Network Working Group
Internet-Draft
Intended status: Informational
Expires: January 2, 2015

V. Lopez
O. Gonzalez de Dios
Telefonica I+D
D. King
Old Dog Consulting
S. Previdi
Cisco Systems, Inc.
J. Tantsura
Ericsson

July 1, 2014

Traffic Engineering Database dissemination for Hierarchical PCE scenarios
draft-lopez-pce-hpce-ted-02

Abstract

The PCE architecture is well-defined and may be used to compute the optimal path for LSPS across domains in MPLS-TE and GMPLS networks. The Hierarchical Path Computation Element (H-PCE) [RFC6805] was developed to provide an optimal path when the sequence of domains is not known in advance. The procedure and mechanism for populating the Traffic Engineering Database (TED) with domain topology and link information used in H-PCE-based path computations is open to interpretation. This informational document describes how topology dissemination mechanisms may be used to provide TE information between Parent and Child PCEs (within the H-PCE context). In particular, it describes how BGP-LS might be used to provide inter-domain connectivity. This document is not intended to define new extensions, it demonstrates how existing procedures and mechanisms may be used.

Status of This Memo

Table of Contents

1. Introduction

In scenarios with multiple domains in both MPLS-TE and GMPLS networks, the hierarchical Path Computation Element (H-PCE) Architecture, defined in [RFC6805], allows to obtain the optimum end-to-end path. The architecture exploits a hierarchical relation among domains.

[RFC6805] defines the architecture and requirements for the end-to-end path computation across domains. The solution draft for the H-PCE [I-D.draft-ietf-pce-hierarchy-extensions] is focused on the

PCEP protocol extensions to support such H-PCE procedures, including negotiation of capabilities and errors. However, neither the architecture nor the solution draft specify which mechanism must to be used to build and populate the parent PCE (pPCE) Traffic Engineering Database (TED).

The H-PCE architecture documents define the minimum content needed in the traffic engineering database required to compute paths. The information required by parent TEDB are identified in [RFC6805] and further elaborated in [I-D.draft-ietf-pce-inter-area-as-applicability]. For instance, [RFC6805] and [I-D.draft-ietf-pce-inter-area-as-applicability] suggest that BGP-LS could be used as a "northbound" TE advertisement. This means that a PCE does not need to listen IGP in its domain, but its TED is populated by messages received (for example) from a Route Reflector. [I-D.draft-ietf-idr-te-pm-bgp] extends BGP-LS to disseminate traffic engineering information. The parameters considered are: delay, packet loss and bandwidth.

This document highlights the applicability of BGP-LS to the dissemination of domain topology within the H-PCE architecture. In particular, it describes how can BGP-LS be used to send the inter-domain connectivity. It also shows how can OSPF-TE and ISIS-TE updates be mapped into BGP-LS.

Note that this document is not intended to define new protocol extensions, it is an informational document and where required it highlights where existing mechanisms and protocols may be applied.

1.1. Parent PCE Domain Topology

The pPCE maintains a domain topology map of the child domains and their interconnectivity. This map does not include any visibility into the child domains. Where inter-domain connectivity is provided by TE links, the capabilities of those links may also be known to the pPCE. The pPCE maintains a TED for the parent domain, the nodes in the parent domain are abstractions of the cPCE domains (connected by real or virtual TE links), but the pPCE domain may also include real nodes and links.

The procedure and protocol mechanism for disseminating and construction of the pPCE TED may be provided using a number of mechanisms, including manually configuring the necessary information or automated using a separate instance of a routing protocol to advertise the domain interconnectivity. Since inter-domain TE links can be advertised by the IGPs operating in the child domains, this information could then be exported to the parent PCE either by the child PCEs or using north-bound export mechanisms.

1.2. Parent PCE TED requirements

The information that would be exchanged includes:

- o Identifier of advertising child PCE.
- o Identifier of PCE’s domain.
- o Identifier of the link.
- o TE properties of the link (metrics, bandwidth).
- o Other properties of the link (technology-specific).
- o Identifier of link endpoints.
- o Identifier of adjacent domain.

2. H-PCE Domain Topology Dissemination and Construction Methods

A variety of methods exist to provide are different alternatives so the parent PCE can get the topological information from the child PCEs (cPCEs):

- o Statically configure all inter-domain link and topology information.
- o Membership of an IGP instance. The necessary topological information could be disseminated by joining the IGP instance of each child PCE domain. However, by doing so, it would break the domain confidentiality principles and is subject to scalability issues.
- o PCEP Notification Messages. Another solution is to send the interconnection information between domains using PCEP Notifications (see section 4.8.4 of [RFC6805]). One approach, followed in research work, is embedding in PCEP Notifications the Inter-AS OSPF-TE Link State Advertisements (LSA) to send the Inter-Domain Link information from child PCEs to the parent PCE and to send reachability information (list of end-points in each domain). However, it is argued that the utilization of PCEP to disseminate topology is beyond scope of the protocol.
- o Separate IGP instance. [RFC6805] points out that in models such as ASON it is possible to consider a separate instance of an IGP running within the parent domain where the participating protocol speakers are the nodes directly present in that domain and the PCEs (parent and child PCEs).

o Use north-bound distribution of TE information. The North-Bound Distribution of Link-State and TE Information using BGP has been recently propose in the IEFT [I-D.draft-ietf-idr-ls-distribution]. This approach is known as BGP-LS and defines a mechanism by which links state and traffic engineering information can be collected from networks and exported to external elements using the BGP routing protocol. By using BGP-LS as northbound distribution mechanism, there would be a BGP speaker in each domains that sends the necessary information to a BGP speaker in the parent domain. This architecture is further elaborated in this document.

3. H-PCE architecture using BGP-LS

As mentioned in [I-D.draft-dugeon-pce-ted-reqs] PCE has to retrieve Traffic Engineering (TE) information to carry out its path computation. This is required not only for intra-domain information, which can be got using IGP (like OSPF-TE or ISIS-TE), but also for inter-domain information in the Hierarchical PCE (H-PCE) architecture.

Figure 1 shows an example of a H-PCE architecture. In this example, there is a parent PCE and three child PCEs, and they are organized in multiple domains. The parent PCE does not have information of the whole network, but is only aware of the connectivity among the domains and provides coordination to the child PCEs. Figure 2 shows which is the visibility that parent PCE has from the network according to the definition in [RFC6805].

Thanks to this topological information, when there is a request to a child PCE with the destination in another domain, this path request is sent to the parent PCE, which selects a set of candidate domain paths and sends requests to the child PCEs responsible for these domains. Then, the parent PCE selects the best solution and it is transmitted to the source PCE.

```
     -----------------------------------------------------------------
     |   Parent PCE Domain                                            |
     |                               -----                            |
     |                              |pPCE |                           |
     |                               -----                            |
     |                                                                |
     |    ----------------    ----------------    ----------------    |
     |   | Domain 1       |  | Domain 2       |  | Domain 3       |   |
     |   |                |  |                |  |                |   |
     |   |      ------    |  |      ------    |  |      ------    |   |
     |   |     |cPCE 1|   |  |     |cPCE 2|   |  |     |cPCE 3|   |   |
     |   |      ------    |  |      ------    |  |      ------    |   |
     |   |                |  |                |  |                |   |
     |   |            ----|  |----        ----|  |----            |   |
     |   |           |BN11+---+BN21|     |BN23+---+BN31|          |   |
     |   |            ----|  |----        ----|  |----            |   |
     |   |                |  |                |  |                |   |
     |   |            ----|  |----        ----|  |----            |   |
     |   |           |BN12+---+BN22|     |BN24+---+BN32|          |   |
     |   |            ----|  |----        ----|  |----            |   |
     |   |                |  |                |  |                |   |
     |    ----------------    ----------------    ----------------    |
      -----------------------------------------------------------------
```

Figure 1: Example of Hierarchical PCE architecture

```
                        ----------------------------
                       | Parent PCE view            |
                       |           ----             |
                       |          |pPCE|            |
                       |           ----             |
                       |                            |
                       |  ----     ----     ----    |
                       | |    |---|    |---|    |   |
                       | | D1 |   | D2 |   | D3 |   |
                       | |    |---|    |---|    |   |
                       |  ----     ----     ----    |
                        ----------------------------
```

Figure 2: Parent PCE topology information

Thanks to the dissemination of inter-domain adjacency information from each cPCE to the pPCE, the pPCE can have a view of reachability between the domains. The H-PCE architecture with BGP-LS is shown in

Figure 3. Each domain has a cPCE that is able to compute paths in the domain. This child PCE has access to a domain TED, which is built using IGP information. In each domain, a BGP speaker has access to such domain TED and acts as BGP-LS Route Reflector to provide network topology to the pPCE. Next to the pPCE, there is a BGP speaker that maintains a BGP session with each of the BGP speakers in the domains to receive the topology and build the parent TED. A policy can be applied to the BGP-LS speakers to decide which information is sent to its peer speaker. The minimum amount of information that needs to be exchanged is the inter-domain connectivity, including the details of the Traffic Engineering Inter-domain Links [RFC6805]. With this information, the parent PCE is able to have access to a domain topology map and its connectivity. Additionally, the BGP-LS speaker can be configured to send some intra-domain information for virtual or candidate paths with some TE information. In this case, the parent PCE has access to an extended database, with visibility of both intra-domain and inter-domain information and can compute the sequence of domains with better accuracy.

BGP-LS [I-D.draft-ietf-idr-ls-distribution] extends the BGP Update messages to advertise link-state topology thanks to new BGP Network Layer Reachability Information (NLRI). The Link State information is sent in two BGP attributes, the MP_REACH (defined in [RFC4670]) and a LINK_STATE attribute (defined in [I-D.draft-ietf-idr-ls-distribution]). To describe the inter domain links, in the MP_REACH attribute, a Link NLRI can be used with the local node descriptors the address of the source, and in the remote descriptors, the address of the destination of the link. The Link Descriptors field has a TLV (Link Local/Remote Identifiers), which carries the prefix of the Unnumbered or Numbered Interface. In case of the message informs about an intra-domain link, the standard traffic engineering information is included in the LINK_STATE attribute.

```
  ----------------------------------------------------------------------
  |  Parent PCE Domain                                                 |
  |                           -------     -----    -----               |
  |                   -------+BGP-LS +---+ TED +--+pPCE |              |
  |                  /       |Speaker|    -----    -----               |
  |                 /         --+---+                                  |
  |                /            |    \_________________                |
  |               /             |                      \               |
  |  ------------/--------    ----+------------     ------+------------  |
  | | Domain 1  /         | |    |  Domain 2 |   |       |   Domain 3 | |
  | |          /          | |    |           |   |       |            | |
  | |   ------+           | |   -+-----      |   |    ---+---         | |
  | |  |BGP-LS |          | |  |BGP-LS |     |   |   |BGP-LS |        | |
  | |  |Speaker|          | |  |Speaker|     |   |   |Speaker|        | |
  | |   ---+---           | |   ---+---      |   |    ---+---         | |
  | |      |              | |      |         |   |       |            | |
  | |   ---+---    ------ | |  ---+-- ------ |   |   ---+---  ------  | |
  | |  |  TED  +-+cPCE 1| | | | TED +-+cPCE| |   |  |  TED +-+cPCE 1| | |
  | |   ---+---    ------ | |  ---+-- ------ |   |   ---+---  ------  | |
  | |      |              | |     |          |   |      |             | |
  | |   ---+---           | |  ---+---       |   |   ---+---          | |
  | |  |  IGP  |          | | |  IGP  |      |   |  |  IGP  |         | |
  | |  |  Peer |          | | |  Peer |      |   |  |  Peer |         | |
  | |   -------           | |  -------       |   |   -------          | |
  | |      |              | |     |          |   |      |             | |
  |  ------+---------------  -----+-----------    ------+------------  |
  |        |                      |                     |              |
  |    -------------------    ----------------    -------------------  |
  |   |     Domain 1      |  |   Domain 2     |  |     Domain 3     |  |
  |    -------------------    ----------------    -------------------  |
  ----------------------------------------------------------------------
```

Figure 3: Example of Hierarchical PCE architecture with BGP-LS

4. Including inter-domain connectivity in BGP-LS

In order for the parent PCE to carry out the path computation tasks it needs the inter-domain topology between the child domain scenarios. This topology is learnt through IGP by each BGP-LS speaker. The Traffic Engineering extensions (OSPF-TE or ISIS-TE) allow IGP to carry link state information that can be used in optimizing technics such as the PCE algorithms. However, the parent PCE does not require such TE information, but just connectivity between the domains. However, TE information within the domain could be disseminated to the parent PCE to reduce the queries to the child PCEs.

4.1. Mapping from OSPF-TE

Carrying TE information in OSPF is a well-known standardized feature [RFC3630]. This section explains how this information can be exported outside one IGP domain using BGP-LS. BGP-LS extends the BGP Update messages to advertise link-state topology thanks to the new BGP Network Layer Reachability Information (NLRI) and BGP-LS attribute.

The BGP NLRI carries the descriptors used to define the element in question (e.g. link or node) and the BGP-LS attribute carries the chosen parameters to characterize the described element. Information is codified using multiple TLV triplets just as the ones used in OSPF-TE making it easy to integrate. For the purpose of this document, we consider a scenario where there is an origin (router) with the correspondent IPv4, a destination with its IPv4 and a link having the following TE parameters: maximum BW, maximum reservable BW and unreserved BW.

4.1.1. Node Descriptors

In the OSPF packet, there are two fields that tell us the origin and destination node IDs. The origin IP is the Source OSPF Router ID in the OSPF header and this is mapped into the IGP Router ID subTLV inside the Local Node Descriptors field [I-D.draft-ietf-idr-ls-distribution]. The destination IP is found as the Link ID field in the MPLS LSA in OSPF. This is mapped into the correspondent IGP Router ID in the Remote Node Descriptors field [I-D.draft-ietf-idr-ls-distribution].

There are other subTLVs inside the Local/Remote Node Descriptors but they are not relevant for this document.

4.1.2. Link Descriptors

The only two TLVs in the Link Descriptors field to map from OSPF are the local and remote interface addresses. This information is mapped directly from the Local/Remote Interface address TLV carried in the MPLS LSA of OSPF into the Local/Remote Interface address subTLV of the Link Descriptors field.

The same procedure must be applied for unnumbered interfaces but utilizing the Link Local/Remote Identifiers TLV.

4.1.3. Mapping OSPF TE parameters into BGP-LS attribute

As mentioned before, these parameters are not required in the H-PCE scenario. They are just required to reduce the number of queries to the children PCEs. The parent PCE can use bandwidth information between two domains to request for some possible connections instead of all.

The BGP-LS attribute will be a set of TLV triplets carrying the desired TE parameters learnt by OSPF. Bandwidth parameters are used to illustrate the example but they are many more (like, available labels).

The BGP-LS attribute is mapped in the following way. The TLVs carried in the MPLS-TE LSA in OSPF are directly translated into the equivalent TLVs in BGP-LS. As such, the Unreserved BW TLV in OSPF is mapped into the Unreserved BW TLV in BGP-LS. The same happens with the Maximum BW TLV and the Maximum Reservable BW TLV.

4.2. Mapping from ISIS-TE

TBD

5. Manageability Considerations

TBD

6. Security Considerations

TBD

7. References

7.1. Normative References

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

[RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, September 2003.

[RFC4670] Nelson, D., "RADIUS Accounting Client MIB for IPv6", RFC 4670, August 2006.

[RFC6805] King, D. and A. Farrel, "The Application of the Path Computation Element Architecture to the Determination of a Sequence of Domains in MPLS and GMPLS", RFC 6805, November 2012.

7.2. Informative References

[I-D.draft-dugeon-pce-ted-reqs]
Dugeon, O., Meuric, J., Douville, R., Casellas, R., and O. Gonzalez de Dios, "Path Computation Element (PCE) Traffic Engineering Database (TED) Requirements", February 2014.

[I-D.draft-ietf-idr-ls-distribution]
Gredler, H., Medved, J., Previdi, S., Farrel, A., and S. Ray, "North-Bound Distribution of Link-State and TE Information using BGP", November 2013.

[I-D.draft-ietf-idr-te-pm-bgp]
Wu, Q., Wang, D., Previdi, S., Gredler, H., and S. Ray, "BGP attribute for North-Bound Distribution of Traffic Engineering (TE) performance Metrics", January 2014.

[I-D.draft-ietf-pce-hierarchy-extensions]
Zhang, F., Zhao, Q., Gonzalez de Dios, O., Casellas, R., and D. King, "Extensions to Path Computation Element Communication Protocol (PCEP) for Hierarchical Path Computation Elements (PCE)", July 2013.

[I-D.draft-ietf-pce-inter-area-as-applicability]
King, D., Meuric, J., Dugeon, O., Zhao, Q., and O. Gonzalez de Dios, "Applicability of the Path Computation Element to Inter-Area and Inter-AS MPLS and GMPLS Traffic Engineering", February 2013.

Authors' Addresses

Victor Lopez
Telefonica I+D
Don Ramon de la Cruz 82-84
Madrid 28045
Spain

Phone: +34913128872
Email: vlopez@tid.es

Oscar Gonzalez de Dios
Telefonica I+D
Don Ramon de la Cruz 82-84
Madrid 28045
Spain

Phone: +34913128832
Email: ogondio@tid.es

Daniel King
Old Dog Consulting
UK

Email: daniel@olddog.co.uk

Stefano Previdi
Cisco Systems, Inc.
Via Del Serafico 200
Rome 00144
IT

Email: sprevidi@cisco.com

Jeff Tantsura
Ericsson
USA

Email: jeff.tantsura@ericsson.com