A collaborative PCE-Based mechanism for E2E recovery within MPLS networks

EL KAMEL Ali[a]

[a] *Research Unit Prince*
Tunisia
Email: ali.el.kamel@hotmail.com

Abstract

This paper deals with the End-to-End recovery issue into switched networks. It offers a solution to support and to recover link/node failure as well into a domain as between domains. The main work is done by specific nodes called PCE(Path Computation Elements), which are deployed once by domain and which should collaborate in order to establish a backup Tunnel that bypassed a point of failure. The proposed mechanism uses a computation procedure presented into [RFC 5298] and known as BRPC (Backward Recursive PCE-based Computation) in order to achieve End-to-end traffic restoration without care on failure location, heterogeneity and autonomy of crossed areas. Experimental results gives an idea about opportunity given by the proposed solution in terms of efficiency and time recovery compared to parallel approaches, regardless divergences of autonomous systems that can be crossed to reach destination.

Key words: MPLS, E2E recovery, PCE, PCEP

1. Introduction

For the last few years, Multi-Protocol Label Switching (MPLS), and also GMPLS (Generalized MPLS), seems to be the technology that all operators should mitigate. Indeed, this technology is able to offer traffic engineering, to support Quality of service, and to enable flexibility into networks which are abusively explored. Moreover, MPLS is able to offer a high aspect of security and to deploy private networks despite the diverging policies and the structural constraints between crossed domains.

Originally, the MPLS and GMPLS networks were limited to single domain environments. Increasingly, it was necessary to think on enabling communication and data exchanging between multiple domains MPLS and GMPLS, where a domain is considered to be any collection of network elements within a common sphere of address management or path computational responsibility. Examples of such domains include Interior Gateway Protocol (IGP) areas and ASs or BGP confederation. As a result, various approaches were introduced and discussed in order to establish Label Switched Paths (LSPs) in multiple

domain networks. The main purpose was considered critical since the established LSP should cross multiple domains discerned usually as heterogeneous and autonomous. Not only that, operators should deliver appropriate service to costumer over G/MPLS networks, which requires the specification of propitious solutions that can support Protection and recovery from failure circumstances. Such solutions have been described into the RFC4428. The [RFC4726] also defines a framework for inter-domain G/MPLS traffic engineering.// As well as research in this area are developed, Service providers point out the definition of several frameworks which can help to offer required quality of service. The Path Computation Element (PCE) [RFC4655] is one of such proposed architecture. Started in 2006 by the working group IETF, it is able to compute paths between two points, in order to bypass a point of failure. Firstly associated with switched networks (G/MPLS), a PCE was not able to define End-to-End backup Tunnel across a bundle of Autonomous Systems, generally equipped with different organizational architectures. This obvious diversity has implied specific extensions as well to current intra-domain routing protocol (IGP) as to inter-domain routing protocol (BGP), leading to the definition of mandatory entities and protocols required to deliver path computation information.

This paper describes an efficient mechanism to support E2E LSP Tunnels protection and restoration in spite of the policy disagreements among crossed domains. It is based on per-domain PCE (Path Computation Element) which should communicate and collaborate in order to establish a backup LSP Tunnel that bypassed any node/link failure.
Mainly, the mechanism is based on a procedure of finding pieces of a tunnel within each domain, thereby, joining them may give a backup tunnel to bypass the point of failure. The procedure of establishing the backup tunnel is backward since pieces joining is done in the backward, from the last domain close to the destination until the domain close to the failed point.

This paper is structured as follows. The section 2 presents a brief overview of related work. Section 3 defines basics of the proposed mechanism and presents an illustrative example. Performances evaluations are presented in Section 4. Conclusion and future work are given in Section 5.

2. Previous Work

First of all, last researches [2] in the recovery issue has proved that the standard protocol of inter-domain routing (BGP: Border Gateway Protocol), is generally associated with long convergence properties leading to a potential latency in paths failover and repair. Indeed, as depicted in [3], inter domain failover may reach over 3 minutes and can cause, therefore, several routing fluctuations up to 15 minutes. AS a result, it can lead to critical end-to-end packet loss rate and delay that may reach respectively a factor of 30 and 4 during path restoration. Therefore, new solutions are developed to deal with the problem of failure handling into wide routed or switched networks. Many approaches such as [RFC3496] have addressed the scope of inter-domain recovery and protection,

but only little work has deal with the End-to-End context. Researchers attested that the proposed solutions for intra domain recovery issue cannot be extended and are not practically feasible to inter-domain context. Indeed, most proposed approaches address one-to-one or peer-to-peer recovery issue. Thus, those solutions are also limited to satisfy inter-domain recovery between at most two ASs.

Generally, routing fluctuations is generally caused by capricious changes in the topology which make routing protocol unstable([RFC2439]). Indeed, the protocol BGP evaluates reach ability basing on advertising AS paths and considering as reached those who are stable and suppressing those considered as flapping networks. Although this feature avoids routing deficiencies, it may cause long convergence times and raises the trade-off between stability and convergence.
Furthermore, a solution has been proposed to solve inter-domain recovery issue. The main objective was to define a backup path for corresponding working LSP regardless the heterogeneity and the autonomy of various crossed domains. The objective has been faced with intra-domain recovery mechanisms inability due to several problems of scalability and inter-provider fault signaling divergence. The proposed solution, defined as IBLBT(Inter-domain Boundary Local Bypass Tunnel)[5], deals with the inter-domain MPLS recovery problem and is based on the establishment of independent protection mechanisms within each domain using concatenated primary and backup LSPs, minimal protection signaling between domains (using local repair bypass tunnels), and local repair at the domain boundaries.
From another hand, IETF has proposed several studies related to the context of inter-domain traffic engineering and fault recovery, generally, within next generation of networks based on emergent technologies like G/MPLS. Likewise, a framework proposed by [6] describes briefly the various failure cases to be addressed by Inter-Domain Fast Rerouting. Indeed, the failure scenarios associated with inter-domain TE may be caused by:

- 1) A crash of a domain edge node that is present in both domains. Recovery mechanisms should then take in consideration the sub-cases of colocating or not of the PSL (Point Switching LSR) and the PML (Point Merging LSR) within the same domain.
- 2) A failure of a domain edge node that is only present in one of the domains and
- 3) A Failure of an inter-domain link.

Finally, [7] proposes a solution that defines independent protection mechanisms within individual domains and which should merged at the domain boundaries. Simulation proved that the solution ensures significant advantages including fast recovery across multiple non-homogeneous domains and big scalability.
Clearly, several approaches have been proposed to deal with the inter-domain TE and recovery issue. However, most solutions are based on a specification

of per-domain computation mechanisms which should merge at domain boundaries. Moreover, no assumptions has been specified by those solutions on how to collaborate at domains frontiers and how to ensure E2E backup path establishment in spite of locally-defined policies and rules. The main issue that should be focused is "how to address E2E recovery"? This issue requires the definition of practical solutions and techniques that are able to deal efficiently with failure protection and recovery in E2E context.

3. Proposed Mechanism

3.1. Mechanism Overview

The proposed mechanism requires to be applied within domains running the REEQoS[12] approach. This approach defines one Master Node (MN) per domain which should communicate to other MNs over a private infrastructure denoted MN-BackBone. This backbone is established at network setup by the definition of high-prior paths joining MNs together. Moreover, an entity called PCE (Path Computation Element) must be associated with the MN at each domain. Communication between PCEs is ensured by the PCE-communication protocol (PCEP) [RFC5440]. The discovery of neighbor MNs or PCEs is ensured by IGP/BGP discovery future defined into RFC5088 and RFC5089. Into a domain, the local configuration of MN or PCE may be done by the network administrator.

3.1.1. REEQoS Overview

The REEQoS approach is defined as a solution for establishing E2E working LSP tunnel. Inter domain and intra domain defined paths are managed by MNs[12]. To be able to send a flow of data, the source should select the closest MN, noted MN_1, to which an admission Request is transmitted. The Path establishment request is then relayed until reaching a MN_n from a domain D_n such as it exists a TE LSP joined to the destination or can be used to reach the final destination. The figure (Figure 1) describes the procedure of establishing an End-to-End working Tunnel:

At each domain D_i, the MN should find two path segments: $I - LSP_i$ (Internal LSP) and $E - LSP_i^{i+1}$ (External LSP) that uses respectively to route flows into the domain D_i and from the domain D_i to a downstream domain D_{i+1}. Next domain can be statically configured or dynamically discovered via IGP/BGP extensions([RFC5088] and [RFC5089]). Thus, a Path Request (XPATH) is transmitted toward all discovered neighbor MN_{i+1}. The selection of the MN_{i+1} can be achieved using local policies or heuristics such as first incoming response or low overloaded one.
When no neighbor MN_{i+1} can satisfy request XPATH coming from an upstream MN_i or when a MN_i cannot find an appropriate I-LSP and/or E-LSP that connects domain D_i to a next domain D_{i+1}, the process of establishment is aborted and a XPathErr message is forward toward the MN_i. Each intermediate MN_j

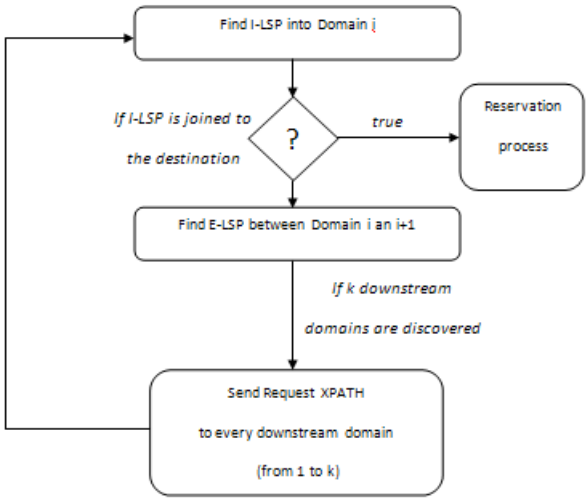

Figure 1: Working LSP Tunnel Establishment: Setup process

between the MN_i and MN_1 restarts another Path Request using the remaining set of relative discovered MNs, if they exist, or propagates conversely the XPathErr message to the upstream MN.
The Reservation process is started when an I-LSP from a domain D_n can be used to reach the final destination. Therefore, using the RSVP-TE or the CR-LDP protocols, an XRESV message is forwarded through the reverse established path in order to reserve required resources. The propagation is ensured using the RRO (Record Routing Object)[12] based on abstract Nodes identifiers such as AS Numbers. Finally, The E2E tunnel has been computed that supports the required QoS of the flow asking for being admitted on an E2E basis.

3.1.2. Mechanism description

When a failure occurs, the local PCE should initiate a recovery procedure. This procedure is denoted BRPC (Backward Recursive PCE-based Computation). The BRPC procedure is based on exchanging information between per-domain PCEs and tries to establish a backup LSP tunnel on a recursive way, using locally stored I-LSPs and E-LSPs(ref REEQoS).
Indeed, the closest node to the failed point should detect the failure using the HELLO message. There, the PCE_i transmits the request to one of neighbouring PCE_{i+1}. Discovery of adjacent PCE can be achieved using IGP/BGP extensions and requests throughout heterogeneous AS can be ensured using the PCEP extension defined into the draft[9]. The PCE of the source domain should receive a PCReq from the PCC (Path Computation Client) ingress node, which is the root of the segment path, part of the whole working LSP tunnel. The

procedure BRPC is described in Figure 2. It is initiated by the PCC(Path Computation Client). The PCC is the closest Node to the point of failure. The PCC should request for establishing a backup Tunnel at E2E scope. A message PCReq is transmitted toward the local PCE. It described associated SLA (Service Level agreement) which must be satisfied by the PCE when computing the backup Tunnel. The local PCE should distribute the establishment request toward all neighbor PCEs. Every PCE relayed the request PCReq until reaching the destination. No resources are allocated at this step. Resources Reservation is started by the Destination which returns a message PCRep through the reverse selected path. The PCRep is forwarded respectively to each PCE participating into the backup Tunnel. Receiving a PCRep means that pre-reserved resources should be allocated and associated with the already established Tunnel. During Backup Tunnel establishment, any PCE may receive more than one response from neighbor PCEs. Therefore, the PCE should compute several MP2P (Mutli-point to point) TE LSP tree. A P2MP intra-domain MPLS-TE LSPs tree is a TE LSP unidirectional tree that is initiated at an ingress Node within a domain and has one or more leaves as Egress nodes from the same domain[RFC4461]. As described by [RFC3209], the TE LSP tree is based on gathering several explicit paths, which may be constructed using strict or loose model. Each branch of a DBPT is returned as an explicit path (in which case, all hops are listed) or a loose path (in which only Ingress and Egress Nodes are specified). The choice between two models is fixed by the Network operator. Indeed, explicit rooting is used when no resource share is planned within a local domain. On the other hand, loose paths allow resources sharing. Finally, an TE LSP can be established using various constraint-based computation algorithm such as cost-based (CSPF) or QoS-based and may use any combination of later listed algorithms such as faire-cost QoS algorithm [RFC4461].

As a response to a PCEReq, a PCE_j may return one or more DBPT to the PCE_{j-1}. Selection of DBPT is based on the fact that at least one of the leaves of DBPTs can be joint to one or more ingress nodes from the downstream domain D_{j+1}. The ingress node of the domain D_j is defined as the node through which a flow admission request is received from the source, or, by which, a request XPATH, is transmitted from MN_{j-1} toward local MN_j(Figure).

In order to not transmit the totality of the DBPT toward upstream PCE, every P2MP tree is assigned a unique identifier, noted P2MP ID or P2ID, as depicted in [RFC4461]. This identifier is constant for the whole LSPs belonging to the same tree. The correspondence between the DBPT and its P2ID is maintained by the local PCE. A downstream PCE should only transmit the identifier of the tree and associated Ingress nodes.

3.2. Backward Recursive PCE-based Computation procedure (BRPC)

The procedure BRPC can be described as follows:

1. Step 1: When a failure occurs, the closest node to the failed point should determine the local-PCE. This node is noted a PCC(Path Computation

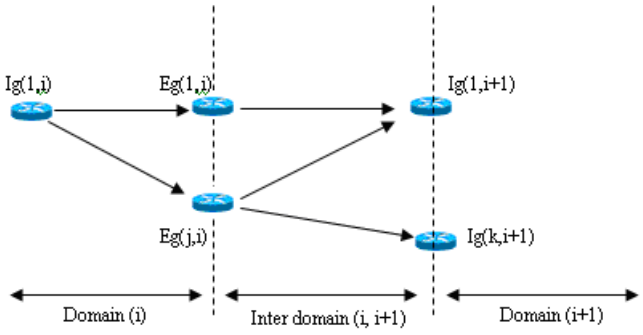

Figure 2: Per-domain DBPT establishment

Client). It transmits a PCReq toward it and requests a backup path computation. The path computation request is then relayed until reaching a PCE_n such that the TE LSP destination resides in the domain D_n.

2. Step 2: When the destination domain D_n is reached, the PCE_n computes the $DBPT_n$, which is made of the list of QoS-constrained I-LSP between every Ingress node I(j,n) and the TE LSP destination using a suitable QoS-based path computation algorithm and returns the resulting tree to PCE_{n-1} as a list of RRO objects.
3. Step i: For k:= n-1 to 1, the local PCE_k receives an association between the DBPT identifier and associated ingress node. This tree is computed by the PCE_{k+1}. Thus, it computes available segments of backup LSP tunnel using local I-LSP and E-LSP databases and the received DBPT. The resulting tree is transmitted toward the upstream PCE.
4. step n: The PCE_1, from which the backup Tunnel has been initiated, may receive a PCRep message containing a RRO object(using loose model) describing the backup LSP tunnel constructed using the procedure BRPC. The PCE_1 selects an appropriate path that it communicate to the PCC which starts the flow rerouting around the point of failure.

3.3. Inter-domain PCEP extension

In order to be able to exchange PCReq/PCRep messages on an inter-domain scope, a PCEP extension was proposed in [9]. A new object called 'PCE Sequence Object' is defined that represents the PCE topology tree. Indeed, every PCE is associated with a public Identifier, such as IPv4/IPv6 prefix, as it has been done with Master-Nodes specification and discovery into the REEQoS approach [12].
Moreover, PCE discovery can be achieved using IGP/BGP extensions as it has been defined into [RFC5073] and [14]. Once discovered, the PCEP messages can be forwarded both in intra and inter domain context. Some relevant problems

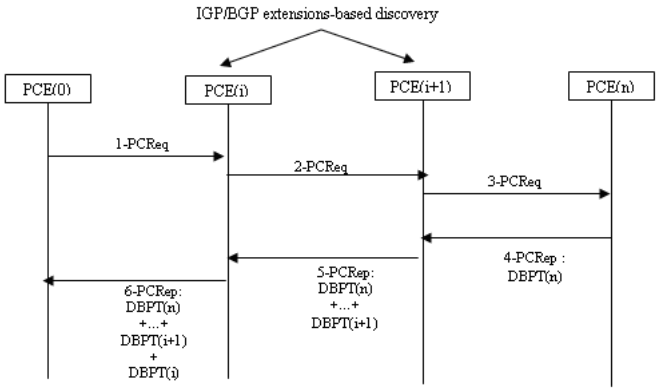

Figure 3: A backup LSP tunnel establishment diagram using BRPC procedure

have been announced, but not treated in the RFC5441. It has been signaled in section 3 [General Assumptions], that a sort of local constraint-mapping based on per-domain policy agreements is usually required in order to overcome constraints disagreements between Operators at border nodes. This can be achieved at PCC-to-PCC communications within domains boundaries, where a sort of translation is required in order to convert Per-domain policy-based constraints, involved into the PCReq message, in order to meet local policies.

4. Performance analysis

In this section, the mechanism is evaluated using analytic model and experimental simulations. Both results prove efficiency and benefits offered by the new solution of path computation in an E2E scope, in terms of establishing optimal backup LSP tunnel at convenient delay and with opportune resources reservation. Simulation is based on the network simulator NS2 (release 2.26) to which some modifications and extensions are introduced in order to be able to deploy specific agents located at dedicated LSRs within the topology.

4.1. Analytic evaluation

A network can be defined as G=(V,E) where N is a set of nodes and E is a set of links. Each component (node or link) can be on fail state or operating state. A failure of a link means that it is removed from the network, while the failure of a node means that the node and all joint links should be removed from the network. Generally, components fail is considered random and independent of one another. Specifically, each failed component i has an associated reliability p_i, describing the probability that it is operational. Let's denote n as the number of unreliable nodes ($n \leq |V|$) and m as the number of unreliable links ($m \leq |E|$). The network G can be in 2^{n+m} states, including the states of no failure.

Let $\Gamma_{ij} =< n_i, \cdots, n_j >$ the working path between source node n_i and destination node n_j. The probability that Γ_{ij} fails, depends on the probability that one of its components (links or nodes) fails. Considering only one component-failure cases, this probability can be written as follows:

$$P_{fail}(\Gamma_{ij}) = \sum_{k \in \Gamma_{ij}} [(1 - p_k) \prod_{j \# k} p_j] \tag{1}$$

In the case of f multiple failures ($f \leq |\Gamma_{ij}|$), the probability is more complicated. Let's denote $S_{fail} =< n_a, \cdots, n_b >$ as the set of failed components ($|S_{fail}| = f$). The probability can be expressed as follows:

$$P_{fail}^{S_{fail}}(\Gamma_{ij}) = \prod_{k \in S_{fail}} (1 - p_k) \prod_{j \notin S_{fail}} p_j \tag{2}$$

for all possible cases with f failures, we define the probability as follows:

$$P_{fail}(\Gamma_{ij}) = \sum_{S_{fail}} P_{fail}^{S_{fail}}(\Gamma_{ij}) = \sum_{S_{fail}} [\prod_{k \in S_{fail}} (1 - p_k) \prod_{j \notin S_{fail}} p_j] \tag{3}$$

Moreover the probability $P_k(\Gamma_{ij})$ that a path Γ_{ij} can support a FEC k is defined basing on ability of its components to support such FEC k. We can write this probability as follows:

$$P_k(\Gamma_{ij}) = \prod_{l \in \Gamma_{ij}} [f_k(l) R(l) p_l] \tag{4}$$

where:

- $f_k(l)$ is defined as follows:

$$f_k(l) = \begin{cases} 1 & \texttt{if the component l can support the FEC k} \\ 0 & \texttt{otherwise} \end{cases}$$

- R(l) is a function defined as follows:

$$R(l) = \begin{cases} 1 & \texttt{if the component l is operational} \\ 0 & \texttt{otherwise} \end{cases}$$

Let Δ_{ij} the set of all paths connecting n_i to n_j, and let Γ_{ij} be the working path. The cost-effective path that can be used to define a backup path of the Γ_{ij} has the probability $P_k^*(\Gamma_{ij})$ given below:

$$P_k^*(\Gamma_{ij}) = \max_{\Gamma'_{ij} \in \Delta_{ij} \setminus \{\Gamma_{ij}\}} (P_k(\Gamma'_{ij})) \tag{5}$$

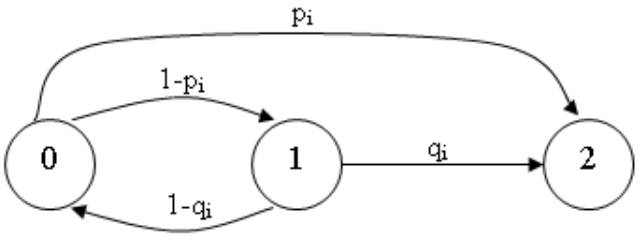

Figure 4: Diagram of the probability of finding a local backup path at Domain D_i

4.1.1. Comparative evaluation between backward Path Computation and Forward Path Computation

In this section, we evaluate probabilities of finding backup paths using respectively the Forward Path Computation and the BRPC. The Forward Path Computation consists of relaying a PCEReq toward downstream PCE until reaching the destination. At each domain, resources must be allocated. The probability of finding an LSP segment that may be used to establish an E2E LSP tunnel is described using the graph (Figure 4). where state 0 is defined as no connection point, at inter domain and intra domain scope, is found. State 1 denotes that a connection point was found at intra-domain scope, and state 2 defines that a connection point is reached at inter-domain scope. Let's denote p_i the probability of finding a connection point at domain D_i. Conversely, let's denote q_i the probability of finding a connection point between D_i and D_{i+1}.

The probability of finding a backup LSP tunnel at domain i is noted Pr_i. Using Forward Path Computation, it can be defined as follows:

$$\begin{cases} Pr_i = p_i + q_i(1-p_i) + (1-p_i)(1-q_i)Pr_{i+1} \forall i < n \\ Pr_n = p_n \end{cases} \tag{6}$$

for simplification, we define respectively

$$\begin{cases} \alpha_i = p_i + q_i(1-p_i) \\ \beta_i = (1-p_i)(1-q_i) \end{cases} \tag{7}$$

there, the probability of establishing a backup path can be simplified as follows:

$$\begin{cases} Pr_i = \alpha_i + \beta_i Pr_{i+1} \forall k < n \\ Pr_n = p_n \end{cases} \tag{8}$$

finally, we can prove that the probability $Pr_i^{(n)}$ of finding a backup path between the domain D_i, considered as the closest upstream domain to the failure

point, and the domain D_n, considered as the destination domain, can be writen as follows:

$$\begin{cases} Pr_i^{(n)} = \sum_{k=i}^{n-1} \alpha_k \prod_{j=i}^{k-1} \beta_j + p_n \prod_{k=i}^{n-1} \beta_k \forall i < n \\ Pr_n = p_n \end{cases} \tag{9}$$

Conversely, Computing an E2E backup LSP tunnel using BRPC requires necessary the computation of the probability of finding it at the starting domain. However, finding a backup LSP segment at domain i requires finding a backup LSP segment at domain i+1. The procedure is recursive since the probability of finding a backup LSP tunnel at domain i depends on the probability of finding a backup path on domain i+1, and so on. The complexity of the procedure at worst case is O(nlogn), where n represents the number of domains connecting the closest upstream domain to the destination domain.
From another hand, the probability of finding a backup LSP tunnel from a domain k to a destination domain n, when the establishment crosses domains to which the working LSP does not belongs to, is defined as the probability of finding both an I-LSP an E-LSP that does not connects the working LSP Tunnel. Explicitly, this probability can be computed without considering the probability of finding an I-LSP or an E-LSP, or both, that connect the working LSP tunnel. Those probabilities are defined as follows:

- The probability of finding both an I-LSP and E-LSP that does not connect the working LSP Tunnel: $(1 - p_i)(1 - q_i)$
- The probability of finding an I-LSP that connects the working LSP tunnel with an E-LSP that does not: $p_i(1 - q_i)$
- The probability of finding an I-LSP that does not connect the working LSP tunnel with an E-LSP that does: $(1 - p_i)q_i$
- Finally, the probability of finding an I-LSP and an E-LSP that both connect the working LSP tunnel: $p_i q_i$

The probability of finding a backup LSP tunnel using the BRPC procedure is defined as:

$$Pr_i = Pr_{i+1}[1 - (p_i(1 - q_i) + q_i(1 - p_i) + p_i q_i)] \tag{10}$$

We can prove that the probability $Pr_k^{(n)}$can be written as follows:

$$\begin{cases} Pr_i^{(n)} = p_n \prod_{k=i}^{n-1} \beta_k \forall i < n \\ Pr_n = p_n \end{cases} \tag{11}$$

Let suppose that $\forall k$; $p_k = p$ and $q_k = q$ the system above can be expressed as follow

$$\begin{cases} Pr_k = p + q(1 - p) + (1 - p)(1 - q)Pr_{k+1} \forall k < n \\ Pr_n = p \end{cases} \tag{12}$$

The probability at a domain k may be simplified as follow

$$\begin{cases} Pr_k = \alpha + \beta Pr_{k+1} \forall k < n \\ \alpha = p + q(1-p) \\ \beta = (1-p)(1-q) \end{cases} \tag{13}$$

Moreover, it is possible to demonstrate that the probability Pr_k can be defined as follow

$$Pr_k = \alpha \sum_{i=0}^{n-(k+1)} \beta^i + \beta^{n-k} p \tag{14}$$

Where n defines the destination domain and k defines the upstream domain closest to the failure point. This can be proved using the recurrence proof principle. First, the assumption is correct for k=n-1;

$$\begin{cases} Pr_n = p \\ Pr_{n-1} = \alpha + \beta Pr_n = \alpha + \beta P \\ Pr_{n-1} = \alpha \sum_{i=0}^{n-((n-1)+1)} \beta^i + \beta^{n-(n-1)} p \end{cases} \tag{15}$$

Let suppose that it remains correct for k=j; we must prove that it remains also correct for k=j-1. Indeed,

$$\begin{cases} Pr_j = \alpha + \beta Pr_{j+1} Pr_j = \alpha \sum_{i=0}^{n-(j+1)} \beta^i + \beta^{n-j} p \\ Pr_{j-1} = \alpha + \beta Pr_j \\ Pr_{j-1} = \alpha + \beta(\alpha \sum_{i=0}^{n-(j+1)} \beta^i + \beta^{n-j} p) \\ Pr_{j-1} = \alpha + \alpha \sum_{i=0}^{n-(j+1)} \beta\beta^i + \beta\beta^{n-j} p \\ Pr_{j-1} = \alpha + \alpha \sum_{i=1}^{n-j} \beta^i + \beta^{n-(j-1)} p \\ Pr_{j-1} = \alpha\beta^0 + \alpha \sum_{i=1}^{n-j} \beta^i + \beta^{n-(j-1)} p \\ Pr_{j-1} = \alpha \sum_{i=0}^{n-(j-1)-1} \beta^i + \beta^{n-(j-1)} p \\ Pr_{j-1} = \alpha \sum_{i=0}^{n-((j-1)+1)} \beta^i + \beta^{n-(j-1)} p \end{cases} \tag{16}$$

That it is, the probability of finding a backup LSP tunnel from domain k to a domain n is defined as follows:

$$Pr_k = \alpha \sum_{i=0}^{n-(k+1)} \beta^i + \beta^{n-k} p \tag{17}$$

In order to find an E2E backup LSP tunnel, it may be necessary to compute the probability of finding it at the starting domain. However, finding a backup LSP segment at domain k requires finding a backup LSP segment at domain k+1. The procedure is recursive since the probability of finding a backup LSP tunnel at domain k depends on the probability of finding a backup path on domain i+1, and so on. The complexity of the procedure at worst case is O(nlogn), where n represents the number of domains connecting the initial domain to the destination domain.
From another hand, the probability of finding a backup LSP tunnel from a

domain k to a destination domain n, when the establishment crosses domains to which the working LSP does not belongs to, is defined as the probability of finding both an I-LSP an E-LSP that does not connects the working LSP Tunnel. Explicitly, this probability can be computed without considering the probability of finding an I-LSP or an E-LSP, or both, that connect the working LSP tunnel. Those probabilities are defined as follows:

- The probability of finding both an I-LSP and E-LSP that does not connect the working LSP Tunnel: (1-p)(1-q)
- The probability of finding an I-LSP that connects the working LSP tunnel with an E-LSP that does not: p(1-q)
- The probability of finding an I-LSP that does not connect the working LSP tunnel with an E-LSP that does: (1-p)q
- Finally, the probability of finding an I-LSP and an E-LSP that both connect the working LSP tunnel: pq

The probability of finding a backup LSP tunnel using the BRPC procedure is defined as:

$$Pr_k = Pr_{k+1}[1 - (p(1-q) + q(1-p) + pq)] \tag{18}$$

considereing same assymptions, we consider that :

$$\begin{cases} Pr_k = \beta Pr_{k+1} \\ \beta = (1-p)(1-q) \end{cases} \tag{19}$$

Similarly, the probability of finding a backup LSP tunnel can be presented as follow:

$$Pr_k = \beta^{n-k} p \forall k \leq n \tag{20}$$

The probability of finding a backup LSP tunnel using the BRPC procedure is less than the probability of finding such backup LSP Tunnel using the Forward Path Computation. Indeed,

$$Pr_k^{(n)}(ForwardPathComputation) - Prk^{(n)}(BRPC) = \sum_{k=i}^{n-1} \alpha_k \prod_{j=i}^{k-1} \beta_j \geq 0 \tag{21}$$

Moreover, let's suppose a failure has occurred between domain i and domain i+1. The number of crossed domains needed to establish the backup LSP tunnel is exactly n-i, where n is the destination domain, using the BRPC procedure. This is due to the fact that the PCReq message is propagated toward the destination before the backup path establishment is started. However, it can be less than n-i using the Forward Path Computation. This depends on the probability of finding a local segment path that connects the original LSP tunnel. In this case, the Path Computation is aborted and the backup LSP tunnel is considered established.

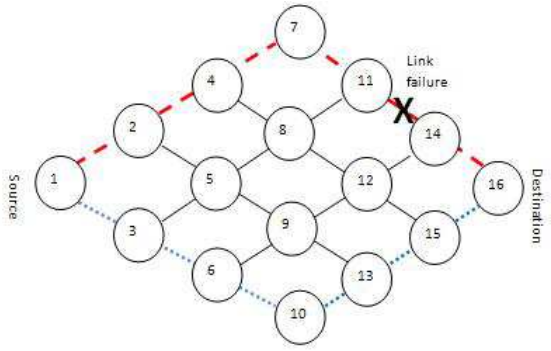

Figure 5: Simulation topology

Let suppose a failure has occurred between a domain k and a domain k+1. Similarly, let denote $\Pi_{i,j}^{m}$ the backup LSP tunnel that joins the LSR i (from domain k) to the LSR j, where j is the first connection point between the established backup LSP tunnel and the original working LSP tunnel and m is the number of crossed domains before reaching the connection point.

Let denote $\Gamma_{i,j}^{m}$ its cost. The shortest backup LSP tunnel is defined as $\Pi_{i,j}^{*}$ such as $\Gamma_{i,j}^{*} = \min_m \Gamma_{i,j}^{m}$. It is obvious that Forward Path Computation ensures less cost than BRPC. However, the previous cost is evaluated at E2E level and does not consider local costs when segment paths are established. Indeed, the BRPC procedure is based on the break-before-make model in which connection is found before making any resources reservation. The Forward Path Computation requires reserving resources at each domain before transmitting requests to next domain. Such process may introduce resources wasting since no guarantee is offered before reaching the destination domain. The propagation of PCReq, toward destination domain, offers a guarantee of establishing backup LSP tunnel without resources wasting nor network performances degradation. Optimisation is more ensured using BRPC procedure.

4.2. Simulation results

In order to test the efficiency of the proposed mechanism, we have run an experimental simulation. We have considered an MPLS network on which we have defined a working Tunnel. The working Tunnel joins the source S to the destination D. The source S wishes to transmit a VBR traffic having a packet size of 250Mb and an exponential variation of the inter leaving period. We have planned a failure to occur on link $(LSR_{11} <--> LSR_{14})$. In this simulation, we have compared the ability of the network to reach a steady state after a failure. We consider the BGP model and the E2E protection model. required definitions are presented in the figure 5. The red Tunnel presents the working Tunnel. The blue one presents the protection Tunnel.

Figure 6 shows that the proposed mechanism is more able to maintain a steady state compared with similar approaches. The BGP model takes more time before reaching the stability that have gone more than 1 min in this example. Otherwise, the proposed mechanism and the E2E recovery are able to ensure stability at nodes by reserving minimum amount of received packets before reaching the stable state. Eventually, the time needed for reaching this stable state is less than 60 s in two models. Furthermore, the proposed mechanism is able to reduce at most the overload of nodes by diminishing the number of received packets. The ratio factor of received packets between the E2E recovery model and the proposed model is over 1,11. It reaches a factor of 2,52 between the proposed mechanism and the BGP model.

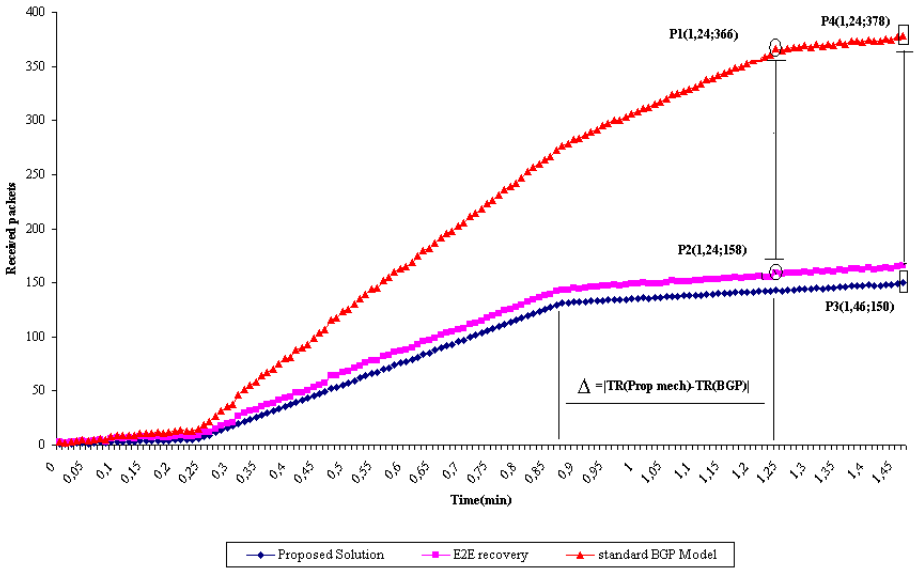

Figure 6: Survey of LSR_{14} overload using various recovery mechanisms

5. Conclusion

This paper presents a new mechanism of path computation at End-to-End scope. This mechanism requires to be explored within MPLS networks which are deployed accordingly to REEQoS approach. This solution is able to ensure failure handling and traffic protection despite heterogeneity and autonomy of crossed autonomous systems (AS). It is based on a recursive procedure called BRPC.
Simulations proved efficiency of the new mechanism in terms of resources utilization and E2E performances. The proposed mechanism offers more reliability

concerning stability of the network. Since the mechanism is applied on End-to-End context, it seems that it can support simultaneous points of failure. Scalability is also guaranteed, since few specific nodes (PCE) are involved into the recovery process. Finally, the proposed mechanism is able to recover from various types of failure: inter-domain link failure or border node failure.

References

[1] Vijayanand, C., Bhattacharya, S. and Kumar, P., *BGP Protocol extensions for PCE Discovery across Autonomous systems*, Work in Progress, June 2007

[2] Y. Rekhter, T. J. Watson, and T. Li, *A border gateway protocol 4 (BGP4)* IETF RFC 1771, Mar. 1995.

[3] C. Labovits et al., *Delayed Internet routing convergence* IEEE/ACM Trans. Networking, vol. 9, no. 3, pp. 293-306, June 2001.

[4] Changcheng Huang and Donald Messier . *A Fast and Scalable InterDomain MPLS Protection Mechanism*. JOURNAL OF COMMUNICATIONS AND NETWORKS, VOL. 6, NO. 1, MARCH 2004.

[5] A.Farrel, J-P Vasseur, A.Ayyangar,*A Framework for InterDomain MPLS Traffic Engineering*. Nov 2006, IETF RFC 4726

[6] Changcheng Huang and Donald Messier *Inter-Domain MPLS Restoration*, Design of Reliable Communication Networks (DRCN) 2003, Banff, Alberta. Canada, October 19-22, 2003

[7] Oki, E.; Inoue, I.; Shiomoto, K., *Path computation element (PCE)-based traffic engineering in MPLS and GMPLS networks* , Sarnoff Symposium, 2007 IEEE, April 30 2007-May,Page(s):1 - 5

[8] Q. Zhao, David Amzallag, Daniel King, *PCE-based Computation Procedure To Compute Shortest Constrained P2MP Inter-domain Traffic Engineering Label Switched Paths*, Internet-Draft, Mars 2009

[9] V. Sharma and F. Hellstrand, *Framework for multi-protocol label swtiching (mpls)-based recovery* 2003, request for comments: 3469

[10] El Kamel Ali and Youssef Habib; *an efficient hybrid mechanism for MPLS-based network*, accepted for publication in the 14th international Symposium on Computers and communications (ISCC09), July 5-8 2009, Tunisia.

[11] A.El Kamel and H. Youssef, *REEQOS: an RSVP-TE based approach for E2E QoS provisioning within MPLS domains*, VECOS2008, (Leeds UK 2-3 July 2008) published online within the British Computer Society(BCS)

[12] K. Kumaki , T. Murai , *PCEP extensions for a BGP/MPLS IP-VPN* , March 8, 2009 , Network Working Group,(Internet Draft) draft-kumaki-murai-pce-pcep-extension-l3vpn-02.txt,

[13] S. Matsushima, T.Murakami, K.Kenechi,*BGP extension for MPLS P2MP-LSP*, IEICE - Transactions on Information and Systems, Volume E89-D, Issue 1 (January 2006) , Pages 211-218

[14] C. Villamizar, R. Chandra, and R. Govindan, *BGP Route Flap Damping*, IETF RFC 2439,Nov. 1998.

[15] A.Awduche, L.Berger, T.Li, V.Srinivasan, G.Swallow,*RSVP-TE: Extension to RSV for LSP tunnels*, December 2001, IETF RFC3209

[16] A.Farrel, J-P Vasseur, A.Ashr, *A path Computation Element (PCE)-based architecture*, August 2006, IETF RFC 4655.

[17] S. Yasukawa, *Signaling Requirements for Point-to-Multipoint Traffic-Engineered MPLS Label Switched Paths (LSPs)*, April 2006,IETF RFC 4461

[18] J.P. Vasseur, J.L. Le Roux , *IGP Routing Protocol Extensions for Discovery of Traffic Engineering Node Capabilities*, December 2007, IETF RFC5073.

[19] Le Roux, J.L., Vasseur, J.-P., Ikejiri, Y., Zhang, R., *OSPF protocol extensions for Path Computation Element (PCE) Discovery*, RFC5088, January 2008.

[20] Le Roux, J.L., Vasseur, J.-P., Ikejiri, Y., Zhang, R., *IS-IS protocol extensions for Path Computation Element (PCE) Discovery*, RFC5089, January 2008.

[21] JP. Vasseur, JL. Le Roux, *Path Computation Element (PCE) Communication Protocol (PCEP)*, March 2009, IETF RFC5440.

[22] Q. Vohra, E. Chen, *BGP Support for Four-octet AS Number Space*, May 2007, IETF RFC 4893