Overlay Network - Path Computation Approaches
draft-bardalai-ccamp-overlay-path-comp-00

Abstract

   This document discusses various path computations approaches which
   are applicable to overlay networks [framework doc ref]. It discusses
   how the customer edge nodes uses the information advertised by the
   provider network to compute a path between two customer edge nodes or
   how it can request the provider network to compute a path and setup
   an end-2-end LSP.

Status of this Memo

Copyright and License Notice

Table of Contents

1 Introduction

This document attempts to describe possible ways to advertise
information required for customer network CE nodes to compute a path
for LSPs between two points in two customer network islands connected
by a provider network, so as to adhere a set of constraints in
provider network without knowledge of the detailed provider network
topology. These constraints could be, but not limited to, diversity,
latency, jitter, skew etc. Connectivity between customer network
islands is presumed to be an "overlay" over provider network.

1.1 Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT",
"SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this
document are to be interpreted as described in RFC 2119 [RFC2119].

2. Path Computation Use-cases

In case of overlay networks it is required to compute a path between
the customer edge nodes for the overlay FA-LSP as shown in the
figure.

```
            |<========= Overlay FA-LSP ============>|
            |                                       |
            V                                       V
 +---+ __ +---+        _           _         +---+ __ +---+
 |   |/  \|   |    +---+ _/ \_ +---+ _/ \_ +---+   |   |/  \|   |
 |C1 |    |CE1|---|   |/     \|   |/     \|   |---|CE2|    |C2 |
 |   |\__/|   |   |   |   |   |   |       |   |   |   |   |\__/|   |
 +---+    +---+   |   |   |   |   |       |   |   |   +---+    +---+
                  |PE1|       |P1 |       |PE2|
 +---+ __ +---+   |   |   |   |   |       |   |   |   +---+ __ +---+
 |   |/  \|   |   |   |   |   |   |       |   |   |   |   |/  \|   |
 |C3 |    |CE3|---|   |\_   _/|   |\_   _/|   |---|CE4|    |C4 |
 |   |\__/|   |   +---+  \_/  +---+  \_/  +---+   |   |   |\__/|   |
 +---+    +---+                                      +---+    +---+
```
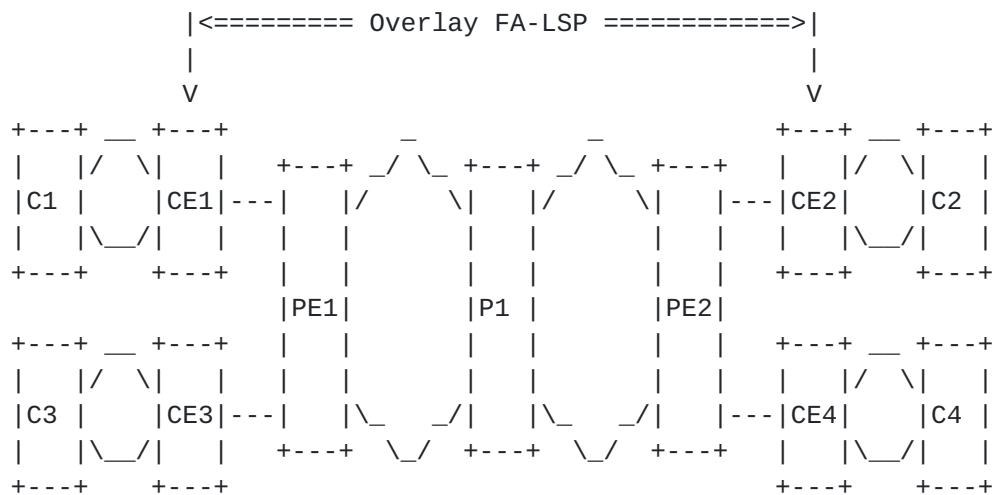
Figure. 1

The typical path computation use-cases are the following:
1. Point-to-point overlay path.
2. Multiple point-to-point diverse overlay paths sharing common LSP
   head and tail ends.
3. Multiple point-to-point diverse overlay paths that do not share
   common LSP head and tail ends.
4. Point-to-multipoint overlay paths.

5. Overlay paths over multi-domain (i.e. Multi-area or multi-AS)
   provider networks.

The typical TE constraints are:
1. Bandwidth or resource (this is technology specific).
2. Include or exclude nodes/links/SRLG or paths identified by
   path-keys.
3. Latency, jitter, max-hop requirements.
4. Optimization options - minimize cost, minimize latency etc.

[3](3). Path Computation Approaches

There are three path computation approaches
1. Virtual-topology approach
2. PCE approach
3. Hybrid approach - combined virtual topology and PCE approach

3.1 **Virtual Topology Approach**

   This path computation approach uses a virtual topology that is
   advertised by the provider network by the customer edge nodes.

```
   Customer Network                              Customer Network
      Island 1                                      Island 2

      __                                               __
     /  \                    Virtual Links            /  \
  C1 <    > CE1-----------PE1======= ========PE3------ CE3 <    >C5
     \C3/     ++++               \/            ++++++    \C4/
     /  \         +              /\             +       /  \
  C2 <    > CE2-----+---- PE2======= ========PE4-+-----CE4 <    >C6
     \__/     +++   +                           +    +    \__/
  ---------------- +--+--------------------------+----+-------
                   +  +        ____      ____     +    +
                   +  +       /    \    /    \    +    +
                   +  ++ PE1 <      >P1<      >PE3++    +
                   +       \_P3_/    \_P4_/          +
   Provider        +       /   \    /   \           +
   Network       ++++++ PE2<      >P2<      >PE4+++++++
                        \____/    \____/
```
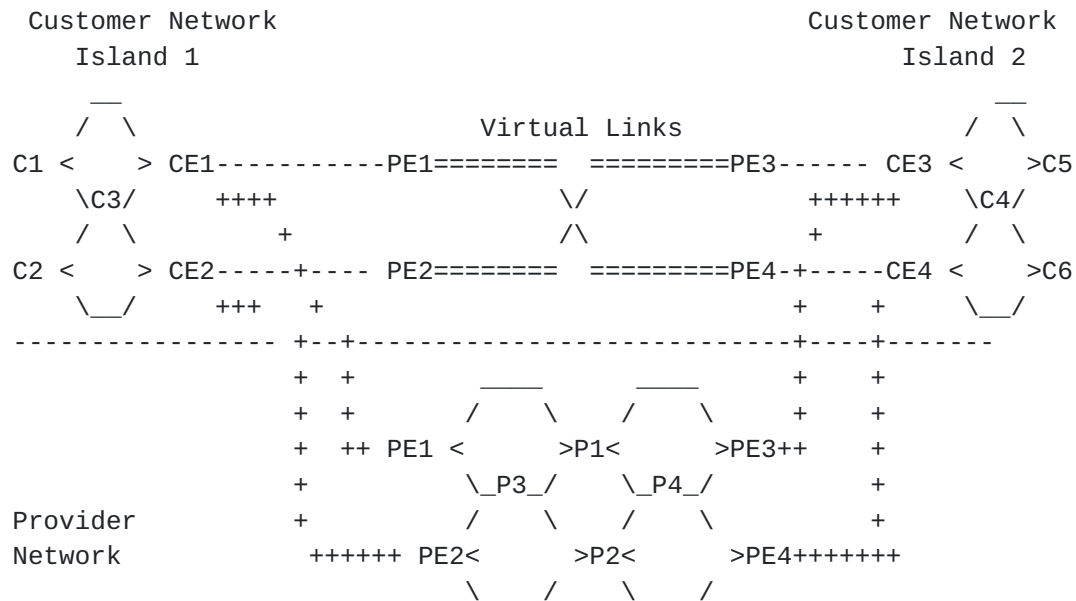
                          Figure. 2

   In Figure 2, Provider Network has 4 interconnected rings supports
   full node diversity to connect any 2 Provider Edge Nodes.

   PE1, PE2, PE3, PE4 are provider edge nodes.
   P1, P2, P3, P4 are internal provider Network nodes, that must not be
   known to the customer network.

   Customer Network has two islands connected by provider network.
   C1, C2, C3, C4, C5, C6 are internal customer network nodes.
   CE1, CE2, CE3, CE4 are customer network edge nodes connected to
   provider network edge nodes PE1, PE2, PE3, PE4.

   Virtual Link Set : Virtual Link set is defined as set of one or more
   virtual links between any two provider edge nodes. The virtual links
   in the virtual link set, when realized may take different paths
   within provider domain, having different SRLGs and other TE metrics.

   Above example topology has following Virtual Link Sets
   a/ [PE1, PE2]
   b/ [PE1, PE3]
   c/ [PE1, PE4]
   d/ [PE2, PE3]

        e/ [PE2, PE4]
        f/ [PE3, PE4]

    The PEs in provider network do full peering with its attached CEs for
    virtual topology. So provider network virtual Links along with its
    SRLG IDs and other TE metrics are advertised into customer network.

    Customer network internal Nodes C1..C6 can see provider network
    virtual TE Links and can compute paths between two points in customer
    network islands across provider network satisfying required diversity
    and TE metrics.

3.2 PCE Approach

    An alternative approach for a CE node to obtain a path to another
    remote CE node would be by making a request to a provider network
    PCE. This approach requires either provider network PE nodes to
    advertise the PCE's IP address to CE nodes or CE Nodes should be
    configured with Provider Network PCE IP address. CE nodes needs to
    advertise the TE link-state of the CE-PE interface. This allows the
    PCE to build the overlay network topology link-state data-base.

```
   |<----------------- Client Layer LSP -------------------->|
   |                                                         |
   |          |<========= Overlay FA-LSP ============>|      |
   |          |                                       |      |
   |          |          |<- Server Layer LSP -->|    |      |
   V          V   x-NNI|                         |    V      V
+---+ __ +---+ |   V         _              _        V    +---+ __ +---+
|   |/  \|   | | V +---+ _/ \_ +---+ _/ \_ +---+    |   |/  \|   |
|C1 |   |CE1|---|   |/     \|   |/     \|   |---|CE2|    |C2 |
|   |\__/|   |  |   |   |         |   |   |    |   |   |\__/|   |
+---+    +---+  |   |   |         |   |   |    |   |   +---+    +---+
              |PE1|       |P1 |       |PE2|
+---+ __ +---+ |   |   |         |   |   |    |   |   +---+ __ +---+
|   |/  \|   | |   |   |         |   |   |    |   |   |   |/  \|   |
|C3 |   |CE3|---|   |\_   _/|   |\_   _/|   |---|CE4|    |C4 |
|   |\__/|   |  |   +---+  \_/  +---+  \_/  +---+    |   |\__/|   |
+---+    +---+                                       +---+    +---+
```
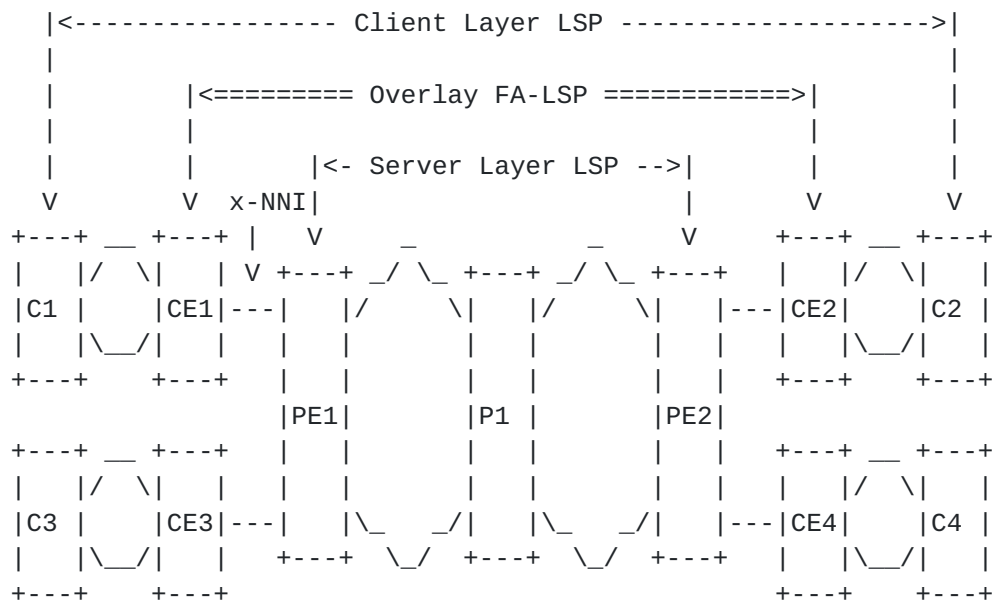
                            Figure. 3

    In Figure. 3 above, the example depicted shows the provider network
    with a single IGP area and the provider network PCE has visibility to
    the detailed topology and TE information representing the server
    layer forwarding plane plus the CE-PE interface link-states that have
    been learned from the CE nodes. The server layer topology in addition
    to the CE-PE interface link-states constitutes the overlay network

topology.

```
       |<----------------- Client Layer LSP -------------------->|
       |                                                         |
       |          |<========= Overlay FA-LSP ============>|      |
       |          |                                |      |      |
       |          |                  Server Layer   |      |      |
       |          |                |<-- LSP -->|    |      |      |
       V          V   x-NNI        |           |    V           V
    +---+ __ +---+ |       _       V    _       V   +---+ __  +---+
    |   |/  \|   | | V +---+ _/ \_ +---+ _/ \_ +---+ |   |/  \|   |
    |C1 |    |CE1|---|   |/     \| |/      \|   |---|CE2|    |C2 |
    |   |\__/|   | |   |   |     |  |       |  |   |   |\__/|   |
    +---+    +---+ |   |   |     |  |       |  |   +---+     +---+
                  |PE1|           |P1 |            |PE2|
    +---+ __ +---+ |   |   |     |  |       |  |   +---+ __  +---+
    |   |/  \|   | |   |   |     |  |       |  |   |   |/  \|   |
    |C3 |    |CE3|---|   |\_   _/|  |\_   _/|   |---|CE4|    |C4 |
    |   |\__/|   | |   +---+  \_/  +---+  \_/  +---+ |   |\__/|   |
    +---+    +---+                                   +---+     +---+
```
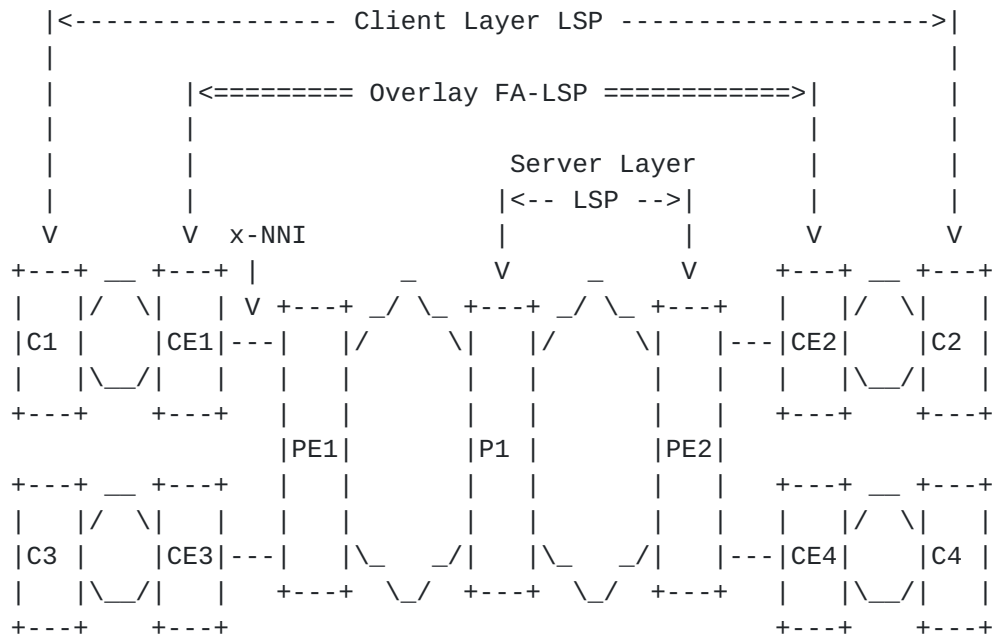
                          Figure. 4

Figure. 4 above shows the case in which the provider network is a
multi-layer network and the server layer boundary does not coincide
with the provider network boundary. Again, the provider network PCE
can have visibility to a single IGP area as described for MLN or
alternatively there could be multiple IGP instances as described in
[RFC6107], one instance for the overlay network and another instance
for the server layer.

```
      |<--------------------- Client Layer LSP ------------------->|
      |                                                            |
      |          |<============= Overlay FA-LSP ==============>|          |
      |          |                                            |          |
      |          |            |<------ Server Layer LSP ----->|          |
      V          V   x-NNI|                                   |      V          V
   +---+ _ +---+ |   V     _                       _    V    +---+ _ +---+
   |   |/ \|   | V +---+ _/ \_ +---+    +---+ _/ \_ +---+    |   |/ \|   |
   |C1 |   |CE1|---|   |/     \|   |    |   |/     \|   |---|CE2|   |C2 |
   |   |\_/|   |   |PE1|  D-1  |PE2|---|   |       |   |   |   |\_/|   |
   +---+   +---+  /|   |\_   _/|   |   |   |       |   |   +---+   +---+
              | +---+  \_/   +---+   |   |       |   |
              |          _        |   |PE5|  D-3  |PE6|
              | +---+ _/ \_ +---+   |   |       |   |
   +---+ _ +---+ | |   |/     \|   |   |   |       |   |   +---+ _ +---+
   |   |/ \|   | |/  |PE3|  D-2  |PE4|---|   |       |   |   |   |/ \|   |
   |C3 |   |CE3|---|   |\_   _/|   |   | ^ |   |\_   _/|   |---|CE4|   |C4 |
   |   |\_/|   |   +---+  \_/   +---+ | +---+  \_/   +---+   |   |\_/|   |
   +---+   +---+                     |                       +---+   +---+
                              x-NNI
```
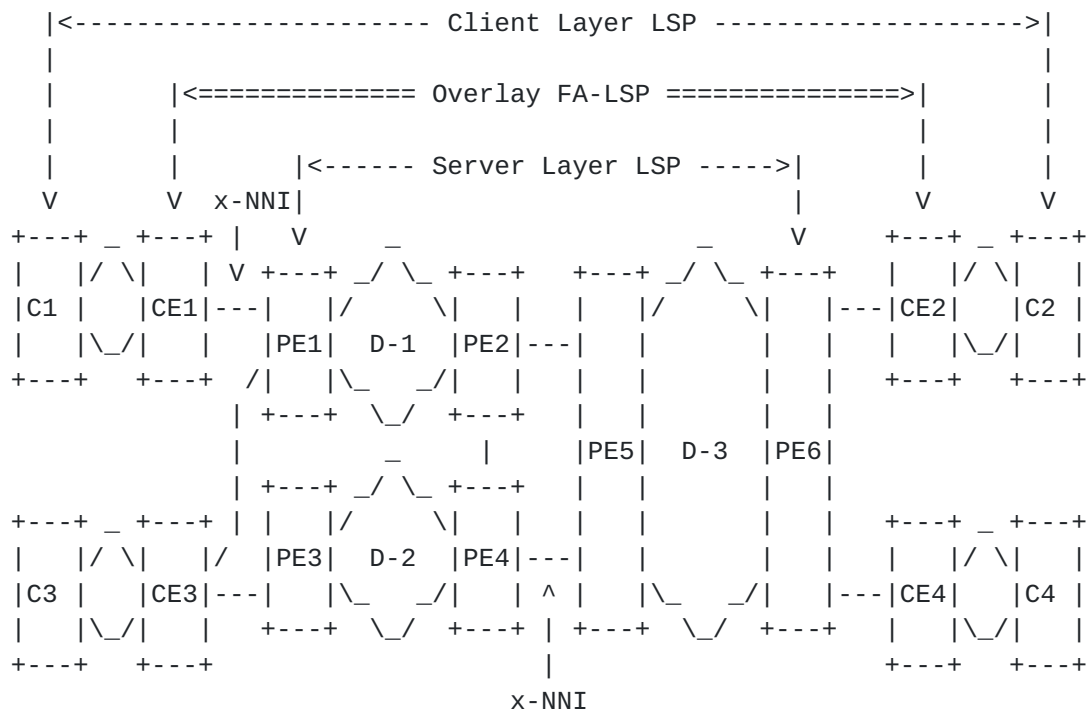
Figure. 4

Figure. 4 above shows a multi-area or multi-AS provider network
(generalized as a multi-domain provider network in this document).
For multi-domain networks a hierarchical PCE could be deployed and
the IP address of the hierarchical PCE is advertised to the CE nodes.
The hierarchical PCE could maintain a multi-domain virtual topology
instead of detailed topology of each domain.

In all three cases the head-end CE node is assumed to be aware of the
address in the remote CE node for which the path is to be computed.
The exact manner by which this knowledge becomes available is beyond
the scope of this document. The head-end CE node then makes a request
to the provider network PCE with the remote address and the required
set of TE constraints that need to be satisfied by the computed
path.

In each case of the provider networks PCE uses the overlay network
topology to compute the path. In case of the provider network example
shown in Figure. 4 the hierarchical PCE computes the domain-level or
inter-domain path first and then computes the intra-domain paths. The
exact mechanism could be using the BRPC procedure in order to compute
optimal intra-domain paths.

Once the computation is complete the PCE responds back with the path.
The path generated by the PCE is expected to contain both real and
virtual links and nodes. In case there is a need to maintain

confidentiality with respect to the details of the provider network
topology from the customer network then the response can include a
path-key. In case there is a need to compute diverse paths one of two
approaches could be followed - simultaneous computation approach in
which case the response will have multiple paths or path-keys or the
request could include the exclude hops or exclude path-key.

3.3 Hybrid Approach

In the absence of a hierarchical PCE for a multi-domain provider
network, it is possible a CE node learns of multiple PCE IP addresses
from multiple PE nodes. This is possible in case each PE node lies in
separate areas or ASs and with PCEs deployed per-area or per-AS. In
such a situation it will be necessary for the CE node to pick one of
the PCEs to send the path computation request. One way to select the
appropriate PCE would be to advertise a virtual-topology associated
with each PCE IP address to provide sufficient information for the CE
node to determine whether a path to the remote CE address can be
computed by the specific PCE.

In Figure. 4 above, CE3 has a dual-homed connectivity with the multi-
domain provider network (i.e. CE3 to D-1 and D-2 via PE1 and PE3
respectively). In the absence of a hierarchical PCE, PE1 can
advertise a virtual topology with connectivity to a set of CE nodes.
Similarly PE3 advertises a virtual topology with connectivity to
another set of CE nodes. This can happen in cases when there is no
available bandwidth to a specific CE node via a specific domain. CE3
can determine using the virtual topologies which PCE should it send
the path computation request.

4. CE-PE / PE-PE Interface

The CE-PE or PE-PE interface requires a routing interface in order to
be able to exchange topology information and a path-computation
interface in order to be able to send path computation requests and
responses. For signaling the overlay LSP a signaling interface is
requried as well.

5 Security Considerations

TBD

6 IANA Considerations

TBD

7 References

7.1 Normative References

[KEYWORDS] Bradner, S., "Key words for use in RFCs to Indicate
           Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC1776] Crocker, S., "The Address is the Message", RFC 1776, April
           1 1995.

[TRUTHS]   Callon, R., "The Twelve Networking Truths", RFC 1925,
           April 1 1996.

7.2 Informative References

[EVILBIT] Bellovin, S., "The Security Flag in the IPv4 Header",
           RFC 3514, April 1 2003.

[RFC5513] Farrel, A., "IANA Considerations for Three Letter
           Acronyms", RFC 5513, April 1 2009.

[RFC5514] Vyncke, E., "IPv6 over Social Networks", RFC 5514, April 1
           2009.

Authors' Addresses

Snigdho Bardalai
sbardalai@infinera.com

Rajan Rao
rrao@infinera.com

Khuzema Pithewan
kpithewan@infinera.com