Sukrit Dasgupta
Jaudelice C. de Oliveira
Drexel University
J.-P. Vassuer
Cisco Systems

Networking Working Group
Internet Draft
Intended Status: Informational
Expires: May 10, 2008

November 7, 2007

Performance Analysis of Inter-Domain Path Computation Methodologies
draft-dasgupta-ccamp-path-comp-analysis-01

Status of this Memo

Copyright Notice

Abstract

This document presents a performance comparison between the per- domain path computation method and the Path Computation Element (PCE) Architecture based Backward Recursive Path Computation (BRPC) procedure. Metrics to capture the significant performance aspects are identified and detailed simulations are carried out on realistic scenarios. A performance analysis for each of the path computation methods is then undertaken.

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].

Contents

1 Terminology

Terminology used in this document

TE LSP: Traffic Engineered Label Switched Path.

CSPF: Constraint Shortest Path First.

PCE: Path Computation Element.

BRPC: Backward Recursive PCE based Computation.

AS: Autonomous System.

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

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

Border LSR: A border LSR is either an ABR in the context of inter- area TE or an ASBR in the context of inter-AS TE.

VSPT: Virtual Shortest Path Tree.

LSA: Link State Advertisement.

LSR: Label Switching Router.

IGP: Interior Gateway Protocol.

TED: Traffic Engineering Database.

PD: Per-Domain.

2 Introduction

The IETF has specified two approaches for the computation of inter- domain (Generalized) Multi-Protocol Label Switching (MPLS) Traffic Engineering (TE) Label Switched Paths (LSP): the per-domain path computation approach defined in [I-D.ietf-ccamp-inter-domain-pd-path-comp] and the PCE based approach specified in [RFC4655]. More specifically we study the PCE based path computation model that makes use of the BRPC method outlined in[I-D.ietf-pce-brpc]. In the rest of this document, we will call PD and PCE the per-domain path computation approach and the PCE path computation approach respectively.

In the per-domain path computation approach, each path segment within a domain is computed during the signaling process by each entry node of the domain up to the next hop exit node of that same domain.

By contrast the PCE-based approach and in particular the BRPC method defined in [I-D.ietf-pce-brpc] relies the collaboration between a set of PCEs to find to shortest inter-domain path after the computation of which the corresponding TE LSP is signaled: path computation is undertaken using multiple PCEs in a backward recursive fashion from the destination domain to the source domain. The notion of a Virutal Shortest Path Tree (VSPT) is introduced. Each link of a VSPT represents the shortest path satisfying the set of required constraints between the border nodes of a domain and the destination LSR. The VSPT of each domain is returned by the corresponding PCE to create a new VSPT by PCEs present in other domains. [I-D.ietf-pce-brpc] discusses the BRPC procedure in complete detail.

This document presents some simulation results and analysis to compare the performance of the above two inter-domain path computation approaches. Two realistic topologies with accompanying traffic matrices are used to undertake the simulations.

Note that although the simulations results discussed in this document have used inter-area networks, they also apply to Inter-AS cases.

Disclaimer: although simulations have been made on different and realistic topologies showing consistent results, the metrics shown below may vary with the network topology.

3 Evaluation Metrics

This section discusses the metrics that are used to quantify and compare the performance of the two approaches.

o Path Cost. The maximum and average path costs are observed for each TE LSP. The distributions for the maximum and average path costs are then compared for the two path computation approaches.

o Signaling Failures. Signaling failures may occur in various circumstances. With PD, the head-end LSR chooses the the downstream border router (ABR, ASBR) according to some selection criteria (IGP shortest path, ...) based on the information in its TED. This ABR then selects the next ABR using its TED, continuing the process till the destination is reached. At each step, the TED information could be out of date, potentially resulting in a signaling failure during setup. In the BRPC procedure, the PCEs are the ABRs that cooperate to form the VSPT based on the information in their respective TEDs. As in the case of the PD approach, information in the TED could be out of date, potentially resulting in signaling failures during setup. Also, only with the PD approach, another situation that leads to a signaling failure is when the selected exit ABR does not have any path obeying the set of constraints toward a downstream exit node or the TE LSP destination. This situation does not occur with the BRPC. The signaling failure metric captures the total number of signaling failures that occur during initial setup and reroute (on link failure) of a TE LSP. The distribution of the number of signaling failures encountered for all TE LSPs is then compared for the PD and BRPC methods.

o Crankback Signaling. In this document we made the assumption that in the case of PD, when an entry border node fails to find a route in the corresponding domain, Boundary re-routing crankback [I-D.ietf-ccamp-crankback] signaling was used. A crankback signaling message propagates to the entry border node of the domain and a new exit border node is chosen. After this, path computation takes place to find a path segment to a new entry border node of the next domain. This causes a additional delay in setup time. This metric captures the distribution of the number of crankback signals and the corresponding delay in setup time for a TE LSP when using PD. The total delay arising from the crankback signaling is proportional to the costs of the links over which the signal travels, i.e., the path which is setup from the entry border node of a domain to its exit border node (the assumption was made that link metrics reflect propagation delays). Similar to above metrics, the distribution of total crankback signaling and corresponding proportional delay across all TE LSPs is compared.

o TE LSPs/Bandwidth Setup Capacity. Due to the different path computation techniques, there is a significant difference in the amount of TE LSPs/bandwidth that can be setup. This metric captures the difference in the number of TE LSPs and corresponding bandwidth that can be setup using the two path computation techniques. The traffic matrix is continuously scaled and stopped when the first TE LSP cannot be setup for both the methods. The difference in the scaling factor gives the extra bandwidth that can be setup using the corresponding path computation technique.

o Failed TE LSPs/Bandwidth on link failure. Link failures are induced in the network during the course of the simulations conducted. This metric captures the number of TE LSPs and the corresponding bandwidth that failed to find a route when one or more links lying on its path failed.

4 Simulation Setup

A very detailed simulator has been developed to replicate a real life network scenario accurately. Following is the set of entities used in the simulation with a brief description of their behavior.

o Topology Description. To obtain meaningful results applicable to present day Service Provider topologies, simulations have been run on two representative topologies. They consists of a large backbone area to which four smaller areas are connected. For the first topology named MESH-CORE, a densely connected backbone was obtained from RocketFuel [ROCKETFUEL]. The second topology has a symmetrical backbone and is called SYM-CORE. The four connected smaller areas are obtained from [DEF-DES]. Details of the topologies are shown in Table 1 along with their layout in Figure 1. All TE LSPs setup on this network have their source and destinations in different areas and all of them need to traverse the backbone network. Table 1 also shows the number of TE LSPs that have their sources in the corresponding areas along with their size distribution.

o Node behavior. Every node in the topology represents a router that maintains states for all the TE LSPs passing through it. Each node in a domain is a source for TE LSPs to all the other nodes in the other domains. As in a real life scenario, where routers boot up at random points in time, the nodes in the topologies also start sending traffic on the TE LSPs originating from them at a random start time (to take into account the different boot-up times). All nodes are up within an hour of the start of simulation. All nodes maintain a TED that is updated using LSA updates as outlined in [RFC3630]. The flooding scope of the Traffic Engineering IGP updates are restricted only to the domain in which they originate in compliance with [RFC3630] and [RFC3784].

o TE LSP Setup. When a node boots up, it sets up all TE LSPs that originate from it in descending order of size. The network is dimensioned such that all TE LSPs can find a path. Once setup, all TE LSPs stay in the network for the complete duration of the simulation unless they fail due to a link failure. Eventhough the TE LSPs are setup in descending order of size from a head-end router, from the network perspective, TE LSPs are setup in random fashion as the routers bootup at random times.

o Inducing Failures. For thorough performance analysis and comparison, link failures are induced in all the areas. Each link in a domain can fail independently with a mean failure time of 24 hours and be restored with a mean restore time of 15 minutes. Both inter-failure and inter-restore times are uniformly distributed. No attempt to re-optimize the path of a TE LSP is made when a link is restored. The links that join two domains never fail. This step has been taken to concentrate only on how link failures within domains affect the performance.

Domain Description					TE-LSP Size	
Domain Name	# of nodes	# of links	OC48 links	OC192 links	[0,20) Mbps	[20,100] Mbps
D_1	17	24	18	6	125	368
D_2	14	17	12	5	76	186
D_3	19	26	20	6	14	20
D_4	9	12	9	3	7	18
MESH Backbone	83	167	132	35	0	0
SYM Backbone	29	37	26	11	0	0

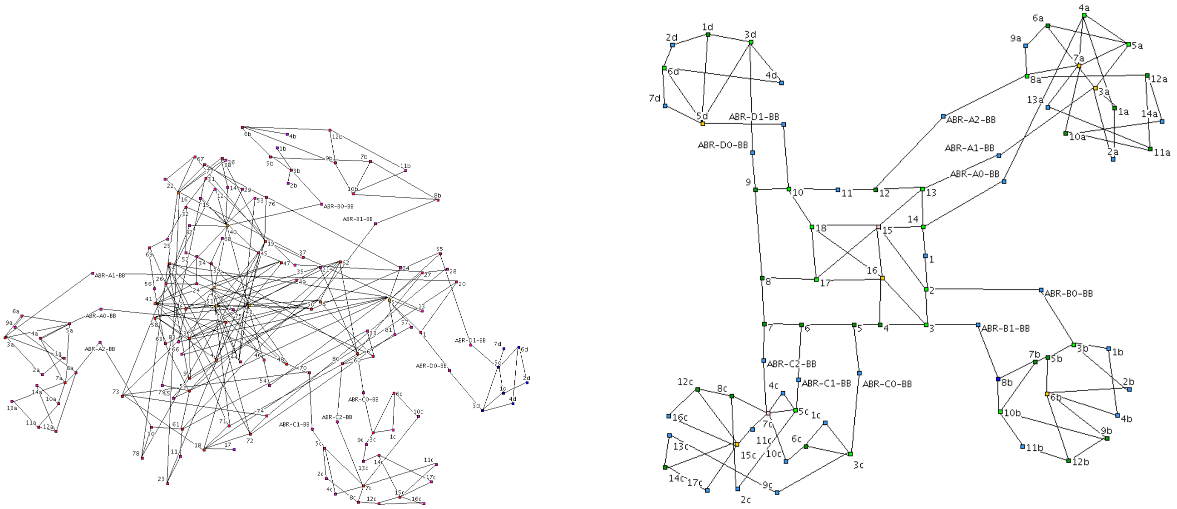
Table 1: Details of all the areas used to create the two topologies.

5 Results and Analysis

Simulations were carried out on the two topologies previously described. The results are presented and discussed in this section. All figures are from the PDF version of this document. In the figures, 'PD-Setup' and 'PCE-Setup' represent results corresponding to the initial setting up of TE LSPs on an empty network using the per-domain and the PCE approach, respectively. Similarly, 'PD- Failure' and 'PCE-Failure' denote the results under the link failure scenario. A period of one week was simulated and results were collected after the transient period. Figure 2 and Figure 3 illustrate the behavior of the metrics for topologies MESH-CORE and SYM-CORE, respectively.

5.1 Path Cost.

Figures 2a and 3a show the distribution of the average path cost of the TE LSPs for MESH-CORE and SYM-CORE, respectively. During initial setup, roughly 40% of TE LSPs for MESH-CORE and 70%



(a) Topology for MESH-CORE.

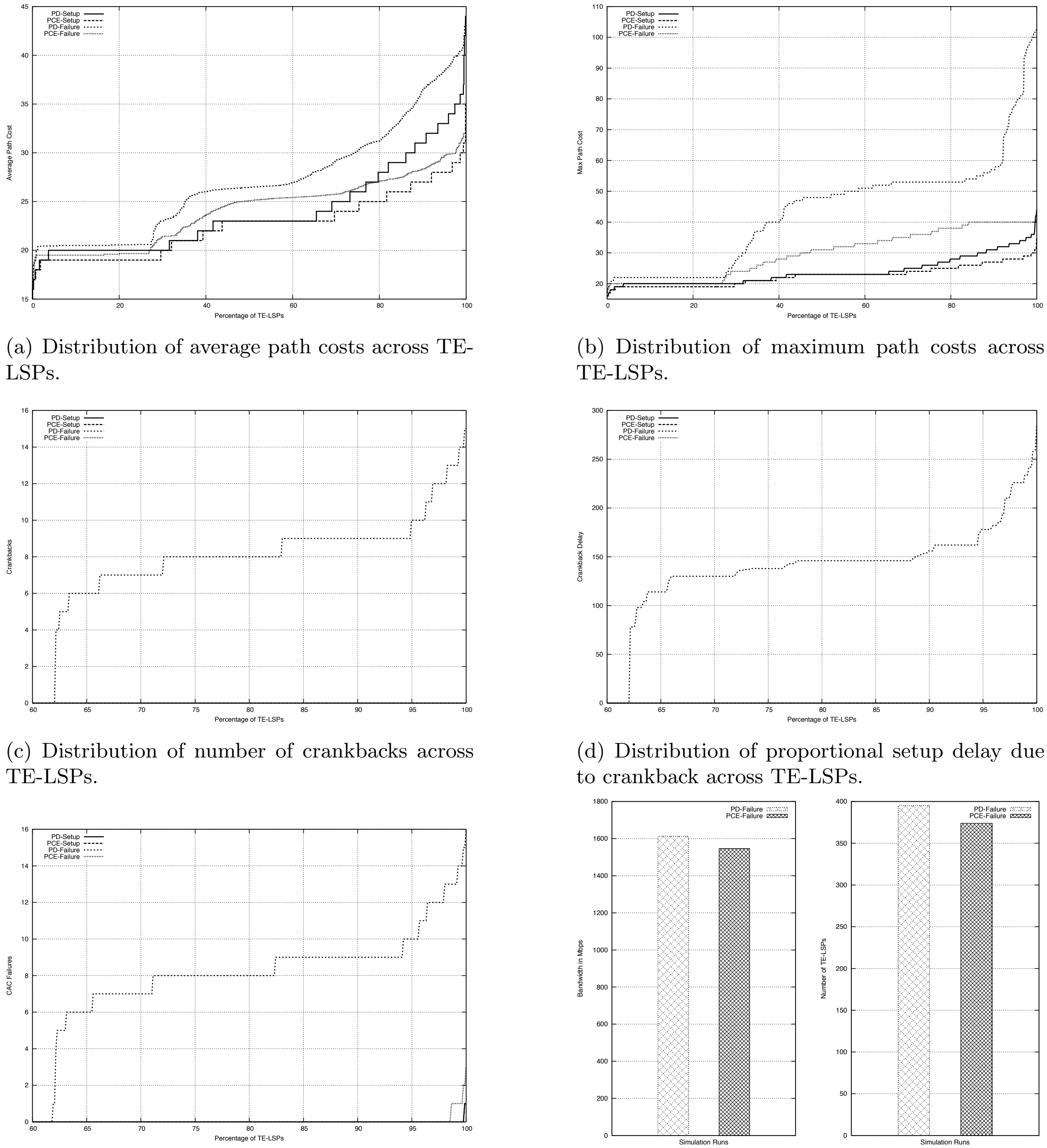
(b) Topology for SYM-CORE.

Figure 1: Topologies used for the simulations

of TE LSPs for SYM-CORE have path costs greater with PD (PD-Setup) than with PCE approach (PCE-Setup). This is due to the ability of the BRPC procedure to select the inter-domain shortest constrained paths that satisfy the constraints. Since the per-domain approach to path computation is undertaken in stages where every entry border router to a domain computes the path in the corresponding domain, the most optimal (shortest constrained inter-domain) route is not always found. When failures start to take place in the network, TE LSPs are rerouted over different paths resulting in path costs that are different from the initial costs. PD-Failure and PCE-Failure in Figures 2a and 3a show the distribution of the average path costs that the TE LSPs have over the duration of the simulation with link failures occurring. Similarly, the average path costs with the PD approach are much higher than the PCE approach when link failures occur. Figures 2b and 3b show similar trends and present the maximum path costs for a TE LSP for the two topologies, respectively. It can be seen that with per-domain path computation, the maximum path costs are larger for 30% and 100% of the TE LSPs for MESH-CORE and SYM- CORE, respectively.

5.2 Crankbacks/Setup Delay.

Due to crankbacks that take place in the per-domain approach of path computation, TE LSP setup time is significantly increased. This could lead to QoS requirements not being met, especially during failures when rerouting needs to be quick in order to keep traffic disruption to a minimum (for example in the absence of local repair mechanisms such as defined in [RFC4090]). Since crankbacks do not take place during path computation with a PCE, setup delays are significantly reduced. Figures 2c and 3c show the distributions of the number of crankbacks that took place during the setup of the corresponding TE LSPs for MESH-CORE and SYM-CORE, respectively. It can be seen that all crankbacks occurred when failures were taking place in the networks. Figures 2d and 3d illustrate the 'proportional' setup delays experienced by the TE LSPs due to crankbacks for the two topologies. It can be observed that for a large proportion of the TE LSPs, the setup delays arising out of crankbacks is very large possibly proving to be very detrimental to QoS requirements. The large delays arise out of the crankback signaling that needs to propagate back and forth from the exit border router of a domain to its entry border router. More crankbacks occur for SYM-CORE as compared to MESH-CORE as it is a very 'restricted' and 'constrained' network in terms of connectivity. This causes a lack of routes and often several cycles of crankback signaling are required to find a constrained path.



(a) Distribution of average path costs across TE-LSPs.

(b) Distribution of maximum path costs across TE-LSPs.

(c) Distribution of number of crankbacks across TE-LSPs.

(d) Distribution of proportional setup delay due to crankback across TE-LSPs.

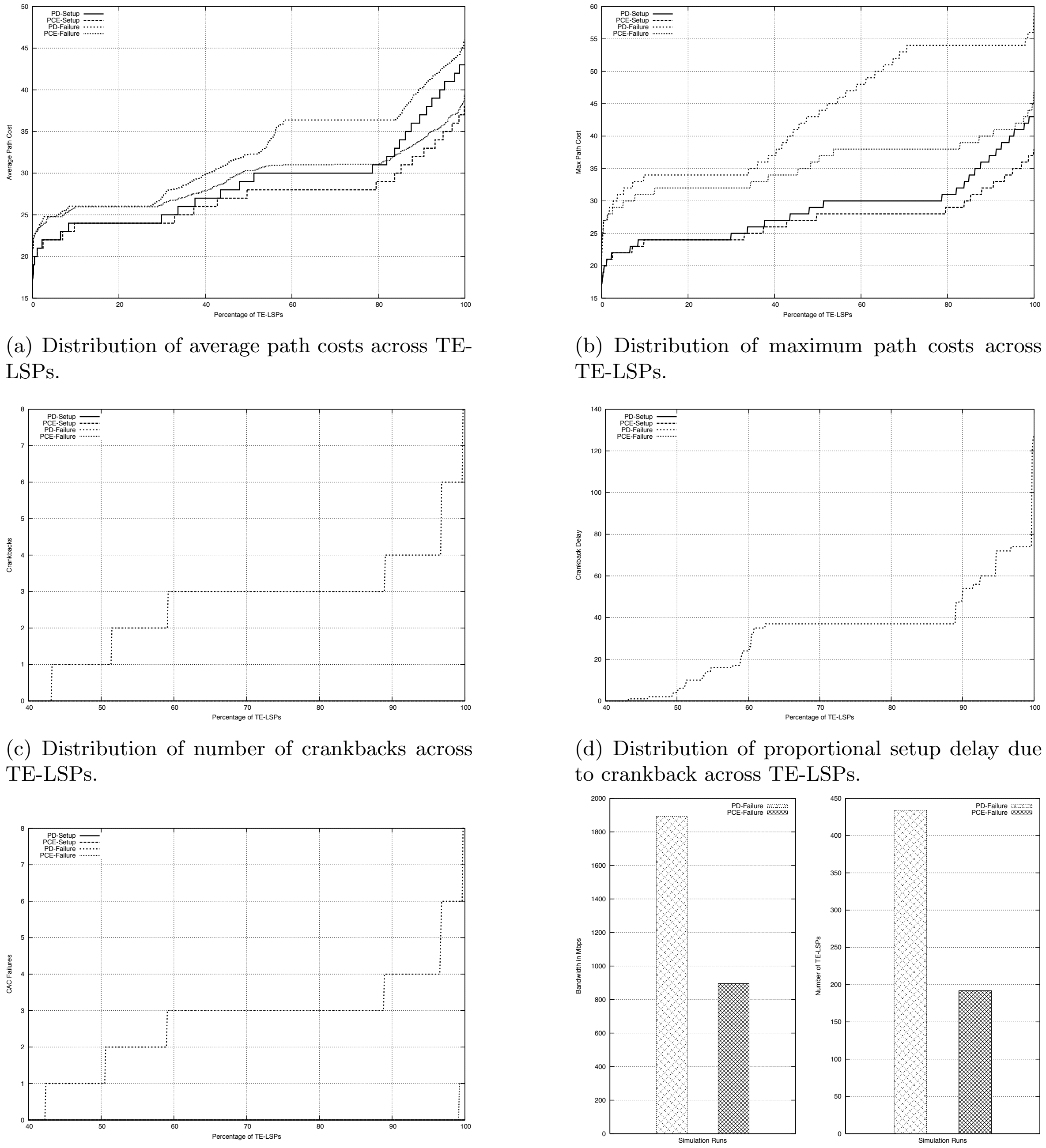
(e) Distribution of CAC failures across TE-LSPs.

(f) TE-LSPs and corresponding bandwidth that failed to find a route.

Figure 2: Results for MESH-CORE

5.3 Signaling Failures.

As discussed in the previous sections, signaling failures occur either due to an outdated TED or when a path cannot be found from the selected entry border router. Figures 2e and 3e shows the distribution of the total number of signaling failures experienced by the TE LSPs during setup. About 38% and 55% of TE LSPs for MESH-CORE and SYM-CORE, respectively, experience a signaling failures with per- domain path computation when link failures take place in the network. In contrast, only about 3% of the TE LSPs experience signaling failures with the PCE method. It should be noted that the signaling failures experienced with the PCE correspond only to the TEDs being out of date.



(a) Distribution of average path costs across TE-LSPs.

(b) Distribution of maximum path costs across TE-LSPs.

(c) Distribution of number of crankbacks across TE-LSPs.

(d) Distribution of proportional setup delay due to crankback across TE-LSPs.

(e) Distribution of CAC failures across TE-LSPs.

(f) TE-LSPs and corresponding bandwidth that failed to find a route.

Figure 3: Results for SYM-CORE

5.4 Failed TE-LSPs/Bandwidth on link failures.

Figures 2f and 3f show the number of TE LSPs and the associated required bandwidth that fail to find a route when link failures are taking place in the topologies. For MESH-CORE, with the per-domain approach, 395 TE LSPs failed to find a path corresponding to 1612 Mbps of bandwidth. For PCE, this number is lesser at 374 corresponding to 1546 Mbps of bandwidth. For SYM-CORE, with the per-domain approach, 434 TE LSPs fail to find a route corresponding to 1893 Mbps of bandwidth. With the PCE approach, only 192 TE LSPs fail to find a route, corresponding to 895 Mbps of bandwidth. It is clearly visible that the PCE allows more TE LSPs to find a route thus leading to better performance during link failures.

5.5 TE-LSP/Bandwidth Setup capacity.

Since PCE and the per-domain path computation approach differ in how path computation takes place, more bandwidth can be setup with PCE. This is primarily due to the way in which BRPC functions. To observe the extra bandwidth that can fit into the network, the traffic matrix was scaled. Scaling was stopped when the first TE LSP failed to setup with PCE. This metric, like all the others discussed above, is topology dependent (therefore the choice of two topologies for this study). This metric highlights the ability of PCE to fit more bandwidth in the network. For MESH-CORE, on scaling, 1556 Mbps more could be setup with PCE. In comparison, for SYM-CORE this value is 986 Mbps. The amount of extra bandwidth that can be setup on SYM- CORE is lesser due to its restricted nature and limited capacity.

6 IANA Considerations

This document makes no request to IANA for action.

7 Security Considerations

This document does not raise any security issue.

8 Acknowledgment

The authors would like to acknowledge Dimitri Papadimitriou for his helpful comments to clarify the text.

9 References

9.1 Normative References

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

9.2 Informative References

[DEF-DES]
J. Guichard, F. Le Faucheur, and J.-P. Vasseur, ”Definitve MPLS Network Designs”, Cisco Press, 2005.

[I-D.ietf-ccamp-crankback]
Farrel, A., ”Crankback Signaling Extensions for MPLS and GMPLS RSVP-TE”, draft-ietf-ccamp-crankback-06 (work in progress), January 2007.

[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-05 (work in progress), April 2007.

[I-D.ietf-ccamp-inter-domain-rsvp-te]
Ayyangar, A., ”Inter domain Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) Traffic Engineering - RSVP-TE extensions”, draft-ietf-ccamp-inter-domain-rsvp-te-06 (work in progress), April 2007.

[I-D.ietf-ccamp-lsp-stitching]
Ayyangar, A., ”Label Switched Path Stitching with Generalized Multiprotocol Label Switching Traffic Engineering (GMPLS TE)”, draft-ietf-ccamp-lsp-stitching-06 (work in progress), April 2007.

[I-D.ietf-pce-brpc]
Vasseur, J., ”A Backward Recursive PCE-based Computation (BRPC) procedure to compute shortest inter-domain Traffic Engineering Label Switched Paths”, draft-ietf-pce-brpc-06 (work in progress), September 2007.

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

[RFC3784]
Smit, H. and T. Li, ”Intermediate System to Intermediate System (IS-IS) Extensions for Traffic Engineering (TE)”, RFC 3784, June 2004.

[RFC4090]
Pan, P., Swallow, G., and A. Atlas, ”Fast Reroute Extensions to RSVP-TE for LSP Tunnels”, RFC 4090, May 2005.

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

[ROCKETFUEL]
N. Spring, R. Mahajan, and D. Wehterall, ”Measuring ISP Topologies with Rocketfuel”, Proceedings of ACM SIGCOMM, 2002.

Authors’ Addresses

Sukrit Dasgupta
Drexel University
Dept of ECE, 3141 Chestnut Street
Philadelphia, PA 19104
USA

Phone: 215-895-1862
Email: sukrit@ece.drexel.edu
URI: pages.drexel.edu/∼sd88

Jaudelice C. de Oliveira
Drexel University
Dept. of ECE, 3141 Chestnut Street

Philadelphia, PA 19104
USA

Phone: 215-895-2248
Email: jau@ece.drexel.edu
URI: www.ece.drexel.edu/faculty/deoliveira

JP Vasseur
Cisco Systems
1414 Massachussetts Avenue
Boxborough, MA 01719
USA

Email: jpv@cisco.com

Full Copyright Statement

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, THE IETF TRUST 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).