

Local Recovery Solutions from Multi-Link Failures in MPLS-TE Networks with Probable Failure Patterns

Andrea Fumagalli, Marco Tacca, Kai Wu
Optical Networking Advanced Research (OpNeAR) Laboratory
Erik Jonsson School of Engineering and Computer Science
The University of Texas at Dallas
Email: {andrea, mtacca, kxw016500}@utdallas.edu

Jean-Philippe Vasseur
Cisco Systems
300 Beaver Brook Road
Boxborough, MA 01719 USA
Email: jpv@cisco.com

Abstract—MPLS TE Fast Reroute proposes a local protection mechanism to quickly reroute protected TE LSPs onto pre-computed and signaled bypass tunnels. This paper explores the case of multiple network element failure scenarios. The undesired complexity inherent to the multiple failure scenario originates from the fact that those failure scenarios are more disruptive, and may require multiple bypass tunnels to cope with.

The objective of this paper is to adapt the MPLS local recovery schemes to multi-failure scenarios, while controlling the number of bypass tunnels that are required. This objective is achieved by mapping multi-failure scenarios onto Probable Failure Patterns (PFP's). PFP's are characterized by their probability (or frequency) of occurrence during the network lifetime. A number of bypass tunnels is then computed to effectively cope with the PFP's according to their frequency or probability of occurrence.

It is shown that by properly choosing how the PFP's are grouped, and how the corresponding bypass tunnels are computed, it is possible to trade the required number of bypass tunnels for their average length and outage probability, i.e., the probability that the local recovery scheme cannot cope with the occurrence of a multi-failure pattern.

I. INTRODUCTION

As more real-time and mission-critical network applications rely on IP/MPLS networks, prompt recovery of data exchange at the IP/MPLS layer from network element failures is becoming increasingly important.

Fast ReRoute (FRR) [1] is a scalable recovery scheme in MPLS networks with Traffic Engineering (TE). FRR is based on *local recovery*, whereby traffic rerouting is performed by the *local nodes*, the nodes immediately upstream to the failing network element. For each network element, one or more bypass tunnels are pre-computed by the local nodes. Upon detection of an element failure, the affected TE LSP's are swiftly rerouted locally via available bypass tunnels. This technique guarantees prompt recovery by: 1) using local signaling for fault detection and 2) making use of pre-computed bypass tunnels. The time required to complete the traffic recovery depends solely on the time required to detect the (local) failure

and to perform the rerouting onto the respective bypass tunnel. Thus, it is not affected by the network size.

Besides single network element failures, today's network are increasingly facing a variety of additional failure patterns. These patterns include multiple network element failures. Multiple outages may be either correlated by a common physical failure and occurring concurrently, or not correlated by a common physical failure and occurring over a short period of time [2].

The extension of the FRR solution to cope with these additional failure patterns represents, however, a challenge. First, a multi-failure pattern is expected to generate greater disruptions. Second, pre-computation of the bypass tunnels must take into account the occurrence of all the probable multi-failure patterns. For example, multiple bypass tunnels may be required to protect the same LSP, each tunnel being chosen to cope with a specific set of failure patterns. An excessive number of bypass tunnels may increase the complexity of the local recovery scheme unnecessarily, without significantly improving the scheme reliability level, along with the challenge to accurately determine the appropriate bypass tunnel to activate. Another important aspect to consider is the delay and other QoS metric of the bypass traffic, e.g., the bypass tunnel length.

In this paper the authors propose an MPLS local recovery scheme and study the use of FRR to cope with multi-failure scenarios. The scheme is based on the concept of *facility*. *Facility* is a path in the network used to route one or more LSP's. The term *facility* is adopted consistently with the term *facility backup* used in the FRR draft [1]. A facility can have multiple links.

The objective here is to provide a framework for choosing which facilities must be protected and how many and which bypass tunnels for each facility must be used by the recovery scheme to protect a facility in the event of the failure of multiple network elements.

The approach is based on a probabilistic model, which assumes that the statistics on network failures are available to the MPLS control plane. This assumption is supported by the fact that, in IP/MPLS networks, some network elements are

more prone to failure than others, and some network element failures are more coupled than others [2], [3]. Using these available statistics, a number of *Probable Failure Patterns (PFP)* is defined. The *PFP* element set is the set of network elements that fail when the *PFP* fails. Each *PFP* is assigned a failure probability that reflects the likelihood, or frequency, of that failure pattern to occur. Multiple *PFP*'s may be logically grouped to form a *PFP cluster*. A bypass tunnel is then pre-computed for each facility that belongs to the *PFP* cluster, taking into account the *PFP*'s failure probabilities.

The advantage of the proposed probabilistic approach, based on *PFP* clusters and failure probabilities, is twofold. By controlling the number of *PFP* clusters, it is possible to control the number of required bypass tunnels, and thus control the complexity of the recovery scheme. By taking into account the *PFP* failure probabilities, it is possible to determine the *outage probability* of the chosen bypass tunnels, defined as the probability that the multi-failure pattern cannot be circumvented by the recovery scheme.

Three solutions to select the *PFP* clusters and the bypass tunnels are presented in the paper. The *Fault-Dependent (FD)* solution assigns each *PFP* to a distinct *PFP* cluster. This is the smallest size possible for each *PFP* cluster. This solution, however, requires the largest number of bypass tunnels. The *Fault Independent (FI)* solution assigns all *PFP*'s to one single *PFP* cluster. This solution requires a single bypass tunnel per facility. The *Hybrid (H)* solution makes use of $k \geq 2$ *PFP* clusters, where k is an integer chosen by the network designer. Hybrid solutions represent intermediate cases between the FD and FI solution.

As discussed in the paper, these three solutions provide a variety of options for designing local recovery schemes against multiple failures. By means of these solutions, it is possible to control the scheme complexity in terms of number of required bypass tunnels, probability of traffic interruption in terms of the outage probability of the bypass tunnels, and QoS performance of the recovery scheme in terms of increased propagation delay, e.g., length of the bypass tunnels.

II. PROBABLE FAILURE PATTERNS (*PFP*) AND *PFP* CLUSTERS

This section defines in detail the Probable Failure Pattern (*PFP*) and the *PFP* cluster.

Consider a MPLS network with arbitrary topology. Let the MPLS network be represented by a directional graph $G(\mathcal{N}, \mathcal{E})$, where \mathcal{N} represents the set of network nodes and \mathcal{E} the set of network links.

A *PFP* is defined for each failure scenario whose probability to occur is non-negligible. A *PFP* is represented by a *PFP* element set and a failure probability level, FPL . The *PFP* element set contains all the network element that concurrently fail in the represented failure scenario. Notice that in a node failure scenario, all the links adjacent to the node are also in the *PFP* element set. The FPL denotes the probability that the represented failure scenario occurs.

The *PFP* may be viewed as the generalization of the well-known Shared Risk Link Group (SRLG) [4]. The typical interpretation of a SRLG implies that all the links that belong to the same SRLG share a common physical failure. The *PFP* have a more general interpretation than the SRLG. The network elements in a *PFP* are those that may fail together, not necessarily due to a common physical failure, within a short period of time with non-negligible probability. It is assumed that all *PFP* represent events that are mutually exclusive with one another, i.e., when a *PFP* fails, all remaining network elements not in the *PFP* element set are operational. With this assumption, the link failure probability is the sum of the failure probabilities of all the *PFP* that contain the link. Finally, a *PFP* carries more information than an SRLG, — i.e., the parameter FPL , which indicates the probability of the failure.

Multiple *PFP*'s may be grouped together to form a *PFP* cluster. A given *PFP* cluster contains all the network elements that belong to the grouped *PFP* element sets. A facility is said to belong to a *PFP* cluster (or a set of *PFP*) if the network elements of the facility entirely belong to a *PFP* in the *PFP* cluster. A facility may belong to one or more *PFP* clusters, for each of which a *bypass tunnel* is required. The bypass tunnel provides an alternative routing solution to all the LSP's that are routed through the facility. The facility-based bypass tunnel can be implemented by means of MPLS label stack.

III. THREE SOLUTIONS TO COMPUTE THE *PFP* CLUSTERS

This section describes three solutions that are proposed to build the *PFP* clusters: the failure independent (FI) solution, the hybrid (H) solution, and the failure dependent (FD) solution.

It is assumed that all the *PFP*'s in the network are given. The two objectives of all three solutions are: (1) minimize each bypass tunnel length, and (2) minimize the probability that the traffic cannot be recovered after a failure of a *PFP*. These two objectives may be in contrast with each other, and a tradeoff must be found.

For each link $e \in \mathcal{E}$, a *PFP* set P_e is defined as the set of *PFP* whose element set contains link e . P_e are mapped on to one or more *PFP* clusters. The upstream node of link e is responsible to reroute traffic on the facilities that originate at the node, pass through link e and belong to P_e . Once a *PFP* cluster is identified (after a *PFP* failure occurs), all and only traffic through those facilities belonging to the *PFP* cluster are rerouted to available bypass tunnels. Depending on the partitioning of P_e into *PFP* clusters, the three solutions are obtained.

A. Failure Independent (FI) Solution

For each link e , the entire *PFP* set P_e is the *PFP* cluster. For each facility belonging to P_e , only one bypass tunnel is computed using the algorithms described in Section IV.

The FI solution requires the minimum fault detection capability for local link failures. When failure of link $e \in \mathcal{E}$

is detected, the only *PFP* cluster, P_e , is considered failed and traffic on all the associated facilities are rerouted. It is not necessary to identify the particular failed *PFP*. The drawback of this solution is that — since each *PFP* cluster possibly contains a large number of *PFP*'s — in some cases it is not possible to find bypass tunnels that avoid all possible *PFP*'s in the *PFP* cluster. In this circumstance, bypass tunnels that have minimal failure probability are preferred. For each facility, the failure probability \mathcal{P} is then given by the sum of the failure probabilities of all the *PFP*'s that affect both the facility and the bypass tunnel.

B. Hybrid (H) Solution

For each link e , P_e is partitioned into k disjoint subsets. Each subset is a *PFP* cluster. Each facility belonging to P_e belongs to one up to k *PFP* clusters, therefore has one up to k bypass tunnels. The challenge of this solution is to find the optimal partitioning of set P_e such that the failure probability of both the facility and the associated bypass tunnels is minimized. The partitioning and the bypass tunnels may be computed using the algorithms described in Section IV. The facility failure probability \mathcal{P} is then the sum of the failure probabilities of all the *PFP*'s that disrupt both the facility and all the k bypass tunnels.

The H solution requires a more complex failure detection mechanism than the FI solution. While it is not necessary to determine exactly which *PFP* fails, it is necessary to determine to which *PFP* cluster the failed *PFP* belongs. The higher the value of k , the smaller the number of *PFP*'s in each of the *PFP* clusters, therefore, the smaller is the combined failure probability of the facility and the associated bypass tunnels. On the other hand, the higher the value of k , the more accuracy is required by the fault detection mechanism in order to correctly identify the *PFP* cluster that contains the failed *PFP*.

C. Failure Dependent (FD) Solution

Set P_e is partitioned such that each *PFP* constitutes a distinct *PFP* cluster. For each *PFP* cluster, each facility is protected by a distinct bypass tunnel. Bypass tunnels may be computed using the algorithms described in Section IV.

This solution requires the most complex failure detection mechanism. Hence, it may require the longest detection time. In order to activate the appropriate bypass tunnel it is necessary to correctly identify the failed *PFP*. On the other hand, since the *PFP* cluster contains only one *PFP* and it is reasonable to assume that the network has enough redundancy to overcome any individual *PFP* failure, the FD solution represents the most reliable option.

IV. ALGORITHMS TO COMPUTE THE BYPASS TUNNELS

This section details the algorithms that are proposed to compute the bypass tunnels in the three solutions described in the previous section. The problem is divided into two subproblems: 1) partitioning of P_e into disjoint *PFP* cluster(s); 2) find bypass tunnel for each facility in each *PFP*

cluster. The mapping is straightforward in the FI and FD solutions, but more complex in H solution. A fast and efficient algorithm partitioning P_e into *PFP* clusters in the H solution is proposed next.

The optimal choice is based on a tradeoff between the hop count of the bypass tunnel¹ and a term associated with the reliability of the bypass tunnel. The function used to evaluate the goodness of any given set of bypass tunnels (that are associated with a facility) is a convex linear combination defined as follows:

$$f(\mathcal{H}, \mathcal{P}) = (1 - a)\mathcal{H} + a \log\left(\frac{\mathcal{P}}{P_{min}}\right) \quad (1)$$

where:

- \mathcal{H} is the average number of hops associated with all the bypass tunnels chosen to protect the facility, i.e., one bypass tunnel for the FD solution, up to k bypass tunnels for the H solution, and one bypass tunnel per *PFP* for the FI solution;
- \mathcal{P} is the probability of facility outage, caused by the inability of the bypass tunnels to circumvent some of the *PFP*'s in the *PFP* cluster;;
- P_{min} is the failure probability associated with the *PFP*'s of the highest *FPL*, i.e., the less likely failures.

The optimal bypass tunnels are those that minimize the ranking function in Eq. 1. With different values of $a \in [0, 1]$, it is possible to explore a wide range of options. When $a = 0$, the shortest bypass tunnel is selected, irrespective of its reliability. When $a = 1$, the most reliable bypass tunnel is selected, irrespective of its hop count.

A. Computing Bypass Tunnels for the FI Solution

Computing the bypass tunnels is more complex when dealing with either the FI, since there may be more than one *PFP* to avoid for a bypass tunnel. Two algorithms are presented to compute bypass tunnels for the FI solution. The first finds the optimal bypass tunnels while the second is a faster heuristic with suboptimal solutions.

1) *Optimal Folding Algorithm*: This algorithm computes the optimal bypass tunnels for each facility in a given *PFP* cluster based on the ranking function in Eq. 1. A folding technique is applied to reduce the search space. Folding is based on the concept of *safe connected components*, defined next.

Each link $e \in \mathcal{E}$ is a *safe link* if e is not contained in any of the *PFP*'s in the given *PFP* cluster. In other words a link is safe with respect to a facility if it is not in the same *PFP* cluster. A safe connected component is a subgraph $S(\mathcal{N}', \mathcal{E}')$ of $G(\mathcal{N}, \mathcal{E})$ where the shortest path between two nodes $i, j \in \mathcal{N}'$ makes use of only safe links. Therefore, if a subpath of a bypass tunnel goes from node $i \in \mathcal{N}'$ to node $j \in \mathcal{N}'$, it is not necessary to consider all possible paths from i to j . In fact, the shortest path, with zero failure probability, is the best

¹In this paper it is assumed that the QoS metric is proportional to the hop length of the bypass tunnel.

subpath that can be found according to the ranking function in Eq. 1.

The algorithm is therefore based on running a k-loopless-shortest path algorithm on an auxiliary graph that is constructed as follows. First, all the maximal safe connected components² are found. Each maximal safe connected component is represented by a distinct node in the auxiliary graph G_a . A link between two nodes i_a and j_a in G_a exists if there is at least one (unsafe) link in graph G that connects two nodes i and j , where i is in the maximal safe connected component that is represented by i_a , and j is in the maximal safe connected component that is represented by j_a . Notice that since each link in G_a corresponds to one or more links in G , and each node corresponds to a component which contains one or multiple subpaths in G , each shortest path on graph G_a can be expanded to one or multiple paths on graph G . All the paths found on G_a are then expanded and ranked based on Eq. 1. The best path is chosen to be the bypass tunnel.

2) *Heuristic Algorithms*: In spite of the folding technique, the algorithm presented in the previous section may be computationally intensive. In this section, a suboptimal algorithm that can be used to compute the bypass tunnel for any given facility and a given PFP cluster is presented. The solution is based on assigning a weight w_e to every link $e \in \mathcal{E}$ using the following expression:

$$w_e = (1 - a) + a \log \frac{\mathcal{P}_e}{P_{min}}, \quad (2)$$

where \mathcal{P}_e is the link failure probability that takes into account only PFP 's in the given PFP cluster. The bypass tunnel for the FI solution is found running a shortest path algorithm using weights w_e . The sub-optimality of this solution consists in the fact that if two network links along the bypass tunnel belong to the same PFP in set P_e , then the failure probability associated with that PFP is erroneously counted twice in determining the bypass tunnel.

B. Computing Bypass Tunnels for the H Solution

The H solution presents the additional problem of partitioning P_e into k PFP clusters. Partitioning is done using the following procedure. First, link e is considered³. The k bypass tunnels associated with facility corresponding to link e can be calculated in two ways:

- 1) all the possible choices for k bypass tunnels are enumerated, the combination of k bypass tunnels that minimizes the function defined by equation 1 is chosen
- 2) all links in the network are associated a cost w_e as in equation 2, then a shortest k disjoint path [5] is run to find the k bypass tunnels.

The found k bypass tunnels associated with the facility corresponding to link e can then be used to determine the k PFP clusters. The k bypass tunnels are analyzed sequentially.

²A maximal safe connected component is a safe component with maximal number of nodes.

³Link e belongs to all k PFP cluster irrespective of the partitioning procedure.

TABLE I
PFP DISTRIBUTION

Fault type	FPL0	FPL1	FPL2	FPL3	Safe
Single Link	90%	10%	0	0	0
2 Adjacent Links	0	33.3%	33.3%	33.3%	0
2 1-hop-away Links	0	0	25%	25%	50%
2 2-hop-away Links	0	0	0	20%	80%
Single Node	0	0	50%	50%	0
3 Random Links	0	0	0.05%	0.05%	99.9%

If a PFP_j in P_e does not disrupt the i^{th} bypass tunnel ($1 \leq i \leq k$) then PFP_j is in the i^{th} PFP cluster.

Once the PFP clusters are determined, all facilities in each cluster must have one bypass tunnel⁴. This can be done using one of the two algorithms described in the previous section.

C. Computing Bypass Tunnels for the FD Solution

In the FD solution, since there is only one PFP in every PFP cluster, all the bypass tunnels associated with each PFP cluster are always available unless the links in the PFP set form a cut, in which case no recovery schemes can recover the failure. The optimal bypass tunnel can be found in this case by using a shortest hop count algorithm with the constraint not to use any of the network links in the PFP cluster.

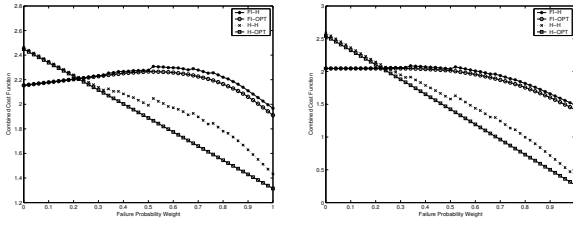
V. SIMULATION AND RESULTS

In this section, a study case is presented to evaluate the performance of the three solutions and the algorithms proposed to compute the bypass tunnels. A 19 node and 39 link European network topology is used in the study. Four different values for FPL are considered. Value $FPL = 0$ corresponds to a failure probability equal to p_0 . Value $FPL = 1$ corresponds to a failure probability equal to $p_0 10^{-1}$. Value $FPL = 2$ corresponds to a failure probability equal to $p_0 10^{-2}$. Value $FPL = 3$ corresponds to a failure probability equal to $p_0 10^{-3}$. With these values, $P_{min} = p_0 10^{-3}$. The PFP 's are randomly created, according to the distribution shown in Table I. The link failures in all PFP are assumed to be bi-directional. Shown curves and values are obtained by averaging results that are obtained over one hundred simulations, each using an independently generated set of PFP 's. For the H solution, $k = 2$ bypass tunnels per facility are considered.

Figure 1(a) plots the average value of the ranking function, that is obtained by the three solutions for the one-link facilities. Similar curves are shown in Figures 1(b) and 2(a), in which two-link facilities and all facilities are taken into consideration, respectively. The curves reveal that the performance gap between the optimal algorithm and the heuristics suboptimal algorithm does not exceed 5%, for both the FI and the H solution.

Figure 2(b) plots the average facility outage probability versus the average tunnel length, obtained for the one-link facilities. Similar curves are shown in Figures 3(a) and 3(b), in which two-link facilities and all facilities are taken into

⁴Except the facility associated with link e , for which the bypass tunnels have already been determined.



(a) One-link facilities

(b) Two-link facilities

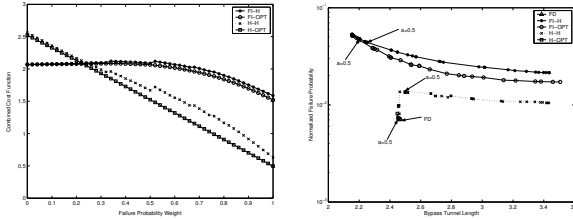
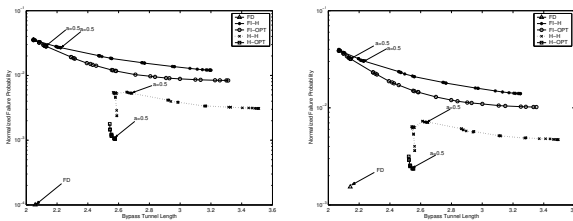
Fig. 1. Function 1 versus a (a) Function 1 versus a for all facilities(b) Failure probability \mathcal{P} versus tunnel length for one-link facilities

Fig. 2. Cost, failure probability, tunnel length

consideration, respectively. The curves indicate a clear tradeoff between the average hop count of the bypass tunnels, and the facility outage probability. (Recall that the latter represents the probability that the restoration scheme cannot restore a facility in the presence of some PFP .) The H solution achieves a better tradeoff, at the cost of using two bypass tunnels per facility, when compared to the single bypass tunnel of the FI solution.

Table II reports the number of bypass tunnels per facility that are required by the three solutions. Shortest bypass tunnels that protect the facility from different PFP clusters are merged into one single tunnel when they follow the same path. The FD solution has an average number of bypass tunnels for



(a) Two-link facilities

(b) All facilities

Fig. 3. Outage probability \mathcal{P} versus tunnel length

TABLE II

COMPARISON: NUMBER OF BYPASS TUNNELS PER FACILITY

Facility	FI	H		FD	
		Avg	Max	Avg	Max
all	1	1.7644	2	2.1137	3
one-link	1	1.8748	2	2.1549	4
two-link	1	1.7417	2	2.1003	4

each facility that is slightly in excess of 2, with maximum values as high as 4. The H solution has an average less than 2 while FI solution remains 1.

VI. CONCLUSION

The paper proposed a probabilistic approach to efficiently handling multi-failure scenarios in MPLS networks that make use of local recovery. The objective of the proposed approach is to provide a way to control the number of bypass tunnels that are required to handle multi-failure patterns, while monitoring the expected length and outage probability of the chosen bypass tunnels.

To assess the validity and the practicality of the proposed approach, three solutions were presented and compared. The fault-independent solution selects only one bypass tunnel for each facility in the network. The hybrid solution selected a predetermined number of bypass tunnels for each facility, being the two-tunnel case the one studied numerically in the paper. The fault-dependent solution selects one bypass tunnel for each probable failure pattern (PFP) and each facility. An increasing number of bypass tunnels are required, while ranging from the fault-independent to the fault-dependent solution. The expected length and outage probability of bypass tunnels are both parameters that can be controlled in the first two solutions.

The proposed approach appears to be flexible and yields a variety of solutions to handling multiple-failures in local recovery schemes. The network designer can thus choose the set of bypass tunnels that yields the desired complexity of the local restoration scheme (likely to be proportional to the number of bypass tunnels), expected length of the bypass tunnels, and outage probability of the bypass tunnels.

REFERENCES

- [1] P. Pan and et al, "Fast reroute extensions to RSVP-TE for LSP tunnels," IETF, June 2003, draft-ietf-mpls-rsvp-lsp-fastreroute-03.txt.
- [2] G. Iannaccone, C.-N. Chuah, R. Mortier, S. Bhattacharyya, and C. Diot, "Analysis of link failures over an IP backbone," in *ACM SIGCOMM Internet Measurement Workshop*, Marseilles, France, Nov. 2002.
- [3] A. Markopoulou, G. Iannaccone, S. Bhattacharyya, C.-N. Chuah, and C. Diot, "Characterization of failures in an IP backbone," in *To appear in IEEE Infocom*, Hong Kong, March 2004, Sprint ATL Research Report.
- [4] J. Strand and A. Chiu, "Issues for routing in the optical layer," *Communications Magazine*, vol. 39, no. 2, pp. 81–87, Feb 2001.
- [5] R. Bhandari, *Survivable Networks: Algorithms for Diverse Routing*, Kluwer Academic, 1998.