Network Working Group Kexin Tang
Internet-Draft Xuerong Wang
Intended status: Informational Xuping Cao
Expires: May 3, 2012 ZTE Corporation
Oct 31, 2011

Stateful PCE
draft-tang-pce-stateful-pce-02.txt

Abstract

A PCE can be either stateful or stateless. The information carried in a stateful PCE is more detailed than that of a stateless PCE. This draft focus on stateful PCE, describes the problems without stateful PCE, and gives the IGP and PCEP extensions to realize stateful PCE.

Table of Contents

1. Introduction

As defined in section 6.8 of [RFC4655], a PCE can be either stateful or stateless. For stateful PCE, there is a strict synchronization between the PCEs not only the network states (in term of topology and resource information), but also the set of computed paths and reserved resources in use in the network. Since stateful PCE has more network information, it can be used to do some complicated work.

However, the existing PCE discovery protocol ([RFC5088], [RFC5089]) and PCEP ([RFC5440]) do not support stateful PCE. And there is no effective synchronization mechanism defined to realize stateful PCE yet. For these sufficient reasons, this document focus on stateful PCE, describes the applicability of stateful PCE and gives the IGP and PCEP extensions to support stateful PCE.

1.1. Conventions used in this document

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

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

- PCE: Path Computation Element. 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.

- PCED: PCE Discovery.

- PCEP: Path Computation Element communication Protocol.

- TED: Traffic Engineering Database, which contains the topology and resource information of the domain. The TED may be fed by Interior Gateway Protocol (IGP) extensions or potentially by other means.

3. Problems

This section lists the typical problems without stateful PCE. As described in [RFC4655], stateful PCE uses information not only from the TED, but also information about existing paths (for example, TE LSPs) in the network when processing new requests, so the information

carried in stateful PCE are more detailed than that of stateless PCE.

Without stateful PCE, there would not be a strict synchronization between PCE and LSP state. Consequently,the PCE may not know the existing path in the network in time,and suppose the network resource taken by the existing path is unoccupied, so it would consider these resource in the other path computations. Resource conflict occurs under this condition.

A PCE may received several PCReq messages, this may occurs when a PCE is required to compute a large number of path for several LSPs,for example when fiber cutting happened or in a sudden accident,which may result in a large number of links down simultaneity,and need to be recovery in a short time.

Figure 1 shows a demonstration topology for the subsequent context.

```
                        +-----+               +-----+
                        |  A  |---------------|  B  |
                        +-----+               +-----+
                                \            /
                                 \          /
                                   +-----+
                                   |  C  |
                                   +-----+
```

Figure 1: demonstration topology

In Figure1,we make the assumptions as follows:the link capacity between all these three nodes is 100M,and the node that hosts the PCE function received a PCReq massage,which required it to compute paths for two LSP respectively:LSP1 and LSP2. LSP1 and LSP2 share the same original and terminal node pair,that is from node A to node B,and require the same link bandwidth:80M.

For LSP 1,to get a shortest path,the PCE may return a path:A--B,while the computation result is not synchronized in its TED in time,the PCE assume the unreserved bandwidth of the link between A and B is still 100M.

Then for LSP2,which require 80M link capacity either,the PCE may still consider the shortest path,that is A--B as before,and it seems that in the TED currently,this shortest path has enough capacity for LSP2,so the PCE returns the optimal path A--B for LSP2.

Once the LSP setup procedure by signaling for LSP1 and LSP2 is running,the setup procedure for the first LSP would successful, while

the latter one would encounter resource conflict:there would not be enough link capacity for the LSP to be setup,and it would return a signaling failure notification message.

As for Inter-area or Inter-AS path computation for a LSP, it is often involved in multiple PCEs cooperation to compute an end-to-end path, for example, BRPC ([RFC5441]) and H-PCE ([H-PCE-FWK]). Resource conflict would even worse in this situation because it takes extra time for communication between the cooperative PCEs.

4. Protocol Extensions

4.1. PCED Extensions

[RFC5088] defines extensions to OSPFv2 ([RFC2328]) and OSPFv3 ([RFC5340]) to allow a PCE in an OSPF routing domain to advertise some information useful to a PCC for PCE selection. It defines a new TLV (named the PCE Discovery TLV (PCED TLV)) to be carried within the OSPF Router Information LSA ([RFC4970]). The type 5 sub-TLV of PCED TLV, which named CE-CAP-FLAGS sub-TLV, used to indicate the capabilities of a PCE. It contains eight capabilities currently, but not includes the stateful capability of a PCE. So the PCE in an OSPF routing domain cannot advertise its state capability information to a PCC，which would help the PCC to make advanced and informed choices in PCE selection.

To discover stateful PCE, a PCC SHOULD know whether PCE is stateful. Therefore, the PCE discovery message SHOULD indicate whether the PCE advertising this message is a stateful PCE. Since PCE-CAP-FLAGS Sub-TLV ([RFC5088] for OSPF, [RFC5089] for IS-IS) contains PCE Capability Flags, this document defines a new flag, Stateful PCE Capability Flag, as follows (need to be assigned by IANA):

Bit	Capabilities
TBD	Stateful PCE

4.2. LSP State Synchronization

With respect to the definition of stateful PCE defined in [RFC4655], a stateful PCE utilizes information from the TED as well as information about existing paths (for example, TE LSPs) in the network when processing new path computation requests. Therefore, a stateful PCE SHOULD know the status (created or deleted) of a LSP. For this reason, there SHOULD be a timely synchronization of LSP state between PCC and PCE, including the path computation result, and the LSP setup or deletion result. For this purpose, this section

makes an extension to the PCNtf message defined in [RFC5440], defines a new Notification-type as follows (need to be assigned by IANA):

o Notification-type=TBD: LSP Status

  * Notification-value=1: end-to-end path computation success(This value is available for distributed path computation only). When a optimal end-to-end path is computed successfully, the head-end PCE SHOULD send a notification message with Notification-type=TBD and Notification-value=1 to the other PCEs collaboratively in the PCE chain which computed the other path segments successfully.

  * Notification-value=2: end-to-end path computation failure(This value is available for distributed path computation only). If an end-to-end path computation is failed,the head-end PCE SHOULD send a notification message with Notification-type=TBD and Notification-value=2 to the other PCEs collaboratively in the PCE chain which computed the other path segments successfully.

  * Notification-value=3: LSP setup success. When a LSP is created successfully by signaling, the PCC SHOULD send a notification message with Notification-type=TBD and Notification-value=3 to all the PCEs.

  * Notification-value=4: LSP setup failure. When a LSP is failed to setup by signaling, the PCC SHOULD send a notification message with Notification-type=TBD and Notification-value=4 to all the PCEs.

  * Notification-value=5: LSP delete success. When a LSP is delete in the network successfully, the PCC SHOULD send a notification message with Notification-type= TBD and Notification-value=5 to all the PCEs.

  * Notification-value=6: LSP delete failure . When a LSP path is failed to delete, the PCC SHOULD send a notification message with Notification-type= TBD and Notification-value=6 to all the PCEs.

This draft extended the PCNtf message, added the Path object which defined for PCRep message [RFC5440] to the PCNtf message ,to identify the ERO and the attribute of a LSP.

The format of the extended PCNtf message is as follows:

```
   <PCNtf Message>::=<Common Header>
                     <notify-list>

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

   <notify>::= [<request-id-list>]
                <notification-list>

   <request-id-list>::=<RP><PATH> [<request-id-list>]

   <PATH>::= <ERO><attribute-list>
```

where:

```
    <attribute-list>::=[<LSPA>]
                       [<BANDWIDTH>]
                       [<metric-list>]
                       [<IRO>]

    <notification-list>::=<NOTIFICATION>[<notification-list>]
```

With the newly added Path object which carries the ERO object and the attribute of a LSP,a PCE can identifies the status of which LSP changes,and the information of the LSP, so as to identifies the resource that taken by the LSP.

5. Security Considerations

The extensions of this draft are based on PCEP and OSPF, only some optional protocol elements are added which will not change the security of existing network.

6. IANA Consideration

The extensions of this draft is baed on PCEP and OSPF, only some optional protocol elements are added which will not change the security of existing network.

7. Normative References

[H-PCE-FWK]
Daniel King, Adrian Farrel, Quintin Zhao, and Fatai Zhang,

"The Application of the Path Computation Element Architecture to the Determination of a Sequence of Domains in MPLS and GMPLS", draft-ietf-pce-hierarchy-fwk-00.txt .

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

[RFC2328] Moy, J., "OSPF Version 2", STD 54, RFC 2328, April 1998.

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

[RFC4970] Lindem, A., Shen, N., Vasseur, JP., Aggarwal, R., and S. Shaffer, "Extensions to OSPF for Advertising Optional Router Capabilities", RFC 4970, July 2007.

[RFC5088] Le Roux, JL., Vasseur, JP., Ikejiri, Y., and R. Zhang, "OSPF Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5088, January 2008.

[RFC5089] Le Roux, JL., Vasseur, JP., Ikejiri, Y., and R. Zhang, "IS-IS Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5089, January 2008.

[RFC5340] Coltun, R., Ferguson, D., Moy, J., and A. Lindem, "OSPF for IPv6", RFC 5340, July 2008.

[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.

Authors' Addresses

Kexin Tang
ZTE Corporation
No.50 Software Avenue, Yuhuatai District
Nanjing, Jiangsu 210012
P.R.China

Phone: +86-025-88014225
Email: tang.kexin@zte.com.cn

Xuerong Wang
ZTE Corporation
R&D Building 3, ZTE Industrial Zone, Liuxian Road,Nanshan District
Shenzhen 518055
P.R.China

Phone: +86-755-26773926
Email: wang.xuerong@zte.com.cn

Xuping Cao
ZTE Corporation
21F, ZTE Plaza, No.19 East Huayuan Road,Haidian District
Beijing 100191
P.R.China

Email: cao.xuping@zte.com.cn