PCE Working Group
Internet-Draft
Intended status: Informational
Expires: July 25, 2015

D. Dhody
Y. Lee
Huawei Technologies
LM. Contreras
O. Gonzalez de Dios
Telefonica I+D
N. Ciulli
Nextworks

January 21, 2015

Cross Stratum Optimization enabled Path Computation
draft-dhody-pce-cso-enabled-path-computation-07

Abstract

Applications like cloud computing, video gaming, HD Video streaming, Live Concerts, Remote Medical Surgery, etc are offered by Data Centers. These data centers are geographically distributed and connected via a network. Many decisions are made in the Application space without any concern of the underlying network. Cross stratum application/network optimization focus on the challenges and opportunities presented by data center based applications and carriers networks together [CSO-DATACNTR].

Constraint-based path computation is a fundamental building block for traffic engineering systems such as Multiprotocol Label Switching (MPLS) and Generalized Multiprotocol Label Switching (GMPLS) networks. [RFC4655] explains the architecture for a Path Computation Element (PCE)-based model to address this problem space.

This document explains the architecture for CSO enabled Path Computation.

Status of This Memo

This Internet-Draft will expire on July 25, 2015.

Copyright Notice

Table of Contents

1. Introduction

Many application services offered by Data Center to end-users make significant use of the underlying networks resources in the form of bandwidth consumption used to carry the actual traffic between data centers and/or among data center and end-users. There is a need for cross optimization for both network and application resources. [CSO-PROBLEM] describes the problem space for cross stratum optimization.

[NS-QUERY] describes the general problem of network stratum (NS) query in Data Center environments. Network Stratum (NS) query is an ability to query the network from application controller in Data Centers so that decision would be jointly performed based on both the application needs and the network status. Figure 1 shows typical data center architecture.

```
                             ---------------
       ----------           |         DC 1  |
      | End-user |. . . . .>|      o o o    |
      |          |          |       \|/     |
       ----------           |        O      |
            |                ----- --|------
            |                        |
            |                        |
            |       -----------------|-----------
            |      /                 |           \
            |     /        ..........O PE1        \     --------------
            |    |       .                         |   | o o o   DC 2 |
            |    | PE4 .                      PE2  |   |  \|/         |
             ----|---O.........................O---|---|---O          |
                 |     .                           |   |              |
                 |      .           PE3            |    --------------
                  \      ..........O   Carrier    /
                   \               |   Network   /
                    ---------------|-------------
                                   |
                           --------|------
                          |        O      |
                          |       /|\     |
                          |      o o o    |
                          |          DC 3 |
                           ---------------
```

Figure 1: Data Center Architecture

Figure 2 shows the context of NS Query within the overarching data center architecture shown in Figure 1.

```
                        --------------------------------------------
                       |                    Application Overlay     |
                       |                    (Data Centers)          |
                       |                                            |
     ----------        |    --------------         --------------   |
    | End-User |       |   | Application  |. . . .| Application  |  |
    |          |. . . >|   | Control      |       |  Processes   |  |
     ----------        |   | Gateway (ACG)|        --------------   |
                       |   |              |        --------------   |
                       |    ------------- . . . . | Application  |  |
                       |          /\              | Related Data |  |
                       |          ||               --------------   |
                        ----------||--------------------------------
                                  ||
                                  ||  Network Stratum Query (First
                                  ||                         Stage)
                                  ||
                        ----------||--------------------------------
                       |          \/         Network Underlay       |
                       |                                            |
                       |    --------------        ----------------  |
                       |   | Network      |. . . |    Network     | |
                       |   | Control      |      |    Processes   | |
                       |   | Gateway (NCG)|       ----------------
                       |   |              |       ----------------  |
                       |    -------------        |    Network     | |
                       |          |------------->|  Related Data  | |
                       |         (Second Stage)   ----------------  |
                        -------------------------------------------
```

Figure 2: NS Query Architecture

NS Query is a two-stage query that consists of two stages:

- A vertical query capability where an external point (i.e., the Application Control Gateway (ACG) in Data Center) will query the network (i.e., the Network Control Gateway (NCG)). The query can be initiated either by ACG to NCG or NCG to ACG depending on the mode of operation. ACG initiated query is an application-centric mode while NCG initiated query is a network-centric mode. It is anticipated that either ACG or NCG can be a final decision making point that chooses the end-to-end resources (i.e., both application IT resources and the network connectivity) depending on the mode of operation.

- A horizontal query capability where the NCG gathers the collective information of a variety of horizontal schemes implemented in the network stratum.

As an example for vertical query (1st stage), [ALTO-APPNET] describes Application Layer Traffic Optimization (ALTO) information model and protocol extensions to support application and network resource information exchange for high bandwidth applications in partially controlled and controlled environments as part of the infrastructure to application information exposure (i2aex) initiative.

For the horizontal query (2nd stage), PCE can be an ideal choice, [CSO-PCE-REQT] describes the general requirement PCE should support in order to accommodate CSO capability. This document is intended to fulfill the general PCE requirements discussed in the aforementioned reference.

This document describes how PCE Architecture as described in [RFC4655] can help in the second stage of NS query.

1.1. Requirements Language

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 [RFC2119].

2. Terminology

The following terminology is used in this document.

ACG: Application Control Gateway.

Application Stratum: The application stratum is the functional block which manages and controls application resources and provides application resources to a variety of clients/end-users. Application resources are non-network resources critical to achieving the application service functionality. Examples include: application specific servers, storage, content, large data sets, and computing power. Data Centers are regarded as tangible realization of the application stratum architecture.

ALTO: Application Layer Traffic Optimization.

CSO: Cross Stratum Optimization.

GMPLS: Generalized Multiprotocol Label Switching.

i2aex: Infrastructure to application information exposure.

LSR: Label Switch Router.

MPLS: Multiprotocol Label Switching.

NCG: Network Control Gateway.

Network Stratum: The network stratum is the functional block which manages and controls network resources and provides transport of data between clients/end-users to and among application resources. Network Resources are resources of any layer 3 or below (L1/L2/L3) such as bandwidth, links, paths, path processing (creation, deletion, and management), network databases, path computation, admission control, and resource reservation capability.

NMS: Network Management System

PCC: Path Computation Client: any client application requesting a path computation to be performed by a Path Computation Element.

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.

PCEP: Path Computation Element Communication Protocol.

TE: Traffic Engineering.

TED: Traffic Engineering Database.

UNI: User Network Interface.

3. CSO enabled PCE Architecture

In the network stratum, the Network Control Gateway (NCG) serves as the proxy gateway to the network. The NCG receives the query request from the ACG, probes the network to test the capabilities for data flows to/from particular points in the network, and gathers the collective information of a variety of horizontal schemes implemented in the network stratum. This is a horizontal query (Stage 2 in Figure 2).

In this section we will describe how PCE fits in this horizontal scheme.

A Path Computation Element (PCE) is 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.

(1) NCG and PCE are co-located.

In this composite solution, the same node implements functionality of both NCG and PCE. When a network stratum query is received from the ACG (stage 1), this query is broken into one or more Path computation requests and handled by the PCE functionality co-located with the NCG. There is no need for PCEP protocol here. In this case, an external PCE interface (e.g., CLI, SNMP, proprietary) needs to be supported. This is out of the scope of this document.

```
          +--------------------------------------------------+
          |   --   --   --   --   --   --   --   --   --     |
          |  |  | |  | |  | |  | |  | |  | |  | |  | |  |    |
          |   --   --   --   --   --   --   --   --   --     |
          |                                                  |
          |               Application Stratum                |
          |                                                  |
          |    +---------------------------------------+     |
          |    |                                       |     |
          +----+                  ACG                  +-----+
               |                                       |
               +------*---*----------------------------+
                      |   |
                      |   |
                      |   |
               +------*---*----------------------------+
               |   +----------+          +----------+  |
          +----+   +          *----------*          *  +-----+
          |    |   |   NCG    |          |   PCE    |  |     |
          |    |   |          *----------*          *  |     |
          |    |   +----------+          +----------+  |     |
          |    |                                       |     |
          |    +---------------------------------------+     |
          |                                                  |
          |                Network Stratum                   |
          |   --   --   --   --   --   --   --   --   --     |
          |  |  | |  | |  | |  | |  | |  | |  | |  | |  |    |
          |   --   --   --   --   --   --   --   --   --     |
          +--------------------------------------------------+
```

Figure 3: NCG and PCE Collocated

(2) NCG and external PCE

In this solution, an external node implements PCE functionality. Network stratum query received from the ACG (stage 1) is converted into Path computation requests at the NCG and relayed to the external PCE using the PCEP [RFC5440]. In this case the NCG includes Path Computation Client (PCC) functionalities.

```
          +--------------------------------------------------+
          |   --   --   --   --   --   --   --   --   --     |
          |  |  | |  | |  | |  | |  | |  | |  | |  | |  |    |
          |   --   --   --   --   --   --   --   --   --     |
          |                                                  |
          |               Application Stratum                |
          |                                                  |
          |    +---------------------------------------+     |
          |    |                                       |     |
          +----+                  ACG                  +-----+
               |                                       |
               +------*---*----------------------------+
                      |   |
                      |   |
                      |   |
               +------*---*-------+
               |   +----------+   |      +----------+
          +----+   |          |   *------*          *--------+
          |    |   |   NCG    |   |      |   PCE    |        |
          |    |   |          |   *------*          *        |
          |    |   +----------+   |      +----------+        |
          |    |                  |                          |
          |    +------------------+                          |
          |                                                  |
          |                Network Stratum                   |
          |   --   --   --   --   --   --   --   --   --     |
          |  |  | |  | |  | |  | |  | |  | |  | |  | |  |    |
          |   --   --   --   --   --   --   --   --   --     |
          +--------------------------------------------------+
```

Figure 4: NCG and external PCE

PCE has the capability to compute constrained paths between a source and one or more destination(s), optionally providing the value of the metrics associated to the computed path(s). Thus it can fit very well in the horizontal query stage of CSO. A PCE MAY have further capability to do multi-layer and/or inter-domain path computation which can be further utilized. NCG which understands the vertical query and the presence of applications constraints can break the application request into suitable path computation request which PCE understands. In this scenario, the PCE MAY have no knowledge of applications and provide only network related metrics to the NCG: the NCG (or the ACG for an application-centric model) is in charge of correlating the network quotations with the application layer information to achieve the global CSO objective.

With this architecture, NCG can request PCE different sets of computation mode that are not currently supported by PCE. For instance, NCG may request PCE a multi-destination and multi-source path computation request. This scenario arises when there are many possible Data Center choices for a given application request and there could be multiple sources for this request. Multi-destination with a single source (aka., anycast) is a default case for multi-destination and multi-source path computation.

In addition, with this architecture, NCG may have different sets of objectives and constraints than typical path computation requests. For instance, multi-criteria objective functions that combine the bandwidth requirement and latency may be very useful for some applications. [PCE-SERVICE-AWARE] describes the extension to PCEP to carry Latency, Latency-Variation and Loss as constraints for end to end path computation.

In a Stateful PCE (refer [PCE-STATEFUL]), there is a strict synchronization of the network states (in term of topology and resource information), and the set of computed paths and reserved resources in use in the network. In other words, the PCE utilizes information from the TED as well as information about existing paths (for example, TE LSPs) in the network when processing new requests. Stateful PCE will be very important tool to achieve the goals of cross stratum optimization as maintains the status of final path selected after cross (application and network) optimization.

As Stateful PCE would keep both LSP ID and the application ID associated with the LSP, it will make path computation more efficient in terms of resource usage and computation time. Moreover, Stateful PCE would have an accurate snapshot of network resource information and as such it can increase adaptability to the changes. This may be important for some application that requires a stringent performance objective.

In conclusion -

- NCG can use the PCE to do path computation based on constrains from multiple sources and destinations.

- Stateful PCE can help in maintaining the status of the final cross optimized path. It can also help NCG in maintaining the relationship of application request and setup path. In case of any change of the path, the Stateful PCE and NCG and cooperate and take suitable action.

4. Path Computation and Setup Procedure

Path computation flow is shown in Figure 5.

1. User for application would contact the application gateway ACG with its requirements.

2. ACG would further query the NCG to obtain the underlying network Status and quotations (offers) for the network connectivity services.

3. NCG would break the vertical request into suitable horizontal path computation request(s).

4. PCE would provide the result to NCG.

5. NCG would abstract the computation result and provide to ACG.

6. NCG and ACG would cooperate to finalize the path that needs to be setup.

7. Note that that the final decision can be made either in ACG or NCG depending on the mode of operation. With application centric mode, minimal data center/IT resource information would flow from ACG to NCG while ACG collects network abstracted information from NCG to choose the optimal application-network resources. With network centric mode, ACG would supply maximal data center/IT resource information to NCG so that NCG in conjunction with PCE would determine the optimal mixed set of application and network resources. In the latter case, the PCE COULD support application/IT- based constrained computation capability beyond network path computation. This requires further PCE capabilities to receive and process data center/IT resource information, possibly in conjunction with network information.

```
    +----------+    1    +---------------------------------------+
    |          |-------->|                                       |
    |   User   |         |                   ACG                 |
    |          |<--------|                                       |
    +----------+    6    +---------------------------------------+
                           ^  |
                           | 2|
                           |  |  +----------+    3     +----------+
                           |  +->|          |--------->|          |
                           |     |   NCG    |          |   PCE    |
                           +-----|          |<---------|          |
                             5   +----------+    4     +----------+
```

Figure 5: Path Computation Flow

In this section we would analyze the mechanisms to finally setup the cross stratum optimized path.

4.1. Path Setup Using NMS

After ACG and NCG have decided the path that needs to be set, NCG can send a request to NMS asking it relay the message to the head end LSR (also a PCC) to setup the pre computed path. Once the path signaling is completed and the LSP is setup, PCC should relay the status of the LSP to the Stateful PCE.

In this mechanism we can reuse the existing NMS to establish the path. Any updates or deletion of such path would be made via the NMS.

Head end LSR (PCC) 'H' is always the owner of the path.

See Figure 6 for this scenario.

```
    +----------+         +---------------------------------------+
    |          |-------->|                                       |
    |   User   |         |                  ACG                  |
    |          |<--------|                                       |
    +----------+         +---------------------------------------+
                          ^  |
        +-----------------+--+------------------------------------+
        |+----------+     |  |  +----------+          +----------+|
        ||          |     |  +->|          |--------->|          ||
        ||   NMS    |     +-----|   NCG    |          |   PCE    ||
        ||          |<----------|          |<---------|          ||
        |+----------+           +----------+          +----------+|
        |    |                                             ^      |
        |    |        +------------------------------------+      |
        |    |        |         Network Stratum                   |
        |    |       --   --   --   --   --   --   --   --   --   |
        |    +----->|H | |  | |  | |  | |  | |  | |  | |  | |  |  |
        |            --   --   --   --   --   --   --   --   --   |
        +---------------------------------------------------------+
```

Figure 6: Path Setup Using NMS

4.2. Path Setup Using a Network Control Plane

A network control plane (e.g. GMPLS) MAY be used to automatically establish the cross optimized path between the selected end points. This control plane MAY be triggered via -

- o NCG to Control Plane: GMPLS UNI or other protocols

- o Control Plane to Head end Router: GMPLS Control Channel Interface (CCI). Suitable protocol extensions are needed to achieve this.

See Figure 7 for this scenario.

```
    +----------+         +---------------------------------------+
    |          |-------->|                                       |
    |   User   |         |                  ACG                  |
    |          |<--------|                                       |
    +----------+         +---------------------------------------+
                          ^  |
        +-----------------+--+------------------------------------+
        |+----------+     |  |  +----------+          +----------+|
        || GMPLS    |     |  +->|          |--------->|          ||
        || Control  |     +-----|   NCG    |          |   PCE    ||
        ||  plane   |<----------|          |<---------|          ||
        |+----------+           +----------+          +----------+|
        |    |                                             ^      |
        |    |        +------------------------------------+      |
        |    |        |         Network Stratum                   |
        |    |       --   --   --   --   --   --   --   --   --   |
        |    +----->|H | |  | |  | |  | |  | |  | |  | |  | |  |  |
        |            --   --   --   --   --   --   --   --   --   |
        +---------------------------------------------------------+
```

Figure 7: Path Setup Using Centralized Control Plane

After cross optimization, ACG and NCG will select the suitable end points, (the path is already calculated by PCE), this path is conveyed to the head end LSR which signals the path and notify the status to the Stateful PCE. Later NCG can send suitable message to tear down the path.

Using centralized control plane can make the NCG responsible for the LSP. Head end LSR signals and maintains the status but the establishment and tear-down are initiated by the control plane. This would have an obvious advantage in managing the setup paths. The Stateful PCE will maintain the TED as well as the status of setup LSP. NCG through centralized control plane can further setup/teardown/modify/re-optimize those paths.

4.3. Path Setup using PCE

A Stateful PCE extension MAY be developed to communicate the cross optimized path to the head end LSR. Current PCEP protocol requires PCC to trigger Path request and PCE to provide reply. Even in Stateful PCE, PCC must delegate the LSP to a PCE, a PCE never initiate path setup. An extension to PCEP protocol MAY let PCE notify to PCC (Head end LSR) to establish the path.

NCG via PCE and PCEP protocol can establish and tear-down LSP as shown in Figure 8. [PCE_INITIATED] is one such attempt to extend PCEP.

```
    +----------+         +---------------------------------------+
    |          |-------->|                                       |
    |   User   |         |                  ACG                  |
    |          |<--------|                                       |
    +----------+         +---------------------------------------+
                          ^  |
        +-----------------+--+------------------------------------+
        |                 |  |  +----------+          +----------+|
        |                 |  +->|          |--------->|          ||
        |                 |     |   NCG    |          |   PCE    ||
        |                 +-----|          |<---------|          ||
        |                       +----------+          +----------+|
        |        +---------------------------------------+ ^      |
        |        |    +------------------------------------+      |
        |        |    |         Network Stratum                   |
        |        |   --   --   --   --   --   --   --   --   --   |
        |        +->|H | |  | |  | |  | |  | |  | |  | |  | |  |  |
        |            --   --   --   --   --   --   --   --   --   |
        +---------------------------------------------------------+
```

Figure 8: Path Setup using PCE

4.4. Path Setup Using a Software Defined Network controller

A logically centralized Software Defined Network (SDN) controller MAY be used to properly configure in an automatic way the traffic forwarding rules that allow the end to end communication across the Network Stratum.

Figure 9 shows this scenario.

```
      +----------+         +---------------------------------------+
      |          |-------->|                                       |
      |   User   |         |                  ACG                  |
      |          |<--------|                                       |
      +----------+         +---------------------------------------+
                            ^  |
          +-----------------+--+------------------------------------+
          |                 |  |  +----------+          +----------+|
          |                 |  +->|          |--------->|          ||
          |                 |     |   NCG    |          |   PCE    ||
          |                 +-----|          |<---------|          ||
          |                       +----------+          +----------+|
          |                           |   ^                         |
          |                           v   |                         |
          |               +----------------------------+            |
          |             +-|       SDN Controller       |--+         |
          |             | +----------------------------+  |         |
          |             |    |    |    |    |    |    |   |         |
          |             v    v    v    v    v    v    v   v         |
          |            --   --   --   --   --   --   --   --        |
          |           |  | |  | |  | |  | |  | |  | |  | |  |       |
          |            --   --   --   --   --   --   --   --        |
          |                                                         |
          |                       Network Stratum                   |
          |                                                         |
          +---------------------------------------------------------+
```

Figure 9: Path Setup using SDN

A direct interface between the SDN Controller and the PCE could be present in the architecture shown in Figure 9.

As result of the interaction between ACG and NCG (including the PCE processing), the NCG is able to instruct the SDN Controller to populate a number of forwarding rules to the network devices for building the end to end path.

5. Other Consideration

5.1. Inter-domain

5.1.1. One Application Domain with Multiple Network Domains

Underlying network connecting the datacenters MAYBE made up of multiple domains (AS and Area). In this case an inter-domain path computation is required.

```
    +----------+         +---------------------------------------+
    |          |-------->|                                       |
    |   User   |         |                  ACG                  |
    |          |<--------|                                       |
    +----------+         +---------------------------------------+
                          ^  |
                          |  |
    +--------------+   +--+--+------------------------------------+
    |  +----------+|   |  |  |  +----------+          +----------+|
    |  |          ||   |  |  +->|          |--------->|          ||
    |  |   PCE    ||   |  |     |   NCG    |          |   PCE    ||
    |  |          ||   |  +-----|          |<---------|          ||
    |  +----+-----+|   |        +----------+          +----+-----+|
    |       |      |   |                                   |      |
    +-------+------+   +-----------------------------------+------+
            |                                              |
            |                                              |
            |<---------------pcep session----------------->|
            |                                              |
```

Figure 10: Multi-domain Scenario

[RFC5441] describes an inter-domain path computation with cooperating PCEs which can be enhanced and utilized in CSO enabled path computation.

5.1.2. Multiple Application Domains with Multiple Network Domains

Underlying network connecting the datacenters MAY be made up of multiple domains (AS and Area) as well as applications domains and ACG MAY be distributed. In such case multiple ACG and NCG will be involved in cross optimizing. This needs to be analyzed further.

5.1.2.1. ACG talks to multiple NCGs

As shown in Figure 11, ACG where the request originates may communicate with multiple NCG to get the network information from multiple domains to be cross optimized.

```
     Application stratum
    +---------------------------+  +---------------------------+
    |                           |  |                           |
    |                           |  |                           |
    |                           |  |                           |
    |                           |  |                           |
    |                           |  |                           |
    |  +----------------------+ |  |  +----------------------+ |
    |  |                      | |  |  |                      | |
    +--+        ACG           +-+  +--+       ACG            +-+
       |                      |       |                      |
       +-+-+-------------+-+--+       +-------+-+------------+
         | |             | +------------+     | |
         | |             +------------+ |     | |
       +-+-+--------+   +-----+       +-+-----+-+--+    +-----+
    +--+            +---+     +-+  +--+            +----+     ++
    |  |    NCG     |---|     | |  |  |     NCG    |----|     ||
    |  |            |---|     | |  |  |            |----|     ||
    |  +------------+   | PCE | |  |  +------------+    | PCE ||
    |                   |     | |  |                    |     ||
    |                   |     |<+--+------------------->|     ||
    |                   +-----+ |  |                    +-----+|
    |Domain 1                   |  |Domain 2                   |
    +---------------------------+  +---------------------------+
                                            Network Stratum
```

Figure 11: ACG talks to multiple NCG

5.1.2.2. ACG talks to the primary NCG, which talks to the other NCG of different domains

As shown in Figure 12, ACG communicated only to the primary NCG, which may gather network information from multiple NCG and then communicate consolidated information to ACG.

```
     Application stratum
    +---------------------------+  +---------------------------+
    |                           |  |                           |
    |                           |  |                           |
    |                           |  |                           |
    |                           |  |                           |
    |                           |  |                           |
    |  +----------------------+ |  |  +----------------------+ |
    |  |                      | |  |  |                      | |
    +--+        ACG           +-+  +--+       ACG          +-+
       |                      |       |                      |
       +-+-+------------------+       +-------+-+------------+
         | |                                  | |
         | |                                  | |
       +-+-+--------+   +-----+       +-------+-+--+    +-----+
    +--+            +---+     +-+  +--+            +----+     ++
    |  |    NCG     |---|     | |  |  |     NCG    |----|     ||
    |  |            |---|     | |  |  |            |----|     ||
    |  +------+-----+   | PCE | |  |  +---+--------+    | PCE ||
    |         |         |     | |  |      |             |     ||
    |         |         |     |<+--+------+------------>|     ||
    |         |         +-----+ |  |      |             +-----+|
    |Domain 1 |                 |  |Domain|2                   |
    +---------+-----------------+  +------+--------------------+
              |                           | Network Stratum
              |                           |
              |<------------------------->|
              |                           |
```

Figure 12: Primary NCG talks to other NCG

5.1.3. Federation of SDN domains

In this case, the Data Centers are federated building a community cloud. In each Data Center, the connection to the network stratum that interconnects the Data Center federation is done by means of one or more devices controllable through an SDN controller particular for that Data Center.

The NCG, then, interacts with a number of separated SDN controllers, orchestrating their operation in order to perform the service requested by the ACG in an optimized way.

Figure 13 shows this scenario.

```
     +----------+         +---------------------------------------+
      |          |-------->|                                       |
      |   User   |         |                  ACG                  |
      |          |<--------|                                       |
      +----------+         +---------------------------------------+
                                      |  ^
                                      |  |
                           +----------+--+-------------------------+
                           |          |  |                         |
                           |          v  |                         |
                           |      +----------+         +----------+|
                           |      |          |-------->|          ||
                           |      |   NCG    |         |   PCE    ||
                           |      |          |<--------|          ||
                           |      +----------+         +----------+|
                           |       | ^    | ^                      |
                           |       | |    | |                      |
                           +-------+-+----+-+----------------------+
                                   | |    | |
                     +-------------+ |    | +-------------+
                     | +-------------+    +-------------+ |
                     | |                                | |
                     v |                                v |
           +--------------------+           +--------------------+
       +-  | SDN Controller DC1 |   . . .   | SDN Controller DCN |  -+
       |   +--------------------+           +--------------------+   |
       |                                                             |
       |                   Federated Data Centers                    |
       +-------------------------------------------------------------+
```

Figure 13: NCG orchestration of separated SDN domains

5.1.4. Nesting of multi-layer SDN domains

A different scenario for multi-domain interconnection could be due to the deployment of multi-layered multi-domain networks (and these domains may be technology, administrative or vendor specific (vendor islands))for supporting end-to-end connectivity at the Network Stratum. Each of those domains can be controlled by a distinct SDN controller adapted to the specifics of the technology under control.

The NCG requests path calculation to a multi-layer PCE which takes into consideration such diversity providing an integrated computation for the best path according to application constraints. The NCG instruct a primary SDN controller which apart of configuring the elements directly controlled by itself, it is able to communicate with other SDN controllers with responsibility over other domains. Such communication can be done through the usage of specific methods

through pre-defined South Bound Interface or East/West Interface (out of the scope of this document).

The following figure shows this scenario.

```
          +----------+        +---------------------------------------+
         |          |------->|                                       |
         |   User   |        |                  ACG                  |
         |          |<-------|                                       |
         +----------+        +---------------------------------------+
                                        |  ^
                                        |  |
                             +----------+--+-------------------------+
                             |          |  |                         |
                             |          v  |                         |
                             |      +----------+       +----------+  |
                             |      |          |------>|  Multi-  |  |
                             |      |   NCG    |       |  Layer   |  |
                             |      |          |<------|   PCE    |  |
                             |      +----------+       +----------+  |
                             |          |  ^                         |
                             |          |  |                         |
                             +----------+--+-------------------------+
                                        |  |
                                        v  |
                                    +----------+
                                    |   SDN    |----------+
                                    |Controller|          |
                                    +----------+          |
                                         ^          +----------+
                                         |          |   SDN    |
                                         v          |Controller|
                                      Layer-N       +----------+
                                     Resources           ^
                                                         |
                                                         v
                                                     Layer-N-1
                                                     Resources
```

Figure 14: Nested multi-layer SDN domains

5.2. Bottleneck

In optical networks all PCE messages are sent over control channel, in Stateful PCE cases its observed that in case of a major link or node failure lot of PCEP messages are sent from all PCC to PCE. This use lot of bandwidth of the control channel.

PCE MAY become a common point of failure and bottleneck. PCE/NCG/ACG failure as well as the link-failure disrupting connectivity could be highly disruptive to the system.

The solution should focus on reducing such bottleneck.

5.3. Relationship to ABNO

[ABNO] demonstrates cross-stratum application/network optimization for the data center use case with PCE as the heart of Application-Based Network Operations (ABNO) architecture. It further highlights the interaction between various ABNO components and PCE to achieve this use-case.

5.4. Relationship to ACTN

[ACTN] describes the framework for abstraction and control of transport networks (ACTN) using hierarchy of controllers. The Physical Network Controller (PNC) or Virtual Network Controller (VNC) is equivalent to the NCG in the CSO framework, both rely on PCE for the network optimization.

6. IANA Considerations

None, This is an informational document.

7. Security Considerations

TBD

8. Manageability Considerations

TBD

9. Acknowledgements

Part of the work in this document has been funded by the European Community's Seventh Framework Programme projects XIFI (L.M. Contreras and O. Gonzalez), under grant agreement n. 604590, and GEYSERS (N. Ciulli and L.M. Contreras), under grant agreement n. 248657.

10. References

10.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

10.2. Informative References

[RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.

[RFC5440] Vasseur, JP. and JL. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.

[RFC5441] Vasseur, JP., Zhang, R., Bitar, N., and JL. Le Roux, "A Backward-Recursive PCE-Based Computation (BRPC) Procedure to Compute Shortest Constrained Inter-Domain Traffic Engineering Label Switched Paths", RFC 5441, April 2009.

[CSO-DATACNTR]
Lee, Y., Bernstein, G., So, N., Kim, T., Shiomoto, K., and O. Gonzalez-de-Dios, "Research Proposal for Cross Stratum Optimization (CSO) between Data Centers and Networks. (draft-lee-cross-stratum-optimization-datacenter-00)", March 2011.

[CSO-PROBLEM]
Lee, Y., Bernstein, G., So, N., Hares, S., Xia, F., Shiomoto, K., and O. Gonzalez-de-Dios, "Problem Statement for Cross-Layer Optimization. (draft-lee-cross-layer-optimization-problem-02)", January 2011.

[NS-QUERY]
Lee, Y., Bernstein, G., So, N., McDysan, D., Kim, T., Shiomoto, K., and O. Gonzalez-de-Dios, "Problem Statement for Network Stratum Query. (draft-lee-network-stratum-query-problem-02)", April 2011.

[CSO-PCE-REQT]
Tovar, A., Contreras, L., Landi, G., and N. Ciulli, "Path Computation Requirements for Cross-Stratum-Optimization. (draft-tovar-cso-path-computation-requirements-00)", October 2011.

[PCE-SERVICE-AWARE]
Dhody, D., Wu, Q., Manral, V., Ali, Z., and K. Kumaki, "Extensions to the Path Computation Element Communication Protocol (PCEP) to compute service aware Label Switched Path (LSP). (draft-ietf-pce-pcep-service-aware-06)", December 2014.

[PCE-STATEFUL]
Crabbe, E., Medved, J., Varga, R., and I. Minei, "PCEP Extensions for Stateful PCE. (draft-ietf-pce-stateful-pce-10)", October 2014.

[ALTO-APPNET]
Lee, Y., Bernstein, G., Varga, T., Madhavan, S., Dhody, D., and Q. Wu, "ALTO Extensions to Support Application and Network Resource Information Exchange for High Bandwidth Applications. (draft-lee-alto-app-net-info-exchange-04)", October 2013.

[PCE_INITIATED]
Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "PCEP Extensions for PCE-initiated LSP Setup in a Stateful PCE Model. (draft-ietf-pce-pce-initiated-lsp-02)", July 2013.

[ACTN] Ceccarelli, D., Fang, L., Lee, Y., Lopez, D., Belotti, S., and D. King, "Framework for Abstraction and Control of Transport Networks", draft-ceccarelli-actn-framework-06 (work in progress), December 2014.

[ABNO] King, D. and A. Farrel, "A PCE-based Architecture for Application-based Network Operations", draft-farrkingel-pce-abno-architecture-15 (work in progress), January 2015.

Authors' Addresses

Dhruv Dhody
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560037
India

EMail: dhruv.ietf@gmail.com

Young Lee
Huawei Technologies
1700 Alma Drive, Suite 500
Plano, TX 75075
USA

EMail: leeyoung@huawei.com

Luis M. Contreras
Telefonica I+D
Ronda de la Comunicacion, s/n
Sur-3 building, 3rd floor
Madrid 28050
Spain

EMail: lmcm@tid.es

Oscar Gonzalez de Dios
Telefonica I+D
Ronda de la Comunicacion, s/n
Sur-3 building, 3rd floor
Madrid 28050
Spain

EMail: ogondio@tid.es

Nicola Ciulli
Nextworks

EMail: n.ciulli@nextworks.it