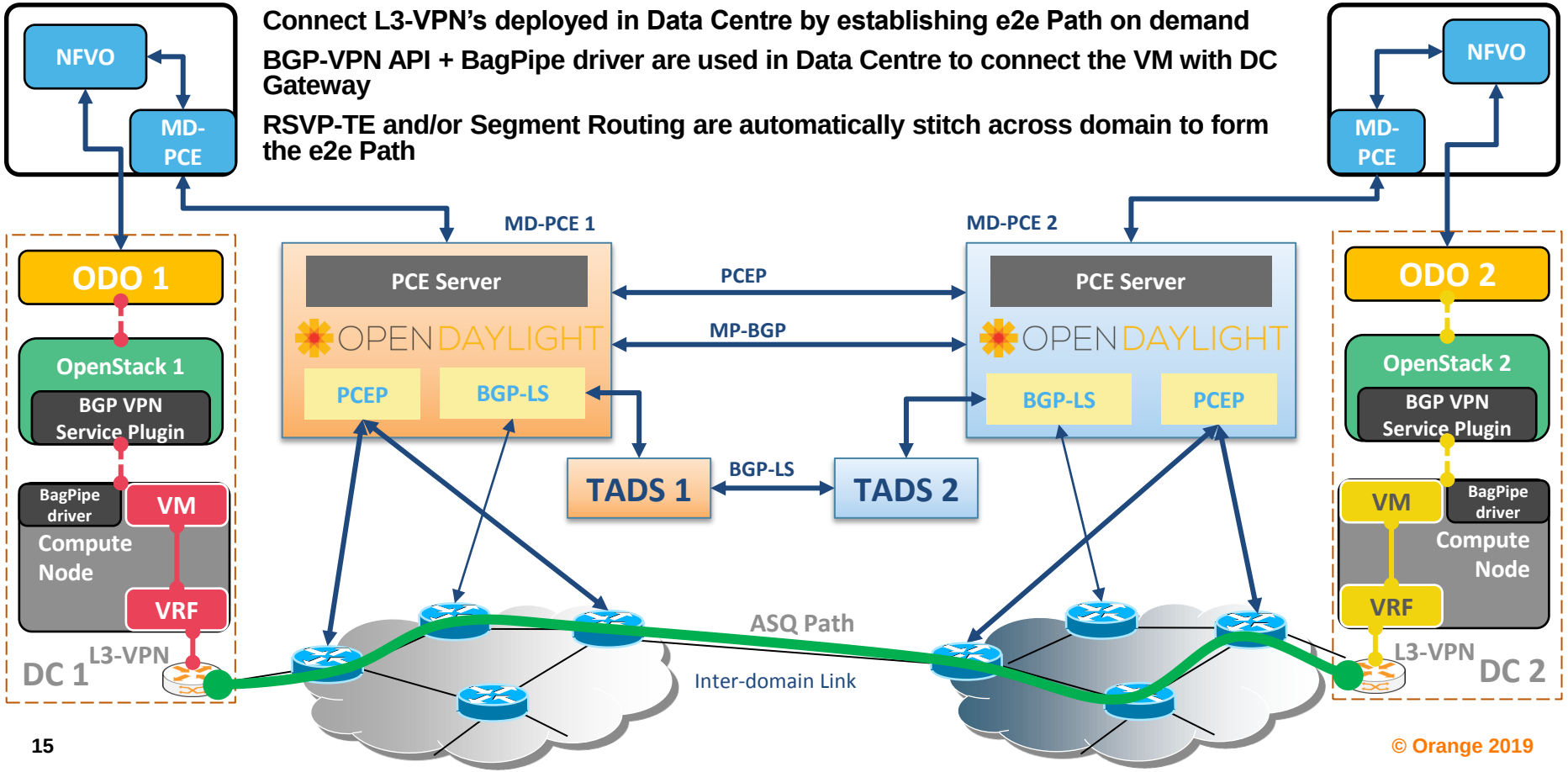


Demo (see it on youtube): Inter-Data Centre Use Case

Connect L3-VPN's deployed in Data Centre by establishing e2e Path on demand

BGP-VPN API + BagPipe driver are used in Data Centre to connect the VM with DC Gateway

RSVP-TE and/or Segment Routing are automatically stitch across domain to form the e2e Path



Development issues (1/2)

PCE to PCE session versus PCC to PCE session

PCE server needs to know if the incoming PCEP session come from a neighbour PCE or a PCC

- **PCEP session capabilities should be help**
- **Need to add Inter-domain capabilities**

PCE Session Manager needs to be enhance with session capabilities

If inter-domain capabilities is received

- **Check that the PCE neighbour is known (similar to BGP)**
- **Send back inter-domain capabilities**

If no capabilities is received

- **This is a standard PCC**

Development issues (2/2)

BRPC needs to wait answer from neighbour PCE

Implementation use synchronization semaphore + timeout

Is Future <> a better solution ?

BRPC Stateful defined new association group

But association group is not yet implemented

Next Steps

Clean & Rebase code on master

Publish alpha version on github prior to submit code to ODL Gerrit to get feedback

Add Association Group

Needed to group local and remote LSPs

- **To manage LSP life cycle including removal**

Add New PCE Session Capabilities

Needed to detect if incoming PCEP session comes from a PCC or a PCE

Improve PCEP Session establishment

Mimic BGP session handler

- **Do we need to separate PCEP Session handler like BGP do?**
- **Be more pro-active**
  - When a PCE neighbour connect, stop trying to connect to it like BGP do

Add Topology Abstraction

- **Automatically computed topology abstraction**
  - Add listener to BGP IPv4/IPv6 to detect AS & Border routers
- **Used BGP facilities to export abstract model**
  - Remove external BGP-LS speaker

Thanks

This work has been performed in the framework of the H2020-ICT-2014 project 5GEx http://www.5gex.eu (Grant Agreement no. 671636), which is partially funded by the European Commission.