PCE-based Computation Procedure To Compute Shortest Constrained P2MP
Inter-domain Traffic Engineering Label Switched Paths
draft-zhao-pce-pcep-inter-domain-p2mp-procedures-01.txt

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

This Internet-Draft will expire on December 3, 2009.

Copyright Notice

Copyright (c) 2009 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust’s Legal Provisions Relating to IETF Documents in effect on the date of publication of this document (http://trustee.ietf.org/license-info). Please review these documents carefully, as they describe your rights and restrictions with respect to this document.
Abstract

Point-to-multipoint (P2MP) Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) Traffic Engineering Label Switched Paths (TE LSPs) may be established using signaling techniques, but their paths must first be determined. The Path Computation Element (PCE) has been identified as an appropriate technology for the determination of the paths of P2MP TE LSPs.

This document describes the procedures and extensions to the PCE communication Protocol (PCEP) to handle requests and responses for the computation of inter-domain paths for P2MP TE LSPs.

Table of Contents

1. Terminology ..............................
2. Introduction ............................
   2.1. Requirements Language ..............
3. Problem Statement .......................:
4. Assumptions ............................
5. Requirements ...........................
6. Objective Functions ....................
7. Protocol Procedures ....................
   7.1. Per Domain Path Computation .......
   7.2. Backwards Recursive Path Computation
   7.3. Core Tree Based Path Computation ...
8. Protocol Extensions for Core Tree Based Computation ......
   8.1. The Extension of RP Object .........
   8.2. The PCE Sequence Object ..........
9. Manageability Considerations ...........
   9.1. Control of Function and Policy ....
   9.2. Information and Data Models ..... 
   9.3. Liveness Detection and Monitoring 
   9.4. Verifying Correct Operation .......
   9.5. Requirements on Other Protocols and Functional Components 
   9.6. Impact on Network Operation .......
10. Security Considerations ............... 
11. IANA Considerations ..................
12. Acknowledgement ........................
13. References ............................
   13.1. Normative References .............
   13.2. Informative References .......... 
Authors’ Addresses ........................
1. Terminology

Terminology used in this document

ABR: Area Border Routers. Routers used to connect two IGP areas (areas in OSPF or levels in IS-IS).

ASBR: Autonomous System Border Routers. Routers used to connect together ASes of the same or different Service Providers via one or more Inter-AS links.

Boundary Node (BN): a boundary node is either an ABR in the context of inter-area Traffic Engineering or an ASBR in the context of inter-AS Traffic Engineering.

Entry BN of domain(n): a BN connecting domain(n-1) to domain(n) along a determined sequence of domains.

Exit BN of domain(n): a BN connecting domain(n) to domain(n+1) along a determined sequence of domains.

Inter-AS TE LSP: A TE LSP that crosses an AS boundary.

Inter-area TE LSP: A TE LSP that crosses an IGP area boundary.

TED: Traffic Engineering Database.

VSPT: Virtual Shortest Path Tree.

P2MP LSP Path Tree: A set of LSRs and TE links that comprise the path of a P2MP TE LSP from its ingress LSR to all of its egress LSRs.

Core Tree: The core tree is a P2MP tree where the root is the ingress LSR, the transit node and branch node are the BNs of the transit domains and the leaf nodes are the leaf BNs of the leaf domains.

Root Boundary Node: The egress LSR from the root domain on the path of the P2MP LSP.

Root Domain: The domain that includes the ingress (root) LSR.

Transit/branch Domain: A domain that has an upstream and one or more downstream neighbour domain.

Leaf Domain: A domain that doesn’t have a downstream neighbor domain.

Leaf Boundary Nodes: The entry boundary node in the leaf domain.

Leaf Nodes: The LSR which is the P2MP LSP’s final
Destination. The lead Nodes can be in Root Domain, Transit Domain and Leaf Domain.

OF: Objective Function: A set of one or more optimization criterion (criteria) used for the computation of a single path (e.g. path cost minimization), or the synchronized computation of a set of paths (e.g. aggregate bandwidth consumption minimization, etc.). See [RFC4655] and [RFC5541].

Path Domain Sequence: The known sequence of domains for a path between the ingress LSR and a leaf node.

PCE Sequence: The known sequence of PCEs for calculating a path between the ingress LSR and leaf node.

PCE Topology Tree: A list of PCE Sequences which has all the PCE Sequence for each path of the P2MP LSP path tree.

This document also uses the terminology defined in [RFC4655], [RFC4875], [RFC5440], and [RFC5441].

2. Introduction

Multicast services are increasingly in demand for high-capacity applications such as multicast VPNs, IPTV (on-demand or streaming), and content-rich media distribution (for example, software distribution, financial streaming, or data-sharing). The inter-domain P2MP TE LSP is a feature that facilitates the deployment and operation of these services across multi domains.

The need to establish point-to-multipoint (P2MP) traffic engineered (TE) Label Switching Paths (LSPs) in Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) networks is covered in [RFC4461].

The applicability of the Path Computation Element (PCE) for the computation of such paths is discussed in [PCE-P2MP-APP], and the requirements placed on PCEP for this are given in [PCE-P2MP-REQ].

This document defines a PCE-based solution which will compute the optimal inter-domain P2MP TE LSP.

2.1. 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 RFC 2119 [RFC2119].
3. Problem Statement

The Path Computation Element (PCE) defined in [RFC4655] is an entity that is capable of computing a network path or route based on a network graph, and applying computational constraints. A Path Computation Client (PCC) may make requests to a PCE for paths to be computed.


The PCE is identified as a suitable application for the computation of paths for P2MP TE LSPs [PCE-P2MP-APP].

The draft-ietf-pce-brpc-09.txt specifies a procedure relying on the use of multiple Path Computation Elements (PCEs) to compute point-to-point (P2P) inter-domain shortest constrained paths across a predetermined sequence of domains, using a backward recursive path computation technique. The technique preserves confidentiality across domains, which is sometimes required when domains are managed by different Service Providers.

The PCE communication protocol (PCEP) [RFC5440] is extended as a communication protocol between PCCs and PCEs for point-to-multipoint (P2MP) path computations and is defined in [PCE-P2MP-EXT]. However, that specification does not provide a mechanism to request path computation of inter-domain P2MP TE LSPs.

This document presents a solution, and procedures and extensions to PCEP to support P2MP inter-domain path computation.

4. Assumptions

It is assumed that due to deployment and commercial limitations (e.g., inter-AS peering agreements) the sequence of domains for a path (the path domain tree) will be known in advance.

The examples and scenarios used in this document are also based on the following assumptions:

- The PCE that serves each domain in the path domain tree is known and the set of PCEs and their relationships is propagated to each PCE during the first exchange of path computation requests;

- Each PCE knows about any leaf LSRs in the domain it serves;

- The boundary nodes to use on the LSP are pre-determined and form path of the path domain tree. In this version of the document we do not consider multi-homed domains.
5. Requirements

This section summarizes the PCEP requirements specific to computing inter-domain P2MP paths. In these requirements we note that the actual computation times by any PCE implementation are outside the scope of this document, but we observe that reducing the complexity of the required computations has a beneficial effect on the computation time regardless of implementation. Additionally, reducing the number of message exchanges and the amount of information exchanged will reduce the overall computation time for the entire P2MP tree. We refer to the "Complexity of the computation" as the impact on these aspects of path computation time as various parameters of the topology and the P2MP LSP are changed.

1. The requirements specified in [RFC5376];

1.1 PCEP must allow an SP to hide from other SPs the set of hops within its own ASes that are traversed by an inter-AS inter-provider TE LSP for each inter-AS TE LSP path segment an inter-AS PCE computes, it may return to the requesting inter-AS PCE an inter-AS TE LSP path segment from its own ASes without detailing the explicit intra-AS hops.

2. A number of additional requirements have also been identified in [RFC4461].

3. The computed P2MP LSP should be optimal when only considering the paths among the BNs.

4. Grafting and pruning of multicast destinations in a domain should have no impact on other domains and on the paths among BNs.

5. The complexity of the computing for each sub-tree within each domain should be only dependent on the topology of the domain and it should be independent of the domain sequences.

6. The number of PCEP request and reply messages should be independent of the number of multicast destinations in each domain.

6. Objective Functions

During the computation of a single or a set of P2MP TE LSPs a request to meet specific optimization criteria, called an Objective Function (OF), may be requested.
The computation of one or more P2MP TE-LSPs maybe subject to an OF in order to select the "best" candidate paths. A variety of objective functions have been identified as being important during the computation of inter-domain P2MP LSPs. These are:

1. The sub-tree within each domain should be optimized.

1.1 Minimum cost tree [PCE-P2MP-REQ].

1.2 Shortest path tree [PCE-P2MP-REQ].

2. The P2MP LSP paths should be optimal while only considering the entry and exit nodes of each domain as the transit, branch and leaf nodes of the P2MP LSP path. (That is, the Core Tree should be optimized.)

3. It should be possible to limit the number of entry points to a domain.

4. It should be possible to force the branches for all leaves within a domain to be in that domain.

7. Protocol Procedures

The following sections describe the procedures to satisfy the requirements specified in the previous section.

7.1. Per Domain Path Computation

Computing P2P LSPs individually is an acceptable solution for computing a P2MP tree. Per domain path computation [RFC5152] can be used to compute P2P multi-domain paths, but it does not guarantee to find the optimal path which crosses multiple domains. Furthermore, constructing a P2MP tree from individual source to leaf P2P LSPs does not guarantee to produce a least-cost tree. This approach may be considered to have scaling issues during LSP setup. That is, the LSP to each leaf is signaled separately, and each border node must perform path computation for each leaf. This solution does suit simply-connected domains and where the preferred points of interconnection are known.

7.2. Backwards Recursive Path Computation

Backward recursive path computation (BRPC) [RFC5441] provides a mechanism to compute optimal P2P LSPs. This overcomes one of the issues raised in Section 3.1, but nevertheless, a P2MP tree constructed from individually computed optimal P2P LSPs does not guarantee to produce a least-cost tree; it does produce a least-cost-to-destination tree.
This solution suits environments where multiple connections exist between domains and there is no preference for the choice of points of interconnection. This approach may have scaling issues during path computation as each the path to each leaf must be computed separately. Although path fragments already used for other leaves may be re-used by stateful PCEs, the information must still be transmitted in a full exchange of PCEP messages for each leaf.

7.3. Core Tree Based Path Computation

A core tree based solution provides an optimal inter-domain P2MP TE LSP and meets the requirements and OFs outlined in previous sections.

A core tree is a path tree with nodes from each domain corresponding to the PCE topology which satisfies the following conditions:

- The root of the core tree is the ingress LSR in the root domain;
- The leaf of the core tree is the entry node in the leaf domain;
- The transit and branch nodes of the core tree are from the entry and exit nodes from the transit and branch domains.

Computing the complete P2MP LSP path tree is done in two phases:

Procedure Phase 1: Build the P2MP LSP Core Tree.

The algorithms to compute the optimal large core tree are outside scope of this document. In the case that the number of domains and the number of BNs are not big, the following extended BRPC based procedure can be used to compute the core tree.

BRPC Based CoreTree Path Computation Procedure

(1). Using the BRPC procedures to compute the VSPT(i) for each leaf BN(i), i=1 to n, where n is the total number of entry nodes for all the leaf domains. In each VSPT(i), there are a number of P(i) paths.

(2). When the root PCE has computed all the VSPT(i), i=1 to n, take one path from each VSPT and form a set of paths, we call it a PathSet(j), j=1 to M, where M=P(1)xP(2)...xP(n);

(3). For each PathSet(j), there are n S2L (Source to Leaf BN) paths and form these n paths into a CoreTree(j);

(4). There will be M number of CoreTrees computed from step3. Apply the OF to each of these M CoreTrees and find the optimal CoreTree.
Procedure Phase 2: Grafting destinations to the P2MP LSP Core Tree.

Once the core tree is built, the grafting of all the leaf nodes from each domain to the core tree can be achieved by a number of algorithms. One algorithm for doing this phase is that the root PCE will send the request with C bit set for the path computation to the destination(s) directly to the PCE where the destination(s) belong(s) along with the core tree computed from the phase 1.

8. Protocol Extensions for Core Tree Based Computation

The following section describes the protocol extensions for Core Tree based inter-domain P2MP path calculation.

8.1. The Extension of RP Object

The extended format of the RP object body to include the C bit as follows:

The C bit is added in the flag bits field of the RP object to signal the receiver of the message that the request/reply is for inter-domain P2MP coreTree or not.

```
0                   1                   2                   3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Reserved   | Flags                           |C|O|B|R| Pri |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        Request-ID-number                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
//                      Optional TLV(s)                        //
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: RP Object Body Format

The following flags are added in this draft:

- C (P2MP coreTree bit - 1 bit):
  - 0: This indicates that this is normal PCReq/PCRrep for P2MP.
  - 1: This indicates that this is PCReq or PCRep message for inter-domain coreTree P2MP. When the C bit is set, then the request message should have the coretree passed along with the destinations which are needed to be graphed.
8.2. The PCE Sequence Object

The PCE Sequence Object is added to the existing PCE protocol. A list of this objects will represent the PCE topology tree. A list of Sequence Objects can be exchanged between PCEs during the PCE capability exchange or on the first path computation request message between PCEs. In this case, the request message format needs to be changed to include the list of PCE Sequence Objects for the PCE inter-domain P2MP calculation request.

Each PCE Sequence can be obtained from the domain sequence for a specific path. All the PCE sequences for all the paths of P2MP inter-domain form the PCE Topology Tree of the P2MP LSP.

The format of the new PCE Sequence Object for IPv4 (Object-Type 3) is as follows:

```
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Object-Class | OT | Res | P | I | Object Length (bytes) |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| IPv4 address for root PCE |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| IPv4 address for the downstream PCE |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| IPv4 address for the downstream PCE |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| IPv4 address for the PCE corresponding to the leafDomain |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: The New PCE Sequence Object Body Format for IPv4

The format of the new PCE Sequence Object for IPv6 (Object-Type 3) is as follows:
9. Manageability Considerations

[PCE-P2MP-REQ] describes various manageability requirements in support of P2MP path computation when applying PCEP. This section describes how manageability requirements mentioned in [PCE-P2MP-REQ] are supported in the context of PCEP extensions specified in this document.

Note that [RFC5440] describes various manageability considerations in PCEP, and most of manageability requirements mentioned in [PCE-P2MP-P2MP] are already covered there.

9.1. Control of Function and Policy

In addition to configuration parameters listed in [RFC5440], the following parameters MAY be required.

- P2MP path computations enabled or disabled.
- Advertisement of P2MP path computation capability enabled or disabled (discovery protocol, capability exchange).

9.2. Information and Data Models

As described in [PCE-P2MP-REQ], MIB objects MUST be supported for PCEP extensions specified in this document.

---

Figure 3: The New PCE Sequence Object Body Format for IPv6
9.3. Liveness Detection and Monitoring

There are no additional considerations beyond those expressed in [RFC5440], since [PCE-P2MP-REQ] does not address any additional requirements.

9.4. Verifying Correct Operation

There are no additional considerations beyond those expressed in [RFC5440], since [PCE-P2MP-REQ] does not address any additional requirements.

9.5. Requirements on Other Protocols and Functional Components

As described in [PCE-P2MP-REQ], the PCE MUST obtain information about the P2MP signaling and branching capabilities of each LSR in the network.

Protocol extensions specified in this document does not provide such capability. Other mechanisms MUST be present.

9.6. Impact on Network Operation

It is expected that use of PCEP extensions specified in this document will not have significant impact on network operations.

10. Security Considerations

As described in [PCE-P2MP-REQ], P2MP path computation requests are more CPU-intensive and also use more link bandwidth. Therefore, it may be more vulnerable to denial of service attacks. Therefore it is more important that implementations conform to security requirements of [RFC5440], and the implementor utilize those security features.

11. IANA Considerations

A number of IANA considerations have been highlighted in previous sections of this document. This section will highlight those requests in future versions of this document.

12. Acknowledgement

The authors would like to thank Adrian Farrel and Dan Tappan for their valuable comments on this draft.
13. References

13.1. Normative References


13.2. Informative References


[PCE-P2MP-EXT]

Authors’ Addresses

Quintin Zhao (editor)
Huawei Technology
125 Nagog Technology Park
Acton, MA 01719
US

Email: qzhao@huawei.com

David Amzallag
British Telecommunications plc
UK

Email: David.Amzallag@bt.com

Daniel King
Old Dog Consulting
UK

Email: daniel@olddog.co.uk

Fabien Verhaeghe

Email: fabien.verhaeghe@gmail.com