Network Working Group   JP Vasseur
Internet Draft   Peter Psenak
Document: draft-vasseur-mpls-ospf-te-cap-00.txt   Cisco Systems, Inc

IETF Internet Draft
Expires: May, 2003

October 2002

OSPF Traffic Engineering capability TLVs

draft-vasseur-mpls-ospf-te-cap-00.txt

Status of this Memo

Abstract

This draft proposes OSPF traffic engineering capability TLVs. Two capability TLVs are defined in the current draft: the Path Computation Server Discovery (PCSD) TLV that allows a router to announce its Path Computation Server capability to other LSRs within an OSPF area or a routing domain and the Mesh-group TLV used by an LSR to indicate its desire to participate to a mesh of Traffic Engineering Label Switched Path (this mesh of TE LSPs is identified by a mesh-group number). They are both used in the context of MPLS Traffic Engineering. Additional OSPF TE capability TLVs may be added in further revision of this draft. Those OSPF TE capability TLVs will be carried within the OSPF router information LSA (opaque type

of 4, opaque ID of 0) defined in [18].

1. Where does this draft fit in the picture of the Sub-IP and OSPF WG ?

This document specifies OSPF extensions in support of MPLS Traffic Engineering. It will be discussed in the CCAMP Working Group with a review of the OSPF Working Group.

2. Terminology

Terminology used in this draft

LSR: Label Switch Router

PCS: Path Computation Server (may be an LSR (ABR, ASBR, ...) or a dedicated path computation server (typically a UNIX machine) not forwarding packet.

PCC: Path Computation Client (any head-end LSR) requesting a path computation to the Path Computation Server.

TE LSP: Traffic Engineering Label Switched Path

Head-end TE LSP: head/source of the TE LSP

Tail-end TE LSP: tail/destination of the TE LSP

Intra-area TE LSP: TE LSP whose head-end and tail-end reside in the same area

Inter-area TE LSP: TE LSP whose head-end and tail-end reside in different areas (the TE LSP spans areas)

Inter-AS TE LSP: TE LSP whose head-end and tail-end reside in different Autonomous Systems (the TE LSP spans AS)

3. Introduction

This section describes the usage of the two OSPF capabilities TLV: the Path Computation Server Discovery TLV and the Mesh-Group TLV. Those OSPF TE capability TLVs will be carried within the OSPF router information LSA (opaque type of 4, opaque ID of 0) defined in [18].

3.1 Path Computation Server Discovery (PCSD) TLV

This draft specifies a new OSPF TE capability TLV called the PCSD TLV for the Auto-discovery of one or more Path Computation Server(s). In various situations (GMPLS, inter-area TE, ...), an LSR may send a request to a Path Computation Server (PCS) to compute one or more Traffic Engineering LSP paths obeying a set of specified constraints.

[4] specifies RSVP extensions:
- for a PCC to send path computation requests to a PCS to compute TE LSP(s) obeying a set of specified constraints,
- for the PCS to provide to the PCC one or more computed paths obeying the set of constraints (or to return an indication mentioning no path obeying the constraints could be found).

The scope of this document is to define a new OSPF TE capability TLV carried within an OSPF router information LSA such that a LSR/centralized path computation tool may announce its capability to be a Path Computation Server within an area or an Autonomous System. This allows every LSR in the network to automatically discover the Path Computation Server(s), which substantially simplifies head-end LSRs configuration. Moreover, this allows to detect dynamically any new PCS or that a PCS is no longer active.

3.2 Mesh-group TLV

As of today, there are different approaches in deploying MPLS Traffic Engineering:

(1) the systematic approach consisting of setting up a full

mesh of tunnels between P or PE routers, with the objective of optimizing the bandwidth usage in the core,
(2) the "by exception" approach where a set TE LSPs are set up on hot spots to alleviate a congestion resulting in an unexpected traffic growth in some part of the network.

The systematic approach requires setting up a full mesh of TE LSPs, which implies the configuration of a large number of tunnels on every Hean-End LSR (P or PE LSR). A full TE mesh of n LSRs requires the set up of O(n^2) TE LSPs. Furthermore, the addition of any new LSR in the mesh implies to configure n TE LSPs on the new LSR and to add a new TE LSP on every LSR ending to this new LSR, which gives a total of 2*n TE LSPs. This is not only time consuming but also not a low risk operation for Service Providers. So a more automatic way of setting up full mesh of TE LSP might be desirable. This requires to define a new TE capability TLV (called the Mesh-group TLV) such that an LSR can announce its desire to join a particular TE LSP mesh, identified by a mesh-group number.

4. PCSD and Mesh-group TLV formats

This section defines the PCSD and the Mesh-group TLV formats carried in an OSFP router information LSA as defined in [18].

The PCSD and the Mesh-group TLV have the following format:

```
 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             |             length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
//                            Value                            //
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Where

Type: identifies the TLV type
Length: length of the value field in octets

The format of the TLV is the same as the TLV format used by the Traffic Engineering Extensions to OSPF [5]. The TLV is padded to four-octet alignment; padding is not included in the length field (so a three octet value would have a length of three, but the total size of the TLV would be eight octets). Nested TLVs are also 32-bit aligned. Unrecognized types are ignored. All types between 32768 and 65535 are reserved for vendor-specific extensions. All other undefined type codes are reserved for future assignment by IANA.

Note that a sub-TLV is similar to a TLV: TLV are carried within an LSA as sub-TLVs are carried within TLVs. Each sub-TLV describes a particular Path Computation Server capability. In the rest of this document both terms will be used interchangeably.

The PCSD TLV type is 2. The PCSD TLV is made of a set of non ordered TLVs each having the same format as described above.

The Mesh-group TLV type is 3. The Mesh-group TLV does not have any sub-TLV currently defined.

4.1 PCSD sub-TLVs

This section defines the sub-TLVs carried within the PCSD TLV payload.

The PCSD TLV is made of various non ordered TLVs defined bellow:

TLV type	Length	Name
1	4	Path computation server scope TLV
2	variable	Path computation server address TLV
3	8	Path computation server capability TLV
4	8	AS-domain TLV

Any non recognized TLV must be silently ignored.

4.1.1 Path computation server scope TLV

The path computation server scope TLV specifies the zone for which the path computation server is capable of performing TE LSP path computation.

The path computation server scope TLV type is 1, its length is 4, and the value is a set of flags:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              1                |               4               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Reserved            |     Flags     |         |A|I|L|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Path computation server scope TLV format

Flags: no flags are currently defined.

Scope

L (local scope). When set, this flag indicates that the PCS can compute paths for the area the LSA is flooded into (i.e the PCS can compute TE LSP path for intra-area TE LSPs).

I (inter-area scope). When set, the PCS can perform TE LSP path computation for inter-area TE LSPs (i.e TE LSP whose destination IP address belongs to another area of the head-end LSR) but within the same AS.

A (multi-domain scope). When set, the PCS can perform path computation for inter-domain TE LSP. In this case, the PCSD TLV must contain one or more AS-domain TLV(s) each describing the domain for which the PCS can compute TE LSPs paths having their destination address in this domain.

Note that a PCS may set one or more flags.

See section 5 for a detailed description of the elements of procedure.

4.1.2 Path Computation Server address TLV

The PCS address TLV specifies the IP address to be used to reach the PCS described by this PCSD TLV. This address will typically be a loop-back address, always reachable, provided the router is not isolated. The Path Computation Server Address TLV is mandatory.

The PCS address TLV type is 2, length is 8 octets for an IPv4 address and 20 octets for an IPv6 address, and the value is the PCS IPv4 or IPv6 address.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              2                |      variable (8 or 20)       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     address-type              |          Reserved             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
//                      PCS IP address                         //
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Path Computation Server address TLV format

Address-type:
- 1 IPv4
- 2 IPv6

The Path Computation Server address TLV MUST appear exactly once in the PCSD TLV originated by a router. The only exception is when the PCS has both an IPv4 and IPv6 address; in this case, two path computation server address TLVs might be inserted: one for the IPv4 address, one for the IPv6 address.

4.1.3 Path Computation Server capability TLV

The PCS capability TLV is used by the PCS to signal its path computation server capabilities. This TLV is optional.

The PCS capability TLV type is 3 and the length is 8 octets.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              3                |            8                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                           Reserved                            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|P|M|D|E|G|                Reserved                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

PCS capability TLV format

P bit
The notion of request priority allows a PCC to specify how urgent is the request setting a flag in the REQUEST_ID object of the Path computation request message. See [4] for more details.

P=1: the PCS takes into account the ''request priority'' in its scheduling of the various requests.
P=0: the PCS does not take the request priority into account

M bit
M=1: the PCS is capable of computing more than one paths obeying a set of specified constraints, provided that they exist.
M=0: the PCS cannot compute more than one path obeying a set of specified constraints.

D bit
The PCC may request the PCS to compute N diversely routed paths obeying a set of specified constraints.
Such N paths may not exist of course depending on the current state of the network. See [4] for more details.
D=1: the PCS is capable of computing diversely (link, node, SRLG) routed paths.
D=0: the PCS is not capable of computing diversely routed paths.
The D bit is relevant if and only if the M bit has been set to 1. It must be set to 0 if the M bit is set to 0.

E bit
The PCC may request the PCS the computation of a path obeying a set of constraints one of those constraints being that one or more specified network element object must not be traversed by the LSP (a network element may be a link, an LSR or an Autonomous System). See [4] for more details.
E=1: the PCS is capable of computing TE LSP paths excluding some network elements.
E=0: the PCS is not capable of computing paths excluding network elements.

G bit
As defined in [4], the PCC may send a request specifying the metric to be used by the PCS when computing the shortest path during the CSPF.
G=1: the PCS supports the computation of CSPF with various metrics
G=0: the PCS just computes the CSPF based on the TE metric

Note that for future capability, it may be desirable to introduce new flags or may be new TLV to be carried in the PCSD capability TLV if the capability needs more than just a single flag to be described.

4.1.4 AS-domain TLV

When the PCS can perform path computation for inter-domain TE LSP, the A bit of the path computation server scope TLV must be set. Moreover, one or more TLVs MUST included within the PCSD TLV, each TLV identifying an AS number. Each TLV will have the following form:


```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              4                |            4                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                           AS Number                           |
```

```
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

AS-domain TLV format

The AS-domain TLV type is 4, length is 4 octets, and the value is the AS number identifying the AS for which the PCS can compute inter-domain TE LSP paths (TE LSP having their destination address in this domain). When coded on two bytes (which is the current defined format as the time of writing), the AS Number field must have its left two bytes set to 0.

The set of AS-domain TLVs specifies a list of ASes (AS1, ... , ASn). This means that the PCS can compute TE LSP path such that the destination address of the TE LSP belong to this set of ASes.

4.2 Mesh-group TLV format

The mesh-group TLV has the following format:

```
 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
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |              3                |Length: Variable (N*8 octets)  |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                        mesh-group-number                      |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                        Tail-end address                       |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                                                               //
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Mesh-group TLV format

N is the number of mesh-groups.

For each Mesh-group announced by the LSR, the TLV contains:
- a mesh-group-number: identifies the mesh-group number,
- a Tail-end address: IP address to be used as a tail-end address by other LSR belonging to the same mesh-group.

5. Elements of procedure

The PCSD and Mesh-group TLVs are carried within an OSPF router information opaque LSA (opaque type of 4, opaque ID of 0) as defined in [18].

A router must originate a new OSPF router information LSA whenever the content of the PCSD or Mesh-group TLV changes or whenever required by the regular OSPF procedure (LSA refresh (every LSRefreshTime, ...)).

As defined in RFC2370, an opaque LSA has a flooding scope determined by its LSA type:

- link-local (type 9),
- area-local (type 10)
- entire OSPF routing domain (type 11). In this case, the flooding scope is equivalent to the Type 5 LSA flooding scope.

A router may generate multiple OSPF router information LSA with different flooding scope.

The PCSD TLV may be carried within a type 10 or 11 router information LSA depending on the path computation server scope.

- If the PCS can compute intra-area TE LSP, the L bit of the path computation server scope sub-TLV of the PCSD TLV must be set and the PCSD TLV must be generated within a Type 10 router information LSA,

- If the PCS can compute inter-area TE LSP, the I bit of the path computation server scope sub-TLV of the PCSD TLV must be set. The PCSD TLV must be generated:
    - within a Type 10 router information LSA if the PCS can compute inter-area TE LSP path for the LSRs in the area it is attached to (for instance the PCS is an ABR computing inter-area TE LSP path for its attached areas)
    - within a Type 11 router information LSA is the PCS can compute inter-area TE LSP path for the whole domain.

- If the PCS can compute inter-AS TE LSP, the A bit of the path computation server scope sub-TLV of the PCSD TLV must be set and the PCSD TLV must be generated within a Type 11 router information LSA,

The Mesh-group TLV may be carried within a type 10 or 11 router information LSA depending on the MPLS TE mesh-group profile:

- If the MPLS TE mesh-group is contained with a single area (all the LSR have their Head-End and Tail-End within the same area), the Mesh-group TLV must be generated within a Type 10 router information LSA,
- If the MPLS TE mesh-group spans multiple OSPF areas, the Mesh-group TLV must be generated within a Type 11 router information LSA,

Note: if the PCS can both compute intra and inter-area TE LSP paths, both the L and I bits of the path computation server scope TLV must be set. The flags are not exclusive. This only applies to the PCSD TLV carried within the type 10 router information LSA.

If a PCS can compute intra-area TE LSP and inter-area or inter-AS TE LSP path, it must originate:

- a type 10 OSPF router information LSA with a PCSD TLV having L=1 and the I and A flags of its Path Computation Server scope TLV set as described above ,
- a type 11 OSPF router information LSA with a PCSD TLV having L=0 and the I and A flags of its Path Computation Server scope TLV set as described above,

Example

```
<-----------------AS1----------------->

<---area 1--><----area 0-----><-area 2->

R1---------ABR1*------------ABR3*-----|                 ------------
 |           |                |       |                 |          |
 |     S1    |      S2        |     ASBR1*--eBGP--ASBR2-|    AS2   |
 |           |                |       |                 |          |
R2---------ABR2*------------ABR4------|                 ------------
```

The areas contents are not detailed.

Assumptions:

- area 1 and area 2 are regular areas
- the * indicates a Path computation server capability
- ABR1 is a PCS for area 1 only
- ABR2 is a PCS for intra and inter-area TE LSP path computation in area 0 and 1
- ABR3 is a PCS for only inter-area TE LSP path computation for the whole domain,
- S1 is a PCS for area 1 only
- S2 is a PCS for the whole domain,
- ASBR1 is a PCS for inter-AS TE LSP only whose destination resides in AS2 (not for intra or inter-area area TE LSPs).

In the example above:

- S1 originates a type 10 router information LSA with a PCSD TLV such that:
    - o The L bit of the path computation server scope TLV is set,
    - o The I and A bits of the path computation server scope TLV are cleared.

- ABR1 originates in area 1 a type 10 router information LSA with a PCSD TLV such that:

- o The L bit of the path computation server scope TLV is set,
    - o The I and A bits of the path computation server scope TLV are cleared,

- ABR2 originates in both area 0 and 1 a type 10 router information LSA with a PCSD TLV such that:
    - o The L and I bits of the path computation server scope TLV are set,
    - o The A bit of the path computation server scope TLV is cleared

- ABR3 originates a type 11 router information LSA with a PCSD TLV such that:
    - o The L bit of the path computation server scope TLV is cleared,
    - o The I bit of the path computation server scope TLV is set,
    - o The A bit of the path computation server scope TLV is

cleared,

- S2 originates:
    - in area 0 a type 10 router information LSA with a PCSD TLV such that:
        - o The L and I bits of the path computation server scope TLV are set,
        - o The A bit of the path computation server scope TLV is cleared,
    - a type 11 router information LSA with a PCSD TLV such that:
        - o The L bit of the path computation server scope TLV is cleared,
        - o The I bit of the path computation server scope TLV is set,
        - o The A bit of the path computation server scope TLV is cleared,

- ASBR1 originates a type 11 router information LSA with a PCSD TLV such that:
    - o The L bit and the I bit of the path computation server scope TLV are cleared,
    - o The A bit of the path computation server scope TLV set,
    - o One AS-domain TLV within the PCSD TLV with AS number = AS2

The receipt of a new router information LSA carrying a PCSD TLV never triggers an SPF calculation.

When an LSR or a dedicated path computation server is newly configured as a PCS, the corresponding router information LSA must be immediately flooded.

When a PCS capability changes, the corresponding router information LSA must be immediately flooded.

When an LSR or a dedicated path computation server looses its path computation server capability, the corresponding router information LSA must be immediately flooded with LS age = MaxAge.

5. Interoperability with routers non supporting this capability

There is no interoperability issue as a router non-supporting the PCSD and Mesh-Group TLVs should just silently discard those TLVs as specified in RFC2370.

6. Security Considerations

No new security issues are raised in this document.

7. Acknowledgments

The authors would like to thank Abhay Roy, Dan Tappan, Robert Raszuk and Vishwas Manral for their comments.

8. References

[1] Moy, J., "OSPF Version 2", RFC 2328, April 1998.

[2] Awduche, D., et al, "Requirements for Traffic Engineering Over MPLS," RFC 2702, September 1999.

[3] Coltun, R., "The OSPF Opaque LSA Option," RFC 2370, July 1998.

[4] Vasseur et al, "RSVP Path computation request and reply messages ", draft-vasseur-mpls-computation-rsvp-te-03.txt, work in progress.

[5] Katz, D., Yeung, D., "Traffic Engineering Extensions to OSPF", draft-katz-yeung-ospf-traffic-04.txt

[6] Smit, H. and T. Li, "ISIS Extensions for Traffic Engineering," draft-ietf-isis-traffic-03.txt, work in progress.

[7] Kompella, K., and Rekhter, Y., "LSP Hierarchy with MPLS TE", Internet Draft, draft-ietf-mpls-lsp-hierarchy-02.txt, Feb., 2001.

[8] Kompella, K., Rekhter, Y., "Signalling Unnumbered Links in RSVP-TE", Internet Draft, draft-ietf-mpls-rsvp-unnum-01.txt, February 2001

[9] Kompella, K., Rekhter, Y., Banerjee, A. et al, "OSPF Extensions in Support of Generalized MPLS", draft-ietf-ccamp-ospf-gmpls-extensions-06.txt (work in progress)

[10] Ashwood-Smith, P. et al, "Generalized MPLS Signaling - CR-LDP Extensions", Internet Draft, draft-ietf-mpls-generalized-cr-ldp-01.txt, February 2001.

[11] Ashwood-Smith, P. et al, "Generalized MPLS - Signaling Functional Description", Internet Draft, draft-ietf-mpls-generalized-signaling-02.txt,
February 2001.

[12] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels," RFC 2119.

[13] Braden, R. Ed. et al, "Resource ReserVation Protocol-- Version 1 Functional Specification", RFC 2205, September 1997.

[14] Awduche, et al, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.

[15] Berger L., Gan D., Swallow G., Pan P., Tommasi F., Molendini S., "RSVP Refresh Overhead Reduction Extensions", RFC 2961, April 2001.

[16] Le faucheur, "Use of IGP Metric as a second TE Metric", Internet draft, draft-lefaucheur-te-metric-igp-00.txt.

[17] Fedyk D., Ghanwani A., Ash J., Vedrenne A. "Multiple Metrics for Traffic Engineering with IS-IS and OSPF", Internet draft, draft-fedyk-isis-ospf-te-metrics-01.txt

[18] Aggarwal et all, ''Extensions to IS-IS and OSPF for Advertising Optional Router Capabilities'', Internet draft, draft-raggarwa-igp-cap-01.txt, October 2002.

9. Author's Addresses

JP Vasseur
CISCO Systems, Inc.
300 Apollo Drive
Chelmsford, 01824
Email: jpv@cisco.com

Peter Psenak
CISCO Systems, Inc.

Pegasus Parc
De Kleetlaan 6A
1831, Diegem
BELGIUM
Email: ppsenak@cisco.com