PCE Working Group
Internet Draft
Category: Standards track

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

Expires: April 30, 2018 October 30, 2017

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

draft-zhang-pce-resource-sharing-05.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 is 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 responsible for path computation with such requirement. 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

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 [RFC8231]. 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: new LSP may request for resource sharing with one or multiple existing LSPs. Furthermore, if there is resource sharing between new LSP and existing LSP, the two LSPs cannot exist simultaneously, the new LSP will replace the existing LSP(s).

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. More specifically, this draft describes the resource sharing issue during the procedure when a new LSP is required to replace an existing LSP, which can be used together with Make-before-break (MBB) described in [RFC3209]. There are a few objects which indicate the resource sharing/disjoint relationships, such as SRLG and ASSOCIATE. However, these objects are used to describe the relationship with two simultaneous LSPs, instead of a new one and an old one, which is different with the object proposed in this draft.

As mentioned in [RFC8231], the PLSP-ID is unique during a PCEP session between PCC and PCE. Such identification is helpful in supporting the above resource sharing requirement for standardization of stateful PCEs. With a unique identifier, the configuration of PCCs is greatly simplified. Instead of determining all the resources to be shared, the PCC could request resource sharing directly from PCE.

The resource sharing can also be required in an inter-layer PCEP session. 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. Single PCE Use Case

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 should be applied for the path re-computation. For example, it might value resource sharing and prefer to share as much resource with the working LSP as possible and specify this policy in the PCReq message. If resources are shared between the old and new LSPs, there will be some 'interruption' when the traffic is switched from the old LSP to the new LSP. Here the resources to be shared mean the LSP information, which includes the node, link and corresponding SRLG information, etc.

On the other hand, in some scenarios there are different policies, for example the LSP should be restored without any interruption with best effort. An example can be found in Fig. 1 without failure on N2-N3 link, instead, an online re-optimization is needed for the working LSP (N1-N2-N3) from the stateful PCE. In such cases, the best choice is to set up a backup LSP for the working LSP with totally separate routing (for example N1-N5-N4-N3), and move the traffic to that backup LSP. After that the working LSP can be torn down, which will not result in any interruption during the optimization procedure. 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 PCC prefer to have less interruption, PCE 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 LSP may be different with the existing LSP, while resource sharing is still preferred by the PCC. 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. Multiple PCEs Use Case

Figure 2 shows a two-layer network example, with each layer managed by a PCE. 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

An inter-layer path computation example is shown in Fig. 2, assume a LSP (LSP1: H2-H3) has been established already, visible as H2-H3 from view of higher-layer PCE and H2-L1-L2-H3 from the global view (or from the view of lower-layer PCE). 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 higher-layer PCE will be forwarded to lower-layer PCE since there is no resource readily available in the higher layer. So it leaves the lower-layer PCE to compute a path in the lower layer in order to support the higher layer request. In this case, lower-layer PCE is required to compute a path between H2 and H5 under the constraint that it can share the resource with that of the LSP1. At this moment the lower-layer PCE has the knowledge on the explicit routing that LSP1 go through (H2-L1-L2-H3), and therefore can map the lower layer LSP with the higher-layer one. So when lower-layer PCE computes the path for LSP2, it can consider the resource used by LSP1 as available with higher priority. For example, lower-layer PCE may choose H2-L1-L2-L4-H5 as the computation result. On the other hand, if the path computation policy is to have a separate path with LSP1, the lower-layer PCE may choose H2-L1-L3-L4-H5.

During this procedure higher-layer PCE can only use LSP1 information (such as its five-tuple LSP information) as the information, an issue to solve is how lower-layer PCE 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 lower-layer 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 lower-layer PCE 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 layer LSP information itself, such as mapping it to the corresponding LSP in the lower layer, in this way lower-layer PCE does 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 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. Association group and type

According to the definition in [ietf-pce-association-group], the association group is used to associate multiple LSPs into one group for further path computation considerations, such as disjointness and resource sharing. An association ID will be used to identify the resource sharing group. In this draft, a new association type is defined as:

    Association type = TBD1 ("Sharing Association Type").

A sharing group should have multiple LSPs. The number of LSPs and the criteria for how LSPs share among each other are implementation dependent. Local path computation policies apply to different PCE and PCC, some examples can be found in section 2.

3.2. Resource Sharing TLV

The PCEP Resource Sharing group MUST carry the following TLV. It MAY be carried within a PCReq message from the network element (or other PCCs) so as to indicate the desired resource sharing requirements to be applied by the stateful PCE during path computation.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Type = TBD2           |            Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 Flags                                   |S|N|L|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          Optional TLVs                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Currently the following flags have been defined:

* L (Link share) bit: when set, this flag indicates that the PCE should prioritize the links that shared by existing LSPs within the sharing group for path computation.

* N (Node share) bit: when set, this flag indicates that the PCE should prioritize the nodes that shared by existing LSPs within the sharing group for path computation.

* S (SRLG share) bit: bit: when set, this flag indicates that the PCE should set the SRLG (Shared Risk Link Group) of the computed LSP to the same as existing LSPs within the sharing group for path computation.

Optional TLVs may be needed to indicate the LSP(s) with which the resource is shared. If multiple LSPs are required, the PCE may need to consider different sharing policies, which is implementation dependent and may result in a different computing result. The selection policy among multiple computation result is out of the scope of this draft.

3.3. Processing Rules

To request a path allowing sharing resource with one or multiple existing LSPs, a PCC includes a Resource Sharing TLV in the association group object in the PCReq message.

On receipt of a PCReq message with a Resource Sharing TLV, a stateful PCE MUST proceed as follows:

- If the Resource Sharing TLV 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 Resource Sharing TLV 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 TLV is extracted correctly, the PCE MUST apply the requested resource sharing requirement.

The procedure of setting flags follows the rules defined in Section 3.1. The RSO flags 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.

It is worth noting that the Resource Sharing TLV can be used together with other path indication objects like IRO/XRO. The difference is, the use of Resource Sharing TLV is to setup an alternative path, instead a new path. It is also dependent on the knowledge of PCC, e.g., if the PCC have a full knowledge of the path information and have strong preference on the route, it may send the PCReq with IRO message to specify the route. On the other hand, if the PCC does not know how the path should go but just want to set up a new LSP to replace the old one, it may use the Resource Sharing TLV instead of IRO.

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 Resource Sharing TLV in association group object 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 a Resource sharing TLV 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 for resource sharing, and such 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 share 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. Association Object Type Indicators

This document defines a new association type, with the following information:

Object Class	Name	Object Type	Reference
TBA1	Sharing-group	Association Type	[this document]

5.2 PCEP TLV Definitions

This document defines the following TLVs to support the resource sharing scenario:

Value	Name	Reference
TBA2	Resource-sharing TLV	[this document]

IANA is requested to allocate the following bit numbers in the flag spaces of Resource-sharing TLV:

Bit	Flag name	Reference
0	Link Share	[this document]
1	Node Share	[this document]
2	SRLG Share	[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.

[RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and Swallow, G., "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.

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

[RFC8231] Crabbe, E., Medved, J., Minei, I., and R. Varga, "PCEP Extensions for Stateful PCE", RFC8231, June 2017.

[ietf-pce-association-group] Minei, I., Crabbe E., Sivabalan S., Ananthakrishnan H., Dhody D., Tanaka Y., "PCEP Extensions for Establishing Relationships Between Sets of LSPs".

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.

[RFC4655] Farrel, A., Vasseur, J.-P., and Ash, J., "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 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
Huawei Technologies
Email: Igor.Bryskin@huawei.com