PCE Working Group
Internet Draft
Category: Standards track

Xian Zhang
Haomian Zheng
Huawei
Oscar Gonzales de Dios
Victor Lopez
Telefonica I+D

Expires: January 3, 2015
July 3, 2014

Extensions to Path Computation Element Protocol (PCEP) to Support Resource Sharing-based Path Computation

draft-zhang-pce-resource-sharing-01.txt

Status of this Memo

Copyright Notice

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].

Abstract

Resource sharing in a network means two or more Label Switched Paths (LSPs) use common piece(s) of resource along their paths. This can help save network resource and useful in scenarios such as LSP recovery or when two LSPs do not need to be active at the same time. A Path Computation Element (PCE) is a centralized entity, responsible for path computation. Given this feature and its access to the network resource information and possibly active LSPs information, it can be used to support resource-sharing-based path computation with better efficiency.

This document extends the Path Computation Element Protocol (PCEP) in order to support resource sharing-based path computation.

Table of Contents

1. Introduction and Motivation

A Path Computation Element (PCE) provides an alternative way for providing path computation function, and it is especially useful in the scenarios where complex constraints and/or a demanding amount of computation resource are required [RFC4655]. The development of PCE standardization has evolved from stateless to stateful. A stateful PCE has access to the LSP database information of the network(s) it serves as a computation engine [Stateful-PCE]. Unless specified, this document assumes a PCE mentioned is a stateful PCE (either passive or active).

Resource sharing denotes that two or more Label Switched Paths (LSPs) share common piece(s) of resource, (such as a common time slot of a link in an Optical Transport Network (OTN)). This is usually useful in the scenario where only one LSP is active and the benefit herein is to save network resources. A simple example of this is dynamically calculating a LSP for an existing LSP undergoing a link failure. Note that the resource sharing can be worked out using a statelss PCE, but the mechanism may be complex and is out the scope of this draft.

This document considers the following requirement: resource sharing with one or multiple existing LSPs.

In a single domain, this is a common requirement in the recovery cases especially in order to increase traffic resilience against failure while reducing the amount of network resource used for recovery purpose [RFC4428].

The current protocol supporting the communication between a PCE and a Path Computation Client (PCC), i.e. PCE Protocol (PCEP), allows for re-optimization of an existing LSP [RFC5440]. This is achieved by setting R bit in the Request Parameter (RP) object, together with some additional information if applicable, in the Path Computation Request (PCReq) message sent from a PCC to the PCE. To support this type of resource sharing, a PCC needs to ask a PCE to compute a new path with the constraints of sharing resource with one or multiple existing LSPs. It is worth noting the 'resource sharing' in this draft not only means one LSP re-using the same link(s) of another LSP, but also the same slice of bandwidth. This may occur when an LSP is required for re-routing, or online re-optimization. Current PCEP specifications do not provide such function.

As mentioned in [stateful-PCE], the standardization of stateful PCEs also facilitates PCEP to meet this requirement since a LSP can be identified using a unique number. This simplifies configuration of PCCs by making it simpler to for a PCC to request resource sharing without having to determine all of the resources to be shared.

The resource sharing can also be required across layers. This is similar to the previous requirement. However, it is more complex and therefore deserves a more detailed explanation here.

In a multi-layer network, Label Switched Paths (LSPs) in a lower layer are used to carry higher-layer LSPs across the lower-layer network [RFC5623]. Therefore, the resource sharing constraints in the higher layer might actually relate to the resource sharing in the lower layer. Thus, it is useful to consider how this can be achieved and whether additional extensions are needed using the models defined in [RFC5623].

In the next sections, use cases are provided to show what information needs to be exchanged to fulfill these requirements. This memo then provides extensions to PCEP to enable this function.

2. Motivation

2.1. Use Case 1

Figure 1 shows a single domain network with a stateful PCE. Assume a working LSP (N1-N2-N3) exists in the network. When there is failure on the link N2-N3, it is desired to set up a restoration path for this working LSP. Suppose N1 serves as the PCC and sends a request to the stateful PCE for such an LSP. Before sending the request, N1 may need to check what policy is configured locally on N1. For example, it might value resource sharing and prefer to share as much resource with the working LSP as possible and specify this in the PCReq message.   Effectiveness here denotes whether the traffic can be diverted back to the working LSP immediately once the failure on the working LSP is repaired. In the case where resource sharing is more important, it would prefer to share as much resource with the working LSP as possible and specify this in the PCReq message.

On the other hand, in some case the LSP should be restored without any interruption with best effort, for example the online re-optimization. In such cases, it would prefer to share as few resources as possible. The best result for such case is to find a separate path that can make the LSP before break, which means no resource sharing involved. This can actually be implemented with existing PCEP mechanism. However, if there is no such separate path, existing PCEP will reply error. A secondary option for this case is to set up an LSP and complete such re-optimization with resource sharing, even if some interruption introduced. Given the resource from the LSP to be interrupted, there may be some solutions instead of Path Compute error due to the lack of resource.

A simple illustration is provided below:

```
              +--------------+
              |              |
              | Stateful PCE |
              |              |
              +--------------+



         +-------+          +-------+          +-------+
         |  N1  +-----------+  N2  +------X---+  N3  |
         +--+---+          +---+--+          +---+--+
            |                  |                 |
            |                  +----------+      |
            |                             |      |
            |     +-------+          +-------+   |
            +-----+  N5  +-----------+  N4  +------+
                  +-------+          +-------+
```

Figure 1: A Single Domain Example

Available recovery paths computed by the stateful PCE:

LSP1: N1-N2-N4-N3
LSP2: N1-N5-N4-N3

If resource sharing is preferred, the stateful PCE will reply with LSP1 information. Instead, if effectiveness is valued higher, it will reply with LSP2 information.

Another piece of information that needs to be conveyed to the PCE is the information about the working path LSP. Note this simple use case assumes end-to-end recovery. But in order to be applicable to use cases such as shared mesh protection purpose, where the head-end or tail-end nodes may be different, this information is necessary in the message exchange between PCCs and PCEs, so that the stateful PCE knows which LSP the path computation request wants to share the resource.

Besides, parameter changes during the resource sharing computation also need to be considered. For example, the bandwidth of the request may be different with the existing LSP, but still ask for resource sharing. PCE should consider the sharing request together with the policy and available resource(s) in the network. Details can be found in Section 3.3.

2.2. Use Case 2

Figure 2 shows a two-layer network example, with each layer managed by a PCE (shown in the graph as PCE Hi (for higher layer) and PCE Lo for PCE lower layer). As Discussed in Section 3 of [RFC5623], there are three models for inter-layer path computation. They are single PCE computation, multiple PCE with inter-PCE communication and multiple PCE without inter-PCE communication, respectively. For the single PCE computation, the process would be similar to that of the use case in Section 2.1. Thus, this model is not discussed further.

```
                                                               -----
                             .................................| LSR |
                           .:                                 | H5  |
                         .:                                   /-----
                       .:                                    /   |
       -----    -----.:                       -----    -----/    |
      | LSR |--| LSR |.......................| LSR |--| LSR |   /
      | H1  |  | H2  |                       | H3  |  | H4  |  /
       -----    -----\                       /-----    -----  /
                      \                     /                /
                       \                   /                /
                        \                 /                /
                         \               /                /
                          \-----   -----/                /
                          | LSR |-| LSR |               /
                          | L1  | | L2  |              /
                           -----   -----\             /
                             |           \           /
                             |            \         /
                             |             \       /
                           -----            \-----/
                          | LSR |-----------| LSR |
                          | L3  |           | L4  |
                           -----             -----
```

Figure 2: A Two-layer Network Example

In this example, assume a LSP (LSP1: H2-H3) has been established already. A new request comes at H2 to establish a new LSP (LSP2:

from H2 to H5), given the constraint it can share resource with LSP1. This requirement is possible if only one of the LSPs needs to be active and resource sharing is the target.

If multiple PCE with inter-PCE communication model is employed, the path computation request sent by H2 to PCE Hi will be passed to PCE Lo since there is no resource readily available in the upper layer. So it leaves to the PCE Lo to compute a path in the lower layer in order to support the upper layer request. In this case, PCE Lo is required to compute a path between H2 and H5 under the constraint that it can share the resource with that of the LSP1. Assume here LSP1 goes from H2, via L1-L2 to H3. So when PCE Lo computes the path for LSP2, it can view the resource used by LSP1 available. For example, PCE Lo may choose H2-L1-L2-L4-H5 as the computation result.

The issue to solve during this procedure is that PCE Hi can only use LSP1 information (such as its five-tuple LSP information) as the information, how PCE Lo can resolve this information to the actual resource usage in its own layer, i.e. lower layer. This could be solved by edge LSR L1 reporting this higher-lower layer LSP correlation to the Lo PCE as part of the LSP information during the LSP state synchronization process. If needed, it can be later updated when there is a change in this information. Alternatively, the PCE Lo can get this information from other sources, such as network management system, where this information should be stored.

If multiple PCE without inter-PCE communication model is employed, the path computation request in the lower layer will be initiated the border LSR node, i.e., L1. The process would be similar to that of the previous scenario. A point worth noting is that the border LSR node may be able to resolve the higher LSP information itself, such as mapping it to the corresponding LSP in the lower layer, thus PCE Lo do not need to perform this function. Otherwise, the mapping method mentioned above can still be used.

3. Extensions to PCEP

This section provides PCEP extensions. Currently the text focuses only on passive stateful PCE and corresponding PCReq. But if active stateful PCE delegation is used, we would like to convey the same information via RSO in PCRpt. In the passive stateful PCE architecture, a PCC is allowed to specify resource sharing when sending a PCReq message. It also details the processing rule and error codes needed.

3.1. Resource Sharing Object

The PCEP Resource Sharing Object (RSO) is optional. It MAY be carried within a PCRep message so as to indicate the desired resource sharing requirements to be applied by the stateful PCE during path computation.

The RSO object format is compliant with the PCEP object format defined in [RFC5440].

The RSO Object-Class is TBA.

The RSO Object-type is 1.

The format of the RSO object body is:

```
 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 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| RSO Flags                 |R|D|       Reserved                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                        Optional TLVs                          ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3: RSO Object Format

RSO codes (16 bits): the objective of the resource sharing. Currently, the following objectives are defined:

D (1 bit): sharing as little as possible.

R (1 bit): sharing as much as possible

It is possible that multiple computation results satisfy the request. Among these results, D set to 1 will select the most separate one, while R set to 1 will select the most sharing one. Both D and R set to 0 don’t specify any constraint and will result in a random selection among these results. The combination of D=1 and R=1 is not allowed.

Reserved (2 bytes): This field MUST be set to zero on transmission and MUST be ignored on receipt.

Optional TLVs may be needed to indicate the LSP with which the resource is shared. The LSP Info TLV, include the IPv4-LSP-

IDENTIFIERS TLV and IPv6-LSP IDENTIFIERS TLV, are defined in the same way as in [stateful-pce].

3.2. Processing Rules

To request a path allowing sharing resource with one or multiple existing LSPs, a PCC includes a RSO object in the PCReq message.

On receipt of a PCReq message with a RSO object, a stateful PCE MUST proceed as follows:

- If the RSO object is unknown/unsupported, the PCE will follow procedures defined in [RFC5440]. That is, the PCE sends a PCErr message with error type 3 or 4 (Unknown / Not supported object) and error value 1 or 2 (unknown / unsupported object class / object type), and the related path computation request is discarded.

- If TLV(s) present in the RSO object are unknown/unsupported and the P bit is set, the PCE MUST send a PCErr message with error type 3 or 4 (Unknown / Not supported object) and error value 4 (Unrecognized/Unsupported parameter), and the related path computation request MUST be discarded as defined in [RFC5440].

- If the resource sharing information is extracted correctly, the PCE MUST apply the requested resource sharing requirement.

The procedure of setting R and/or D bit follows the rules defined in Section 3.1. The RSO codes may be locally configured on the requesting nodes via external entities, such as a network management system or the entity that impose the resource sharing requirement.

3.3. Carrying RSO in a PCEP Message

The RSO is applied to an individual path computation request and the format of the PCReq message is updated as follows:

```
<PCReq Message> ::= <Common Header>

                    [<svec-list>]

                    <request-list>
```

where:

```
        <svec-list> ::= <SVEC>

                        [<OF>]

                        [<metric-list>]

                        [<svec-list>]


       <request-list> ::= <request> [<request-list>]


       <request> ::= <RP>

                     <END-POINTS>

                     [<LSPA>]

                     [<BANDWIDTH>]

                     [<metric-list>]

                     [<OF>]

                     [<RRO>[<BANDWIDTH>]]

                     [<IRO>]

               [<RSO>[<BANDWIDTH>]]

                     [<LOAD-BALANCING>]
```

and where:

```
      <metric-list> ::= <METRIC>[<metric-list>]
```

4. Security Considerations

Security of PCEP is discussed in [RFC5440] and [RFC6952]. The extensions in this document do not change the fundamentals of security for PCEP.

However, the introduction of the RSO provides a vector that may be used to probe for information from a network. For example, a PCC that wants to discover the path of an LSP with which it is not involved, can issue a PCReq with an RSO and may be able to get back quite a lot of information about the path of the LSP through issuing multiple such requests for different endpoints and analyzing the received results. To protect against this, a PCE should be configured with access and authorization controls such that only authorized PCCs (for example, those within the network) can make computation requests, only specifically authorized PCCs can make requests using the RSO, and resource sharing requests relating to specific LSPs are further limited to a select few PCCs. How such access controls and authorization is managed is outside the scope of this document, but it will at the least include Access Control Lists.

Furthermore, a PCC must be aware that setting up an LSP that shares resources with another LSP may be a way of attacking the other LSP, for example by depriving it of the resources it needs to operate correctly. Thus it is important that, both in PCEP and the associated signaling protocols, only authorized resource sharing is allowed.

5. IANA Considerations

5.1. New Object Type

IANA manages the PCEP Objects code point registry (see [RFC5440]). This is maintained as the "PCEP Objects" sub-registry of the "Path Computation Element Protocol (PCEP) Numbers" registry.

This document defines a new PCEP object, the RSO object, to be carried in PCReq messages. IANA is requested to make the following allocation in the ''PCEP Objects'' sub-registry:

Object Class	Name	Object Type	Name	Reference
TBA	RSO		Resource Sharing	[this document]

5.2 RSO codes

IANA is requested to create and maintain a new sub-registry named ''RSO codes''. The following codes are defined in this document:

Bit	Code Name	Meaning	Reference
0	D	sharing as much as possible	

[this document]

1	R	sharing as little as possible

[this document]

6. References

6.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to indicate requirements levels", RFC 2119, March 1997.

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

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

[Stateful-PCE] Crabbe, E., Medved, J., Minei, I., and R. Varga, "PCEP Extensions for Stateful PCE", draft-ietf-pce-stateful-pce-07 (work in progress), October 2013.

6.2. Informative References

[RFC4428] Papadimitriou, D., Mannie., E., ''Analysis of Generalized Multi-Protocol Label Switching (GMPLS)-based Recovery Mechanisms (including Protection and Restoration)'', RFC4428, March 2006.

[RFC5623] Oki., E., Takeda, T., Le Roux, JL., Farrel, A., ''Framework for PCE-Based Inter-Layer MPLS and GMPLS Traffic Engineering'', RFC5623, September 2009.

[RFC6952] Jethanandani, M., Patel, K., Zheng, L., ''Analysis of BGP, LDP, PCEP, and MSDP Issues According to the Keying and Authentication for Routing Protocols (KARP) Design Guide'', RFC6952, May 2013.

7. Authors' Addresses

Xian Zhang
Huawei Technologies

Email: zhang.xian@huawei.com

Haomian Zheng
Huawei Technologies

Email: zhenghaomian@huawei.com

Oscar Gonzalez de Dios
Telefonica I+D
Don Ramon de la Cruz 82-84
Madrid   28045
Spain
EMail: ogondio@tid.es

Victor Lopez
Telefonica I+D
Don Ramon de la Cruz 82-84
Madrid   28045
Spain
EMail: vlopez@tid.es

Contributor’s Address :

Dhruv Dhody
Huawei Technologies

Email: dhruv.dhody@huawei.com

Igor Bryskin
ADVA Optical

Email: IBryskin@advaoptical.com