Restoration in Multi-Domain GMPLS-based Networks

Anna Manolova, Sarah Ruepp, Lars Dittmann

DTU Fotonik Technical University of Denmark, Byg. 343/130, 2800 Kgs. Lyngby, Denmark

Eusebi Calle, Jose Marzo

Institute of Informatics and Applications (IIiA), University of Girona, 17071 Girona, Spain

Abstract

In this paper we evaluate the efficiency of using restoration mechanisms in a dynamic multi-domain GMPLS network. Major challenges and solutions are introduced and two well-known restoration schemes (End-to-End and Local-to-End) are evaluated. Additionally, new restoration mechanisms are introduced: one based on the position of a failed link, called Location-Based, and another based on minimizing the additional resources consumed during restoration, called Shortest-New. A complete set of simulations in different network scenarios show where each mechanism is more efficient in terms, such as, resource overbuild or recovery delay.

Keywords:
multi-domain network, Generalized Multiprotocol Label Switching, restoration, location-based recovery.

1. Introduction

Current network technologies enable the transportation of a huge volume of information. Therefore, the consequence of a failure becomes more pronounced. Network reliability is seen as a key requirement for the next generation networks. Reliability can be provided through different fault management mechanisms applied at different network levels and time scales. A crucial aspect in developing a fault management system is the creation and routing of backup paths. Recovery paths can be created before a failure occurs (protection), in this case the connection backup paths are pre-
established or after the failure (restoration), in this case the backup paths are signalled and/or routed in reaction to network faults. Protection schemes provide faster recovery of the failed path but require preplanned and/or pre-established protection paths, which increase the needed capacity in the network. On the other hand, restoration techniques do not require additional allocation of protection resources. Both rely on the existing infrastructure and utilize the available spare capacity pool at the time of the failure.

With the increasing of the available transmission capacity and the advances in providing new transport services, new requirements for failure handling have emerged. In particular the requirements for differentiated reliability provisioning have attracted attention and have resulted in extensive research work [1, 2]. Providing differentiated failure handling can be based not only on the service type but on other factors as well. Correlation between the impact of a failure and the position of a failed link on the path is made in [3] and a novel routing mechanism which improves the availability of the established connections is proposed. Unlike most of the research work in the area of differentiated reliability the work presented in this paper focuses on multi-domain link failure scenarios due to the specific implications that the multi-domain environment pose on the reliability mechanisms. In this context, we propose and investigate the efficiency of standard restoration proposals and introduce a position-based failure handling technique, which provides differentiation based on the actual position of the failure along the path of an affected connection. Furthermore, we also introduce a new technique based on minimizing the additional resources consumed during restoration, called Shortest-New.

This work is an extension of the work presented in [4] and is focused on survivability in multi-domain single-layer connection-oriented network such as Generalized Multi-Protocol Label Switched networks (GMPLS) [5]. However, it can be easily extended to any other connection-oriented network technology.

The paper reviews standard survivability principles in Section II, then introduces what are the main challenges to apply restoration in multi-domain networks in Section III. In Section IV, two novel algorithms are described. Section V analyzes the results, obtained via simulations, and in section VI conclusions are presented.
2. Network Survivability Principles

There are many literature sources related to survivability provisioning (protection and restoration) in single-domain networks, among which the works in [6, 7]. In this section, the standard recovery mechanisms are reviewed. End-to-end and local-to-end schemes [8] are introduced as the main proposals to deploy restoration in multi-domain networks. A standard recovery process is also depicted in order to introduce some concepts used by our proposed algorithms.

2.1. Standard connection recovery schemes

Considering connection-oriented networks such as GMPLS, the availability of an established Label Switched Path (LSP) depends on the applied recovery mechanism [9], the availability of spare resources and the network topology as a whole [10]. Three main recovery strategies exist. Figure 1 a) depicts an End-to-End (E2E) recovery, also called path recovery. In this case, once the failure is detected a notification is sent to the node responsible for the recovery (in this case node 1, ingress node) and an E2E disjoint path is used to send the traffic. This scheme provides global path recovery, but during the recovery process there are traffic losses. This is due to the fact that the failure notification has to be transmitted from the node detecting it (node 5) to the ingress node (node 1). To minimize data losses, local recovery, described in Figure 1 b), is used. In this case the path is not protected end-to-end. Instead, each link along the path is protected separately. A combination of these two mechanisms is referred to as Local-to-Egress (L2E), Figure 1 c). In multi-domain failure scenarios, where inter-domain link failures are considered, generally only the E2E and the L2E schemes can be applied due to the limited topological visibility of the border nodes. Local recovery is possible only if there are parallel links available between border nodes. Furthermore, applying L2E recovery does not require traffic merging capabilities in each intermediate node. Consequently, in restoration scenarios, computing a path to the egress node has more advantages, than using local recovery. Another reason to use L2E in restoration is the potentially shorter backup path, making better resource consumption.

2.2. Connection recovery process (recovery time)

In the recovery process there are different phases between the moment the failure occurs and when the working path is fully recovered [7, 11]. Each phase involves a method or a process.
The main components/phases are:

(a) A method for selecting the working and protection paths (routing algorithm).
(b) A method for signaling (setting up) the paths (for instance, LDP/RSVP-TE).
(c) A mechanism for fault detection.
(d) A hold off time (to avoid duplications in the recovery process).
(e) A fault notification method.
(f) A switchover mechanism to move traffic from the working path to the backup path.
(g) A repair detection mechanism (optional), to detect that a fault along a path has been repaired.
(h) A switchback mechanism (optional), for switching traffic back to the primary working path, once it is discovered that the fault has been corrected or has been repaired.

Each phase depends fully on the used survivability methods. Further details of these phases could be found in [6, 7, 8, 11]. In this paper we focus on the methods explained in the above section which involve phases (a),
Figure 2: Recovery Phases.

(b), (c), and (e). Figure 2 shows the recovery time process and the times (delays) associated with each phase. The notification time typically includes any queuing delays at the nodes, involved in the recovery of the path. Thus, the load in the network has specific effect on it. In order to minimize possible negative effects (unacceptable delay values), priority queues for the failure notification packets can be used.

3. Survivability in Multi-Domain Networks

Survivability in single-domain networks has been heavily investigated during the past years, but the multi-domain case has not received much attention. Even though it is widely considered that in the context of multi-domain restoration the resilience principles should not be significantly different to the single domain case, new relevant challenges appear.

A major problem comes from the limited visibility of the nodes regarding the multi-domain connectivity and the full network topology. The preservation of topological and state information within the domain is based on confidentiality preservation policies between domains. These policies have a very important implication: border nodes have information only about reachable destinations in the other domains, but no information about core or border nodes. Thus, unless there are parallel links between border nodes there is no possibility of applying local restoration techniques for failed inter-domain links.
On the other hand, there are no standards for multi-domain routing in GMPLS networks. Several approaches are being evaluated: the Border Gateway Protocol (BGP) [12], the Path Computation Element architecture (PCE) [13] and the E-NNI routing specification (only for intra-carrier application) [14]. BGP requires protocol re-convergence after the failure to obtain the restoration path, since BGP provides only one path per destination. PCE also includes delay for restoration path computation, providing constrained-based path computation to meet QoS requirements. Finally, the E-NNI approach also necessitates protocol re-convergence at different levels of the applied routing hierarchy. It is obvious that the slower the path computation is, the longer it takes to restore a failed connection. A possible solution relies on pre-computed disjoint backup paths. Both the PCE approach and some BGP extensions [15, 16, 17] offer solutions for obtaining such paths. In our work we adopt the principle of backup path pre-computation, where the backup path is computed before failure, but is signalled only after the failure has been detected.

Finally, the confidentiality preservation policies, regarding the network state, are also an obstacle since handling inter-domain link failures is still unclear in terms of how much information the domains are willing to share. For instance, information about neighbors regarding the inter-domain links and how far the failure notification should be propagated. A typical approach is to confine the failure notification only within the domains adjacent to the failed link [18].

4. New Multi-Domain Restoration Proposals

In this section two novel restoration techniques, which provide differentiated failure handling based on different criteria, are outlined. The operation of both schemes is detailed and analysis of the needed protocol extensions is given. Several challenges are outlined and some solutions are proposed. Our work is explicitly focused on the case of inter-domain link failures in connection-oriented multi-domain networks.

4.1. Location-Based restoration

In Location-Based restoration the node upstream the link failure must take a decision: either to restore the failed connection using the L2E technique or to signal the failure upstream to the head-end so that E2E restora-
tion can be applied\textsuperscript{1}. In order to take the most appropriate decision in our location-based restoration scheme a node must know its position in the overall path. We denote this scheme: Simple Location-Based Restoration (SLBR). The delay for recovery can be expressed in terms of the distance $D(i, j)$ between the node detecting the failure $i$ and the source/destination node $j$. Since the involved calculations are not complex, the delay for setting up a restoration connection is dominated by the propagation delay for the signaling packets (failure notification packets and RSVP-TE PATH messages). Thus, the distance between $i$ and $j$ can be expressed in terms of number of hops. The SLBR technique requires the following decision to be taken at the node detecting the failure:

- **SLBR**
  
  \( \text{if } D(i, d) < D(i, s) \text{ apply } L2E, \)
  \( \text{else if } D(i, d) > D(i, s) \text{ apply } E2E, \)
  \( \text{else apply random selection,} \)

  where $s$ denotes the source node, $d$ the destination node and $i$ is the node upstream the link failure.

4.2. **Shortest New restoration (SN)**

A typical objective in transport networks is improved resource utilization. In order to minimize the additional resources consumed during restoration, compared to the resources consumed for the working path, we base our restoration decision on the amount of potentially consumed resources. In a wavelength-division multiplexed (WDM) network, the considered resources are the wavelengths. The total resource consumption of a LSP is the number of links it traverses. This can be expressed in terms of number of hops. The decision, taken in the node detecting the failure, can be expressed as:

- **SN**
  
  \( \text{if } D(s, i) + D_{res}(i, d) > D_{res}(s, d) \text{ apply } E2E, \)
  \( \text{else apply } L2E, \)

\textsuperscript{1}See section 2 for explanation, why local restoration has restricted application for inter-domain link failure in multi-domain networks.
where $i$ indicates the node, detecting the failure, $s$ indicates the source node, $d$ indicates the destination node, $D(i, s)$ is the distance in hops from $i$ to $s$, $D_{res}$ is the length of the restoration path in hops.

If the node detecting the failure takes the decision to apply E2E, then the failure notification is propagated all the way back to the head-end of the connection where an E2E restoration is applied. The required information for taking the decision in case of a failure can be provided using extended routing protocols and piggybacking the existing provisioning protocols. The following subsection discusses the challenges of the proposed mechanisms and possible solutions.

### 4.3. Challenges and solutions

As presented above, applying the proposed restoration mechanisms require the distribution of additional information and the application of specific path computation methods. The first issue to be solved is related to the used multi-domain routing protocol. If a standard BGP is used for multi-domain routing, the node detecting the failure must wait until the protocol re-converges to a stable solution in order to use a new path to restore the affected LSP. The PCE architecture requires either a minimum of two parallel connections between domains [19], or an additional path computation delay if the Backward Recursive Path Computation technique [20] is used dynamically. In both cases (standard BGP or PCE) the restoration time could be very long. In order to avoid this additional delay, and not to bind the network topology to have double connections between domains, a modified BGP protocol can be used, which provides more than one path per destination beforehand. Several solutions for multi-path dissemination in BGP networks exist [15, 17]. Another solution is the application of the E-NNI specification [14], but this requires OSPF-TE to be deployed in all domains. Whatever the chosen mechanism for multi-domain routing is, each border and source node is assumed to have at least two disjoint paths per destination. In this way, the additional delay for restoration path computation ($T_{BPC}$ in Figure 2) will be avoided and the recovery time will mainly depend on the length of the restoration path and the distance from the failure to the point of recovery. The authors of [16] present a practical example of using AS-disjoint paths, computed via modified BGP protocol, for survivability provisioning in WDM networks.
Figure 3: Example of obtaining the exact position of a node along the LSP path.

The second issue to be addressed is how to obtain the information, regarding the actual position of a border node along the overall path of the LSP. Several approaches can be taken. Figure 3 illustrates two of them. The first option is using the Record Route Object (RRO) during the path setup. The RRO carries the full path from the source and is updated on the way at each node, processing the RSVP-TE PATH message. It is also included in the RESV message, indicating the whole path from source to destination. Using the information from the RRO, each node can calculate its relative position in terms of hops to the source and to the destination. This information is stored in the border node per established LSP and is used in case a link attached to the node fails. A possible drawback is the confidentiality preservation requirements between domains according to which no topological information should leak out from the domain borders. The RRO object carries sensitive topological information and is typically either erased, when crossing borders, or encoded [21]. In such a case, a simple hop counter can be used as depicted in Figure 3, which keeps track of the number of nodes and can be added to the PATH and RESV messages of RSVP-TE [22]. An alternative approach is to use time-stamps, but this necessitates time synchronization between the participating domains.

4.4. Operation example

In our implementation we use the modified BGP for obtaining two AS-disjoint paths per destination presented in [16]. Furthermore, a hop counter in the RSVP-TE messages for identifying the actual position of each border
node along a path is used. Figure 4 illustrates the operation of the proposed SLBR scheme. Two options for the restoration of the affected LSP can be applied, E2E or L2E (i.e. Local-to-Destination). It can be seen that if E2E restoration is applied, the setup delay is proportional to 22 hops, and if L2E is applied - only 12 hops. Thus, if the proposed SLBR mechanism is applied, the affected LSPs receive differentiated handling and would be restored with a mechanism which seeks to minimize the recovery time.

Figure 4: Example of the operation of the SLBR mechanism.

5. Simulations

In order to evaluate the efficiency of the proposed schemes we have implemented them in an event-driven simulator [23]. We tested two different Pan-European networks namely COST 266 [24] and NOBEL (see Figure 5). RSVP-TE is used as a resource reservation protocol and a modified BGP as multi-domain routing protocol. The modifications in BGP allow it to automatically discover two AS-disjoint paths to each destination. We investigate
the performance of two standard restoration techniques: Local-to-End (L2E) and End-to-End (E2E), and the two novel proposals: Simple Location-based Restoration (SLBR) and Shortest New (SN). Three performance metrics are investigated:

- Resource Overbuild - indicating the additional resources needed to recover affected connections, compared to the resources needed for the working paths.
- Setup delay - indicating the average time for LSP recovery.
- Recovery success ratio - indicating the ratio between the number of successfully restored LSPs and the number of affected LSPs.

![Network topologies](image)

(a) COST 266 network topology.  
(b) NOBEL network topology.

Figure 5: Network topologies.

The simulation setup is as follows. We model a WDM network. Single optical fiber with 150 wavelength channels is used between nodes. The domain boundaries follow the borders of the countries, with up to 4 source/destination nodes per domain. Link propagation delays depend on the geographical distance between nodes. Traffic is uniformly distributed between all source/destination nodes; the LSP request inter-arrival time is exponentially distributed. All requests have average duration of 600sec. No wavelength conversion is applied in the network during LSP setup. Wavelength assignment is random and is carried out at the destination node based on the label
set supplied by the RSVP-TE protocol. Each signaling packet is subject to 1ms delay for processing per node.

The restoration setup delay is calculated as the difference between the time when the failure occurs and the time the RSVP-TE RESV message is received at the point of repair (Figure 2). If the point of repair is the head-end of the connection, this time includes the time to notify the node of the failure. The setup delay is an averaged result among all successfully restored connections.

Two simulation scenarios are presented: with and without wavelength conversion at the point of repair. Since head-ends can choose from among all their unoccupied resources for recovery, in order to carry out a fair comparison we introduce limited wavelength conversion at the points of repair when local restoration is performed. If no wavelength conversion is allowed, the head-ends of the connections use the same wavelength for the backup path as was used for the workings path. In both scenarios, the same set of links were failed per examined topology.

5.1. Limited wavelength conversion scenario

Figure 6 and Figure 7 illustrate the recovery success ratio, the setup delays, and the resource overbuild for inter-domain link failures in the COST 266 and the NOBEL networks respectively.

Considering the recovery success ratio (subfigures a)), it can be seen that the behavior of all schemes is similar in both topologies with the L2E providing, on average, better recovery success ratio than the E2E scheme, especially at high loads. This is due to the fact that at high loads, the probability for blocking on paths with more hops is higher. Since the Pan-European networks are well-connected\(^2\), the backup paths (both local and global) are not much more longer than the working paths\(^3\). Thus, recovering from the head-end of the connection yields higher blocking. The SLBR outperforms the rest of the schemes because it distributes the connections to be restored throughout the network. Some are restored locally, others from end-to-end. This balances the traffic in the network and consequently, the probability for successful restoration increases.

\(^2\)The average domain connectivity is above 3.15 with domain Germany having connectivity of 8.

\(^3\)The used multi-domain routing protocol is BGP, which chooses paths based on the number of ASes on the way, not on the number of actual nodes.
In terms of setup delay (subfigures b)), the L2E outperforms all mechanisms mainly because of the well-connected topology of the tested networks, their hierarchical structure (they both have an effective core around domain Germany) and the fact that the failed links have been chosen to be in the middle of the network in order for the failure to have the highest impact on the traffic. Furthermore, with the increase of the load, the setup delay decreases, which is due to the fact that connections, which require more resources (i.e. connections with long recovery paths) have lower probability to be recovered.

With respect to the resource overbuild (subfigures c)), it can be seen that
the SN mechanism indeed provides the lowest overbuild. Furthermore, in synchrony with the already observed results for the setup delays and the recovery success ratio, the higher the load in the network, the lower the resource overbuild is, which is due to the fact that only connections with short recovery paths get restored.

Taking into consideration the overall performance of the proposed restoration techniques it can be seen that the SLBR and the SN achieve a good balance between the main performance metrics. L2E provides the fastest restoration, but suffers from increased resource overbuild. E2E, on the other hand, provides low resource overbuild, but has the highest setup delay. Since
the novel mechanisms implement E2E and L2E in a differentiated fashion, their performance balances the usual tradeoff between resource utilization and setup delay. Furthermore, implementing the rather naive SLBR approach splits the restoration traffic in the network, which increases the probability for successful restoration.

Figure 8 illustrates the percentage of occasions where end-to-end restoration is applied under the novel mechanisms. As it can be seen, for the COST 266 network almost in all cases the SN performed end-to-end restoration. This explains the close similarity in the behavior of the SN and the E2E mechanisms for the COST 266 network (see Figure 6). In the rest of the cases, the mechanisms employ end-to-end restoration approximately in half of the cases, which explains why they achieve a good balance between the performance metrics.

Figure 8: Percentage of occasions end-to-end recovery applied within the SLBR and the SN schemes.

5.2. No wavelength conversion scenario

Allowing limited wavelength conversion for restoration results in better performance of the restoration techniques, compared to the case where no wavelength conversion is allowed. Disabling the wavelength conversion at restoration significantly decreases the probability for successful recovery as can be seen from Figure 9 and Figure 10, which illustrate the performance of all restoration schemes for the COST 266 and the NOBEL networks.
Figure 9: Performance results for COST 266 network without wavelength conversion.
Figure 10: Performance results for NOBEL network without wavelength conversion.
In this scenario, head-end restoration becomes very difficult. Thus, the performance of the L2E restoration is the best in terms of recovery success ratio. As observed earlier, the setup delays are shorter with L2E restoration, which is also observed here. With respect to the resource overbuild, an interesting result is observed for the COST 266 network. Both E2E and the SN mechanisms experience negative resource overbuild. This can be explained by the following facts. First, as was outlined earlier, the used multi-domain routing protocol chooses paths based on AS count, not on actual hop count. Thus, some working paths which pass via large domains with several border nodes have higher hop count (node-wise) than their respective backup paths, which pass via smaller domains with only one border node. For example, the path from the source node connected to Bordeaux to the destination node connected to Prague passes via Germany. The full paths is 8 node hops (2 AS hops): Bordeaux - Paris - Strasbourg - Frankfurt - Munich - Berlin - Prague - destination. The backup paths is 7 node hops (4 AS hops): Bordeaux - Marseille - Rome - Zagreb - Vienna - Prague - destination. Second, since the requirement for using the same wavelength for the backup path as was used for the working path is very limiting, only connections with short backup paths are successfully restored. As a result, the resource overbuild for the successfully recovered connections is negative. This result is topology specific and depends on the failed links.

6. Conclusions

In this paper the performance of restoration mechanisms in multi-domain networks is analyzed and evaluated. End-to-End and Local-to-End, together with two novel proposed mechanisms, SLBR and SN, have been tested in two different topologies. When wavelength conversion is allowed, all mechanisms have improved performance in terms of Recovery success ratio, because the strong limitation of the used wavelength is relaxed. The setup delay increases due to the successful recovery of connections with longer backup paths, which leads to increased resource overbuild. The novel techniques, SLBR and SN, report a good balance between the different performance metrics with SN providing the lowest resource overbuild. With no wavelength conversion, L2E performs better in terms of recovery success but the resource overbuild is the worst, whereas E2E presents the opposite behavior. The Recovery success ratio decreases due to limited resource availability. The setup delay decreases because connections requiring long backup paths experience higher
blocking probability. As a consequence, the resource overbuild also decreases. Nevertheless, the proposed SLBR and SN mechanisms still provide a good balance between the performance metrics, particularly with respect to the resource overbuild.

In summary, results have shown that the evaluated mechanisms present different performance depending on the topology, the load and the location of the failure. Employing mechanisms which can provide differentiated failure handling is beneficial for extending the service portfolio of network operators, serving clients with diverse requirements. Network operators could take into account restoration as an alternative resiliency mechanism, in order to design highly flexible next generation multi-domain networks.

7. Acknowledgments

Part of this work was supported by the Government of Catalonia (DURSI SGR-1202 and 2007-FIR-00836), and by the Spanish Science and Innovation Ministry (TRION TEC2009 - 10724).

8. References


