

Networking Working Group
Internet-Draft
Intended status: Informational
Expires: June 17, 2007

JP. Vasseur, Ed.
Cisco Systems, Inc
R. Zhang
BT Infonet
N. Bitar
Verizon
JL. Le Roux
France Telecom
December 14, 2006

**A Backward Recursive PCE-based Computation (BRPC) procedure to compute
shortest inter-domain Traffic Engineering Label Switched Paths
draft-ietf-pce-brpc-02.txt**

Status of this Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with [Section 6 of BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at
<http://www.ietf.org/ietf/1id-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at
<http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on June 17, 2007.

Copyright Notice

Copyright (C) The Internet Society (2006).

Abstract

The ability to compute constrained shortest Traffic Engineering (TE) Label Switched Paths (LSPs) in Multiprotocol Label Switching (MPLS)

and Generalized MPLS (GMPLS) networks across multiple domains (where a domain is referred to as a collection of network elements within a common sphere of address management or path computational responsibility such as IGP areas and Autonomous Systems) has been identified as a key requirement . This document specifies a procedure relying on the use of multiple Path Computation Elements (PCEs) in order to compute such inter-domain shortest constraint paths along a determined sequence of domains, using a backward recursive path computation technique while preserving confidentiality across domains, which is sometimes required when domains are managed by different Service Providers.

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 [RFC 2119](#) [[RFC2119](#)].

Table of Contents

1.	Terminology	4
2.	Introduction	4
3.	General assumptions	5
4.	BRPC Procedure	6
4.1.	Domain path selection	6
4.2.	Mode of Operation	6
5.	PCEP Protocol Extensions	8
6.	Inter-AS TE Links	9
7.	Usage in conjunction with per-domain path computation	9
8.	BRPC procedure completion failure	9
9.	Applicability	10
9.1.	Diverse end-to-end path computation	10
9.2.	Path optimality	10
10.	Reoptimization of an inter-domain TE LSP	11
11.	Path Computation failure	11
12.	Metric normalization	11
13.	Manageability Considerations	12
14.	IANA Considerations	12
14.1.	New flag of the RP object	12
14.2.	New flag of the NO-PATH-VECTOR TLV	12
15.	Security Considerations	12
16.	Acknowledgements	12
17.	References	13
17.1.	Normative References	13
17.2.	Informative References	13
Appendix A.	Proposed Status and Discussion [To Be Removed Upon Publication]	14
	Authors' Addresses	14
	Intellectual Property and Copyright Statements	16

1. Terminology

ABR: routers used to connect two IGP areas (areas in OSPF or levels in IS-IS).

ASBR: routers used to connect together ASs of a different or the same Service Provider via one or more Inter-AS links.

Boundary Node (BN): a boundary node is either an ABR in the context of inter- area TE or an ASBR in the context of inter-AS TE.

Entry BN of domain(n): a BN connecting domain(n-1) to domain(n).

Exit BN of domain(n): a BN connecting domain(n) to domain(n+1).

Inter-AS TE LSP: A TE LSP that crosses an AS boundary.

Inter-area TE LSP: A TE LSP that crosses an IGP area boundary.

LSR: Label Switching Router.

LSP: Label Switched Path.

PCE (Path Computation Element): an entity (component, application or network node) that is capable of computing a network path or route based on a network graph and applying computational constraints.

PCE(i) is a PCE with the scope of domain(i).

TED: Traffic Engineering Database.

VSPT: Virtual Shortest Path Tree.

The notion of contiguous, stitched and nested TE LSPs is defined in [[I-D.ietf-ccamp-inter-domain-framework](#)] and will not be repeated here.

2. Introduction

The requirements for inter-area and inter-AS MPLS Traffic Engineering have been developed by the Traffic Engineering Working Group (TE WG) and have been stated in [[RFC4105](#)] and [[RFC4216](#)], respectively.

The framework for inter-domain MPLS Traffic Engineering has been provided in [[I-D.ietf-ccamp-inter-domain-framework](#)].

[[I-D.ietf-ccamp-inter-domain-pd-path-comp](#)] defines a technique for

establishing inter-domain (G)MPLS TE LSP whereby the path is computed during the signalling process on a per-domain basis by the entry boundary node of each domain (each node in charge of computing a section of an inter-domain TE LSP path is always along the path of such TE LSP). Such path computation technique fulfills some of the requirements stated in [\[RFC4105\]](#) and [\[RFC4216\]](#) but not all of them. In particular, it cannot guarantee to find an optimal (shortest) inter-domain constrained path. Furthermore, it cannot be efficiently used to compute a set of inter-domain diversely routed TE LSPs.

The aim of this document is to describe a PCE-based TE LSP computation procedure to compute optimal inter-domain constrained (G)MPLS TE LSPs.

Qualifying a path as optimal requires some clarification. Indeed, a globally optimal TE LSP placement usually refers to a set of TE LSPs whose placements optimize the network resources with regards to a specified objective function (e.g. a placement that reduces the maximum or average network load while satisfying the TE LSP constraints). In this document, an optimal inter-domain constrained TE LSP is defined as the shortest path satisfying the set of required constraints that would be obtained in the absence of multiple domains (in other words, in a totally flat IGP network between the source and destination of the TE LSP).

3. General assumptions

In the rest of this document, we make the following set of assumptions common to inter-area and inter-AS MPLS TE:

- Each IGP area or AS is assumed to be Traffic Engineering enabled (i.e. running OSPF-TE or ISIS-TE and RSVP-TE).
- No topology or resource information is distributed between domains (as mandated per [\[RFC4105\]](#) and [\[RFC4216\]](#)), which is critical to preserve IGP/BGP scalability and confidentiality.
- While certain constraints like bandwidth can be used across different domains, other TE constraints like resource affinity, color, metric, etc. as listed in [\[RFC2702\]](#) could be translated at domain boundaries. If required, it is assumed that, at the domain boundary nodes, there will exist some sort of local mapping based on policy agreement, in order to translate such constraints across domain boundaries during the inter-PCE communication process.
- The various ASBRs are BGP peers, without any IGP running on the inter-ASBR links.

- Each AS can be made of several IGP areas. The path computation procedure described in this document applies to the case of a single AS made of multiple IGP areas, multiples ASs made of a single IGP area or any combination of the above. For the sake of simplicity, each AS will be considered to be comprised of a single area in this document. The case of an Inter-AS TE LSP spanning multiple ASs where some of those ASs are themselves made of multiple IGP areas can be easily derived from this case by applying the BRPC procedure described in this document, recursively.
- The domain path (set of domains traversed to reach the destination domain) is either administratively pre-determined or discovered by some means (outside of the scope of this document).

4. BRPC Procedure

The BRPC procedure is a Multiple-PCE path computation technique as described in [[RFC4655](#)]. A possible model consists of hosting the PCE function on boundary nodes (e.g., ABR or ASBR) but this is not mandated by the BRPC procedure.

The BRPC procedure does not make any assumptions with regards to the nature of the inter-domain TE LSP that could be contiguous, nested or stitched.

Furthermore, no assumption is made on the actual path computation algorithm in use by a PCE (it can be any variant of CSPF, algorithm based on linear-programming to solve multi-constraints optimization problems and so on).

4.1. Domain path selection

The PCE-based BRPC procedure applies to the computation of an optimal constrained inter-domain TE LSP. The sequence of domains to be traversed can either be determined a priori or during the path computation procedure. The BRPC procedure guarantees to compute the optimal path across a specific set of traversed domains (which constitutes an additional constraint). In the case of an arbitrary set of meshed domains, the BRPC procedure can be used to compute the optimal path across each domain set in order to get to optimal constrained path between the source and the destination of the TE LSP.

4.2. Mode of Operation

Definition of VSPT(i)

In each domain i :

* There is a set of $X\text{-en}(i)$ entry BNs noted $BN\text{-en}(k,i)$ where $BN\text{-en}(k,i)$ is the k th entry BN of domain(i).

* There is a set of $X\text{-ex}(i)$ exit BN noted $BN\text{-ex}(k,i)$ where $BN\text{-ex}(k,i)$ is the k th exit BN of domain(i).

$VSPT(i)$: MP2P (MultiPoint To Point) tree returned by $PCE(i)$ to $PCE(i-1)$:

```
          Root (TE LSP destination)
          /      I      \
    BN-en(1,i)  BN-en(2,i) ... BN-en((j), i).
```

Where $j \leq [X\text{-en}(i)]$

Each link of tree $VSPT(i)$ represents the shortest path between $BN\text{-en}(j,i)$ (identified by its TE Router-ID) and the destination that satisfies the set of required constraints for the TE LSP (bandwidth, affinities, ...). These are path segments to reach the destination from $BN\text{-en}(j,i)$.

Note that $PCE(i)$ only considers the entry BNs that provide connectivity from domain($i-1$). That is, the set $BN\text{-en}(k,i-1)$ is only made of those BNs that provide connectivity from domain ($i-1$) to domain(i). Furthermore, some BNs may be excluded according to policy constraints (either due to local policy or policies signaled in the path computation request).

Step 1: the PCC needs to first determine the PCE capable of serving its path computation request. The path computation request is then relayed until reaching a $PCE(n)$ such that the TE LSP destination resides in the domain(n). At each step of the process, the next PCE can either be statically configured or dynamically discovered via IGP/BGP extensions. If no next PCE can be found or the next hop PCE of choice is unavailable, the procedure stops and a path computation error is returned (see section [Section 8](#)). If multiple PCEs are discovered, the PCE may select a subset of these PCEs based on some local policies or heuristics. Note also that a sequence of PCEs might be enforced by policy on the PCC and this constraint can be either carried in the PCEP path computation request (defined in [\[I-D.ietf-pce-pcep\]](#)).

Step 2: $PCE(n)$ computes $VSPT(n)$ made of the list of shortest

constrained path(s) between every BN-en(j,n) and the TE LSP destination using a suitable path computation algorithm (e.g. CSPF) and returns the computed VSPT(n) to PCE(n-1).

Step i:

- For i=n-1 to 2: PCE(i) concatenates the topology of domain (i) (using its TED) with the received VSPT(i+1). In the case of Inter-AS TE LSP computation, this requires to also add the inter-AS TE links connecting the domain (i) to the domain (i+1). Then PCE(i) computes VSPT(i) (P2MP tree made of the shortest constrained paths between each BN-en(j,i) and the TE LSP destination).

End

Finally PCE(1) computes the end-to-end shortest constrained path from the source to the destination and returns the corresponding path to the requesting PCC.

Each branch of the VSPT tree (path) may be returned in the form of an explicit path (in which case all the hops along the path segment are listed) or a loose path (in which case only the BR is specified) so as to preserve confidentiality along with the respective cost. In the later case, various techniques can be used in order to retrieve the computed explicit paths on a per domain basis during the signaling process thanks to the use of path keys as described in [\[I-D.bradford-pce-path-key\]](#).

BRPC guarantees to find the optimal (shortest) constrained inter-domain TE LSP according to a set of defined domains to be traversed. Note that other variants of the BRPC procedure relying on the same principles are also possible.

Note also that in case of ECMP paths, more than one path could be returned to the requesting LSR.

5. PCEP Protocol Extensions

The BRPC procedure requires the specification of a new flag of the RP object carried within the PCReq message (defined in [\[I-D.ietf-pce-pcep\]](#)), the aim of which is to specify that the shortest path(s) satisfying the constraints from the destination to the set of entry boundary nodes are requested (such set of path(s) forms the downstream VSPT as specified in [Section 4.2](#)).

The following new flag of the RP object is defined: VSPT (V) flag: 0x20. When set, this indicates that the PCC requests the computation

of an inter-domain TE LSP using the BRPC procedure.

Because path segment(s) computed by a downstream PCE in the context of the BRPC procedure must be provided along with their respective path cost(s), the C flag of the RP object carried within the PCReq message MUST be set. It is the choice of the requester to appropriately set the O bit of the RP object.

6. Inter-AS TE Links

In the case of Inter-AS TE LSP path computation, the BRPC procedure requires the knowledge of the traffic engineering attributes of the Inter-AS TE links: the process by which the PCE acquires this information is out of the scope of the BRPC procedure (which is compliant with the PCE architecture defined in [[RFC4655](#)]).

That said, a straightforward solution consists of allowing the ASBRs to flood the TE information related to the inter-ASBR link(s) although no IGP TE is enabled over those links (there is no IGP adjacency over the inter-ASBR links). This allows the PCE of a domain to get entire TE visibility up to the set of entry ASBRs in the downstream domain.

7. Usage in conjunction with per-domain path computation

The BRPC procedure may be used to compute path segments and could be used in conjunction with other path computation techniques (such as the per-domain path computation technique defined in [[I-D.ietf-ccamp-inter-domain-pd-path-comp](#)]) to compute the end-to-end path. In this case end-to-end path optimality can no longer be guaranteed.

8. BRPC procedure completion failure

If the BRPC procedure cannot be completed because a PCE along the domain path does not support the procedure, a PCErr message is returned to the upstream PCE with a Error-Type "BRPC procedure completion failure". The PCErr message MUST be relayed to the requesting PCC.

PCEP-ERROR objects are used to report a PCEP protocol error and are characterized by an Error-Type which specifies the type of error and an Error-value that provides additional information about the error type. Both the Error-Type and the Error-Value are managed by IANA. A new Error-Type is defined that relates to the BRPC procedure.

Error-type	Meaning
10	BRPC procedure completion failure
	Error-value
	1: BRPC procedure not supported by one or more PCEs along the domain path

9. Applicability

As discussed in [section 3](#), the requirements for inter-area and inter-AS MPLS Traffic Engineering have been developed by the Traffic Engineering Working Group (TE WG) and have been stated in [\[RFC4105\]](#) and [\[RFC4216\]](#), respectively. Among the set of requirements, both documents indicate the need for some solution providing the ability to compute an optimal (shortest) constrained inter-domain TE LSP and to compute a set of diverse inter-domain TE LSPs.

9.1. Diverse end-to-end path computation

PCEP allows a PCC to request the computation of a set of diverse TE LSPs thanks to the SVEC object by setting the flags L, N or S to request link, node or SRLG diversity respectively. Such request MUST be taken into account by each PCE along the path computation chain during the VSPT computation. In the context of the BRPC procedure, a set of diversely routed TE LSP between two LSRs can be computed since the paths segment(s) of the VSPT are simultaneously computed by a given PCE. The BRPC path procedure allows for the computation of diverse paths under various objective functions (such as minimizing the sum of the costs of the N diverse paths, etc) thus avoiding the well-known "trapping" problem. Indeed, with a 2-step approach consisting of computing the first path followed by the computation of the second path after having removed the set of network elements traversed by the first path (if that does not violate confidentiality preservation), one cannot guarantee that a solution will be found even if such solution exists. Furthermore, even if a solution is found, it may not be the most optimal one with respect to objective function such as minimizing the sum of the paths costs, bounding the path delays of both paths and so on. Finally, it must be noted that such a 2-step path computation approach is usually less efficient in term of signalling delays since it requires two serialized TE LSP set up.

9.2. Path optimality

BRPC guarantees that the optimal (shortest) constrained inter-domain path will always be found subject to policy constraints. When combined with other local path computation techniques (e.g. in the case of stitched/nested TE LSP) and in the case where a domain has

more than one BR-en or more than one BR-ex, optimality after some network change within the domain can only be guaranteed by re-executing the BRPC procedure.

10. Reoptimization of an inter-domain TE LSP

The ability to reoptimize an existing inter-domain TE LSP path has been explicitly listed as a requirement in [[RFC4105](#)] and [[RFC4216](#)]. In the case of a TE LSP reoptimization request, regular procedures apply as defined in PCEP where the path in use (if available on the head-end) is provided within the path computation request in order for the PCEs involved in the reoptimization request to avoid double bandwidth accounting.

11. Path Computation failure

If a PCE requires to relay a path computation request according to the BRPC procedure defined in this document to a downstream PCE that no such PCE is available, the PCE MUST send a negative path computation reply to the requester using a PCReq message as specified in [[I-D.ietf-pce-pcep](#)] that contains a NO-PATH object. In such case, the NO-PATH object MUST carry a NO-PATH-VECTOR TLV (also defined in [[I-D.ietf-pce-pcep](#)]) with the newly defined bit named "BRPC Path Computation chain unavailable" set.

0x04: BRPC Path computation chain unavailable

12. Metric normalization

In the case of inter-area TE, the same IGP/TE metric scheme is usually adopted for all the IGP areas (e.g. based on the link-speed, propagation delay or some other combination of link attributes). Hence, the proposed set of mechanisms always computes the shortest path across multiple areas obeying the required set of constraints with respect to a well-specified objective function. Conversely, in the case of Inter-AS TE, in order for this path computation to be meaningful, a metric normalization between ASs may be required. One solution to avoid IGP metric modification would be for the SPs to agree on a TE metric normalization scheme and use the TE metric for TE LSP path computation (in that case, this must be requested in the PCEP Path computation request) thanks to the COST object.

13. Manageability Considerations

To be added in a further revision of this document.

14. IANA Considerations

14.1. New flag of the RP object

A new flag of the RP object (specified in [[I-D.ietf-pce-pcep](#)]) is defined in this document.

Name: VSPT (V)

Value: 0x20.

When set, this indicates that the PCC requests the computation of an inter-domain TE LSP using the BRPC procedure.

A new Error-Type is defined in this document (Error-Type and Error-value to be assigned by IANA).

Error-type	Meaning
10	BRPC procedure completion failure
	Error-value
	1: BRPC procedure not supported by one or PCEs along the domain path

14.2. New flag of the NO-PATH-VECTOR TLV

A new flag of the NO-PATH-VECTOR TLV defined in [[I-D.ietf-pce-pcep](#)]) is specified in this document.

Defining RFC: [draft-ietf-pce-pcep](#) (to be replaced by RFC number when published)

Name of bit: BRPC Path computation chain unavailable

Bit number (suggested value): 0x04

15. Security Considerations

The BRPC procedure does not introduce any additional security issues beyond the ones related to inter-PCE communication.

16. Acknowledgements

The authors would like to thank Arthi Ayyangar, Dimitri Papadimitriou

and Siva Sivabalan for their useful comments. A special thank to Adrian Farrel for his useful comments and suggestions.

17. References

17.1. Normative References

- [I-D.ietf-pce-pcep]
Roux, J. and J. Vasseur, "Path Computation Element (PCE) communication Protocol (PCEP) - Version 1",
[draft-ietf-pce-pcep-04](#) (work in progress), December 2006.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", [RFC 4655](#), August 2006.

17.2. Informative References

- [I-D.bradford-pce-path-key]
Bradford, R., "Preserving Topology Confidentiality in Inter-Domain Path Computation using a key based mechanism", [draft-bradford-pce-path-key-01](#) (work in progress), October 2006.
- [I-D.ietf-ccamp-inter-domain-framework]
Farrel, A., "A Framework for Inter-Domain Multiprotocol Label Switching Traffic Engineering",
[draft-ietf-ccamp-inter-domain-framework-06](#) (work in progress), August 2006.
- [I-D.ietf-ccamp-inter-domain-pd-path-comp]
Vasseur, J., "A Per-domain path computation method for establishing Inter-domain Traffic Engineering (TE) Label Switched Paths (LSPs)",
[draft-ietf-ccamp-inter-domain-pd-path-comp-03](#) (work in progress), August 2006.
- [I-D.ietf-ccamp-inter-domain-rsvp-te]
Ayyangar, A. and J. Vasseur, "Inter domain GMPLS Traffic Engineering - RSVP-TE extensions",
[draft-ietf-ccamp-inter-domain-rsvp-te-03](#) (work in progress), March 2006.
- [RFC2702] Awduche, D., Malcolm, J., Agogbua, J., O'Dell, M., and J. McManus, "Requirements for Traffic Engineering Over MPLS",

[RFC 2702](#), September 1999.

[RFC4105] Le Roux, J., Vasseur, J., and J. Boyle, "Requirements for Inter-Area MPLS Traffic Engineering", [RFC 4105](#), June 2005.

[RFC4216] Zhang, R. and J. Vasseur, "MPLS Inter-Autonomous System (AS) Traffic Engineering (TE) Requirements", [RFC 4216](#), November 2005.

[Appendix A](#). Proposed Status and Discussion [To Be Removed Upon Publication]

This Internet-Draft is being submitted for eventual publication as an RFC with a proposed status of Informational. Discussion of this proposal should take place on the following mailing list:
pce@ietf.org.

Authors' Addresses

JP Vasseur (editor)
Cisco Systems, Inc
1414 Massachusetts Avenue
Boxborough, MA 01719
USA

Email: jpv@cisco.com

Raymond Zhang
BT Infonet
2160 E. Grand Ave.
El Segundo, CA 90025
USA

Email: raymond_zhang@bt.infonet.com

Nabil Bitar
Verizon
40 Sylvan Road
Waltham, MA 02145
USA

Email: nabil.bitar@verizon.com

JL Le Roux
France Telecom
2, Avenue Pierre-Marzin
Lannion, 22307
FRANCE

Email: jeanlouis.leroux@orange-ft.com

Full Copyright Statement

Copyright (C) The Internet Society (2006).

This document is subject to the rights, licenses and restrictions contained in [BCP 78](#), and except as set forth therein, the authors retain all their rights.

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Intellectual Property

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in [BCP 78](#) and [BCP 79](#).

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at <http://www.ietf.org/ipr>.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Acknowledgment

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).

