



# Evaluation of post-fault restoration strategies in multi-domain networks

Feng Xu<sup>a</sup>, Tamal Das<sup>b</sup>, Min Peng<sup>c</sup>, Nasir Ghani<sup>a,\*</sup>

<sup>a</sup> University of New Mexico, 87131-0001 Albuquerque, NM, United States

<sup>b</sup> IIT Bombay, India

<sup>c</sup> Wuhan University, China

## ARTICLE INFO

### Article history:

Available online 4 December 2011

### Keywords:

Multi-domain network  
Network survivability  
Multi-failure restoration  
Signaling crankback  
Hierarchical routing

## ABSTRACT

Although multi-domain survivability is a major concern for operators, few studies have considered the design of post-fault restoration schemes. This paper proposes two such strategies, based upon hierarchical routing and signaling crankback, to handle single and multi-link failure events in multi-domain IP/MPLS networks (also extendible to optical DWDM networks). Specifically, the hierarchical routing strategy relies upon abstracted domain information to compute inter-domain loose routes, whereas the crankback scheme applies signaling re-tries to restore paths in a domain-by-domain manner. The performance of these proposed solutions is then analyzed and compared via simulation.

© 2011 Elsevier B.V. All rights reserved.

## 1. Introduction

Network survivability remains a major concern for operators, particularly as expanding user and business applications drive increasingly more traffic across longer ranges. As is well-known, a host of pre-planned *protection* and post-fault *restoration* recovery schemes have been developed for IP *multi-protocol label switching* (MPLS) and optical *generalized MPLS* (GMPLS) networks [1]. In the former types, backup paths are pre-computed and their resources reserved in advance for rapid recovery, i.e., fast switching to a backup path if failure occurs on the primary one. Such pre-provisioned techniques are generally capable of ensuring recovery against single link failures. However, these schemes require the allocation of resources regardless of whether or not there is actually a failure, thereby consuming more resources and overheads. On the other hand, post-fault restoration schemes use active re-routing procedures to set up new working paths after faults [2]. Although these alternative strategies have longer restoration delays (hundreds of ms to seconds

range) and cannot guarantee recovery in all cases [2,3], they are much more resource efficient and well suited for handling multi-failure situations.

Nevertheless, most existing studies on network survivability have focused on single-domain settings [3]. By contrast, broader *multi-domain* scenarios are much more challenging as “global” topology information is lacking and only selected nodes may have partial views of the inter-domain state, i.e., due to the obvious scalability and confidentiality concerns [2]. As a result, multi-domain survivability is starting to see increased attention in recent years [4–11]. Here, almost all of these studies have focused on developing some form of pre-provisioned protection for primary/backup paths traversing across multiple domains. For example, some have considered extending SONET/SDH interconnection strategies to achieve more “localized” dual/multi-homing link protection between border nodes in optical *dense wavelength division multiplexing* (DWDM) multi-domain networks [4]. Meanwhile, some other researchers have proposed more elaborate path computation and setup signaling strategies to compute link-disjoint primary/backup routes across either identical [5–7] or different [8,9] domain sequences. Alternative static protection mechanisms, including *pre-configured cycles* (*p-cycle*) [10] and Hamiltonian cycles [11], have also been studied. How-

\* Corresponding author. Tel.: +1 505 277 1475.

E-mail address: [ghanin@yahoo.com](mailto:ghanin@yahoo.com) (N. Ghani).

ever, multi-domain protection schemes can only guarantee recovery from single link failure events and hence the performance of these schemes would be severely compromised when more than one link fails simultaneously, e.g., as may be the case during larger natural or man-made disaster events.

As a result, it is imperative to develop more adaptive multi-domain recovery solutions for various failure scenarios. Indeed, dynamic post-fault restoration strategies have much potential here to handle randomized failures, particularly those affecting multiple links and/or nodes. Even though restoration schemes may exhibit more latent recovery, they are well suited for current data-centric services with more relaxed recovery times/priority [12], e.g., *virtual private network* (VPN). More importantly, since restoration is more resource efficient than pre-provisioned protection, it can offer a very viable “last-gap” alternative when building tiered survivability frameworks. Therefore, two different multi-domain post-fault restoration schemes are studied here, based upon *hierarchical routing* and *signaling crankback* strategies, respectively. In particular, the hierarchical routing strategy uses the limited “global” state obtained from *topology abstraction* to compute inter-domain *loose routes* (LR) and then expands the restored paths along with signaling. Conversely, the signaling crankback scheme is an enhanced version of that in [13] and performs path computation in a domain-to-domain manner to avoid resource-deficient links. Overall, both of these solutions can be applied in either single-link-failure or multi-failure scenarios.

The rest of this paper is organized as follows. Section 2 first surveys the latest work in network survivability. Next, Section 3 details the proposed restoration solutions using hierarchical routing and signaling crankback. Performance analysis results are then presented in Section 4 for different failure scenarios using discrete simulation, followed by conclusions and future directions in Section 5.

## 2. Background

As noted earlier, a handful of post-fault restoration solutions have been designed for single-domain networks, i.e., where a centralized controller has full visibility of the network and is able to independently compute the restored paths for failed connections. For example, three restoration schemes are introduced in [12], i.e., path restoration, sub-path restoration, and link restoration. In *path restoration*, the source node of each failed connection tries to re-establish the route via a new end-to-end path that bypasses the failed link towards the destination node (upon receiving a failure notification message). In *sub-path restoration* and *link restoration*, the upstream end-point node of the failed link tries to re-route the “sub-path” from itself towards the destination node or the downstream end-point node of the failed link (without notifying the source node). Although these basic restoration strategies can be effectively applied to multi-domain settings in some form, i.e., to be coupled with appropriate provisioning strategies, detailed studies are still lacking. Along these lines, the work in [13] presents an initial study on multi-domain restoration for the case of single inter-domain link

failures. This work is adapted in this paper and further applied for multi-failure restoration (detailed in Section 3).

Meanwhile, a range of multi-domain pre-protection schemes have also been developed. For example, [5–7] present studies on multi-domain protection strategies where the working and backup paths use the same domain sequence, i.e., “per-domain”. In particular, [5] analyzes both distributed path computation and setup signaling issues and then proposes an enhanced sequential signaling scheme to compute diverse path-pairs along the given domain sequence. By contrast, [8] attempts to achieve improved “path diversity” by allowing primary/backup paths to traverse different end-to-end domain sequences. Specifically, this scheme applies full-mesh abstraction to extract critical path-level diversity state based upon Suurballe’s algorithm [14] and build a global view. A distributed path computation algorithm is then proposed to generate link-disjoint LR-pairs, and it is formally shown to be able to find optimal end-to-end path-pairs. However, simulation results are not presented and the abstraction overheads are prohibitively high, i.e.,  $O(N^4)$  for  $N$  border nodes in a domain. Note that some have also considered shared protection for multi-domain network settings using topology abstraction schemes [9]. However, the likelihood of implementing detailed resource sharing across multiple domains, particularly those operated by different carriers, will be generally quite low. Other schemes using p-cycle and Hamiltonian cycle protection have also been extended to multi-domain environments [10,11]. These algorithms basically treat intra- and inter-domain resilience separately by routing pre-defined recovery cycles to achieve more efficient resource utilization. However, these approaches are only specialized for single link failures and have high computational complexities.

Finally, various studies have also investigated multi-failure recovery, mostly in single-domain settings. For example, earlier efforts have looked at specialized *dual near-simultaneous* failures. An example is the work in [15] which develops a *backup re-provisioning* (BR) scheme for DWDM networks to periodically re-compute failed protection lightpaths before a second failure event. This approach gives notable improvement in backup availability but entails high computational overheads. Meanwhile, [16] presents several shared protection schemes to handle dual-link failures, whereas [17] compares post-fault restoration versus pre-provisioned protection for dual-link DWDM failures, showing improved recovery with the former strategy. In addition, the authors in [18] have studied dual link failure recovery in IP-tunneling networks and provisioned fast recovery (for the first failure) via protection routes that can inherently sustain a second link failure. Clearly this assumes some prior knowledge of the location of the second failure. More recent efforts have also looked at survivability for probabilistic multi-failure conditions. For instance, [19] develops a path computation scheme for generalized random failures and builds upon two notions of reliability, i.e., *local* and *global*. The former chooses paths with the lowest number of failure events, whereas the latter selects paths so that a failure affects a minimum number of connections. Moreover, [20] studies correlated and probabilistic failures and tries to maximize

the reliability of pre-computed protection routes. The existing *shared risk link group* (SRLG) definition [3] is further extended here to a more general *probabilistic SRLG* (p-SRLG) concept as well. Nevertheless, none of these dual-/multi-failure recovery schemes have considered the broader and more complex multi-domain setting, as is the focus herein.

In light of the above, there is still significant scope to develop multi-domain restoration-based solutions using “active” post-fault path re-computation to minimize dependence upon pre-provisioned backup schemes. The applicability of these techniques to multi-failure recovery is particularly relevant. Along these lines, two proposed restoration strategies are now presented and analyzed, i.e., using hierarchical routing and signaling crankback.

### 3. Multi-domain restoration solutions

Overall, both hierarchical routing and signaling crankback strategies have been studied for *working-mode* provisioning in multi-domain networks; see [21] and [22]. These setups assume the availability of full intra-domain link-state at interior nodes, e.g., via *open shortest path first-traffic engineering* (OSPF-TE) protocol, and some form of limited inter-domain visibility at selected border nodes as well. Namely, hierarchical inter-domain routing implements a second level of link-state routing between border nodes and achieves a partial “skeleton” view of global domain/link resources. Meanwhile, signaling crankback is a simpler approach in that it operates with very limited path-vector state, as provided by the *border gateway protocol* (BGP) [3]. In both schemes, however, setup signaling is done using the *resource reservation-TE* (RSVP-TE) protocol.

The application of the above strategies for post-fault recovery is now detailed. Carefully note that only generalized bandwidth-provisioning networks are considered here, although the proposed strategies can easily be tailored and adapted for optical wavelength-selective networks. Before introducing the schemes, the requisite notation is first defined. In general, a multi-domain network, comprised of  $D$  domains with the  $i$ -th domain having  $n^i$  nodes and  $b^i$  border/gateway nodes,  $1 \leq i \leq D$ , is modeled as a set of domain sub-graphs,  $G^i(V^i, L^i)$ , where  $V^i = \{v_1^i, v_2^i, \dots\}$  is the set of domain nodes and  $L^i = \{l_{jk}^{ii}\}$  is the set of intra-domain links in domain  $i$  ( $1 \leq i \leq D$ ,  $1 \leq j, k \leq n^i$ ), i.e.,  $l_{jk}^{ii}$  is the link from  $v_j^i$  to  $v_k^i$  with available and maximum capacities  $c_{jk}^{ii}$  and  $C_{jk}^{ii}$ , respectively. Meanwhile, the set of border nodes in domain  $i$  is also given by  $B^i$ . Next, each domain is assumed to have a *path computation element* (PCE) with full access to the interior topology graph [21,22]. The *inter-domain* link between border node  $v_k^i$  in domain  $i$  and  $v_m^j$  in domain  $j$  is further denoted as  $l_{km}^{ij}$ , with available and maximum capacities  $c_{km}^{ij}$  and  $C_{km}^{ij}$ , respectively,  $1 \leq i, j \leq D$ ,  $1 \leq k \leq b^i$ ,  $1 \leq m \leq b^j$ . Each node also maintains a list of traversing connections,  $A_j^i$  for node  $v_j^i$ , where each entry in  $A_j^i$  is a route vector. In addition, the RSVP-TE message fields include a path route vector,  $R$ , and an exclude link vector,  $X$ , to track the congested or failed links. Finally, dual intra/inter-domain counters, i.e.,  $h_1$  and  $h_2$ , are defined for signaling crankback.

#### 3.1. Restoration with hierarchical routing

As mentioned above, hierarchical routing uses topology abstraction to summarize domain-level state. Namely, designated *routing area leader* (RAL) nodes, possibly co-located with the PCE, run specialized graph transformation algorithms to reduce physical domain-level topology/resource graphs, i.e.,  $G^i(V^i, L^i)$ , into compressed topologies. This abstracted state is then flooded between border nodes of all domains, allowing them to build “global” network views and perform inter-domain path computation. In particular, all inter-domain (and intra-domain) link state updates use relative threshold triggering policies, as these are shown to give the best state timeliness [21]. Added hold-off timer mechanisms are also used to prevent excessive flooding [23]. As per [21], two specific topology abstraction schemes are used (and illustrated in Fig. 1):

*Simple-node abstraction.* This scheme reduces each domain to a virtual node emanating just the physical inter-domain links for that domain. This scheme has lower routing overhead as updates are only sent for the actual physical inter-domain links, i.e.,  $l_{km}^{ij}$ ,  $i \neq j$ . However, simple-node abstraction provides no domain-internal visibility.

*Full-mesh abstraction.* This scheme computes/advertises “abstract links” between all domain border nodes in addition to the updates for the physical inter-domain link. The goal is to provide a summarization of domain traversal costs via abstract links  $l_{jk}^{ii}$ ; see [21] for algorithmic details. As expected, full-mesh abstraction has notably higher computational/routing complexities as it generates  $O(|B^i|^2)$  abstract links for domain  $i$  at the inter-domain level.

Using the above formulations, the domain-level RAL/PCE entities in the hierarchical routing setup compute “skeleton path” sequences for inter-domain requests. Namely, these LRs are generated by running modified shortest path algorithms over the “global” topology/resource graphs. Again, leveraging from [21], two different LR path computation approaches are studied here, i.e., *minimum hop* and *minimum relative distance*. Specifically, each physical/abstract link in the inter-domain network graph is assigned with a “cost”,  $\omega_{kmn}^{ij}$ , as follows:

*Minimum hop:*  $\omega_{kmn}^{ij} = 1$ .

*Minimum relative distance:*  $\omega_{kmn}^{ij} = 1/(u \cdot c_{kmn}^{ij}/C_{kmn}^{ij})$ , where  $c$  and  $C$  are the unreserved and maximum link capacities, respectively, and  $u$  is a constant. The former metric tries to achieve (inter-domain) resource minimization whereas the latter aims for load-balancing. Finally, if the above LR computation is successful, full *explicit route* (ER) expansion is attempted along the LR sequence using RSVP-TE signaling.

Now consider the application of hierarchical routing strategy to post-fault restoration, i.e., after an intra- or inter-domain link failure. Here it is assumed that the link end-points can quickly discover the failure via rapid lower layer fault-detection mechanisms [24] (which are deemed out of scope herein). These endpoint nodes then loop through their active connection lists,  $A_j^i$ , to search for any

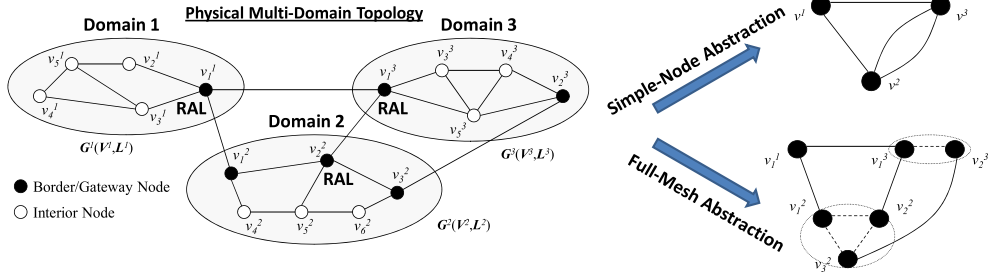


Fig. 1. An example of simple-node and full-mesh abstractions from a multi-domain network.

traversing connections that are using the failed link. Here, all such connections are removed from  $\mathbf{A}_j^i$  and appropriate restoration procedures initiated for each. Specifically, the detecting nodes first transmit resource takedown messages along the failed upstream and downstream path segments. Next, each node on the upstream side of the failed link notifies the source of the connection via a notification message with the failed link noted in exclude route vector,  $\mathbf{X}$ . This message also sets an appropriate restoration flag to indicate that this is a failed connection.

Upon receiving the above failure notifications from downstream nodes, the source nodes query their respective PCEs to trigger (post-fault) hierarchical inter-domain path re-computation and subsequent re-routing. Taking into account the dissemination delays of global link state (via hold-off timers), the source domain PCE extracts the failed inter-domain link from  $\mathbf{X}$  and prunes it from its inter-domain graph before re-computing a revised LR. The subsequent setup procedures (i.e., signaling expansion) are the same as described for working-mode operation. Overall, hierarchical routing restoration is shown in the top part of Fig. 2, where a failure on link  $l_{22}^{56}$  between domains 5 and 6 causes the endpoint nodes to issue takedown sequences to free resources along the failed paths. The source node then initiates ER signaling expansion along the least-cost LR computed via the abstracted topology state and the restored path is successfully signaled and set up via domains 2, 7 and 4.

### 3.2. Restoration with enhanced signaling crankback

Recently, some (working-mode) signaling crankback schemes have also been studied for multi-domain networks, leveraging related RSVP-TE extensions in RFC 4920 [25]. In general, these strategies assume minimal inter-domain state knowledge and rely upon distributed “per-domain” computation. Namely, domain PCEs iteratively select the “next-hop” domain and then initiate intra-domain ER towards the specific border node/egress link. Path computation concludes when the fully expanded route sequence reaches the destination node.

However, as resource deficiencies can arise during signaling setup, various crankback procedures have been defined. In particular, this work leverages earlier contributions by the authors in [22] on developing a joint intra/inter-domain crankback scheme. Consider the working-mode process to begin with. This solution limits the number of intra/inter-domain crankback attempts

to  $H_1$  and  $H_2$ , respectively, and implements two key steps, i.e., *crankbacknotification* and *re-computation*. Specifically, the first node detecting a resource-deficient link sends an RSVP *PATH\_ERR* message to the ingress border node of its own domain. Here the intra-/inter-domain crankback counters values are set to  $h_1 = H_1$  and  $h_2 = H_2$  and the problematic link noted in the exclude link vector,  $\mathbf{X}$ . Upon receiving this *PATH\_ERR* notification, the ingress border node performs *re-computation*. Namely, it decrements  $h_1$  and, if it is not zero, selects another egress border node for ER expansion, i.e., intra-domain crankback. Alternatively, if  $h_1$  expires, inter-domain crankback is done by sending a *PATH\_ERR* message to the ingress node in the upstream domain (and  $h_2$  decremented). The connection request is considered ultimately failed if both counters expire. The work in [22] also adds further intelligence to improve next-hop domain selection, i.e., to avoid random/exhaustive searches. Namely, each PCE maintains a static, pre-computed multi-entry *distance vector table* that lists up to  $K$  next-hop domains/egress links to each destination domain. This table can be generated using a limited inter-domain path state, e.g., from BGP tables; see details in [22].

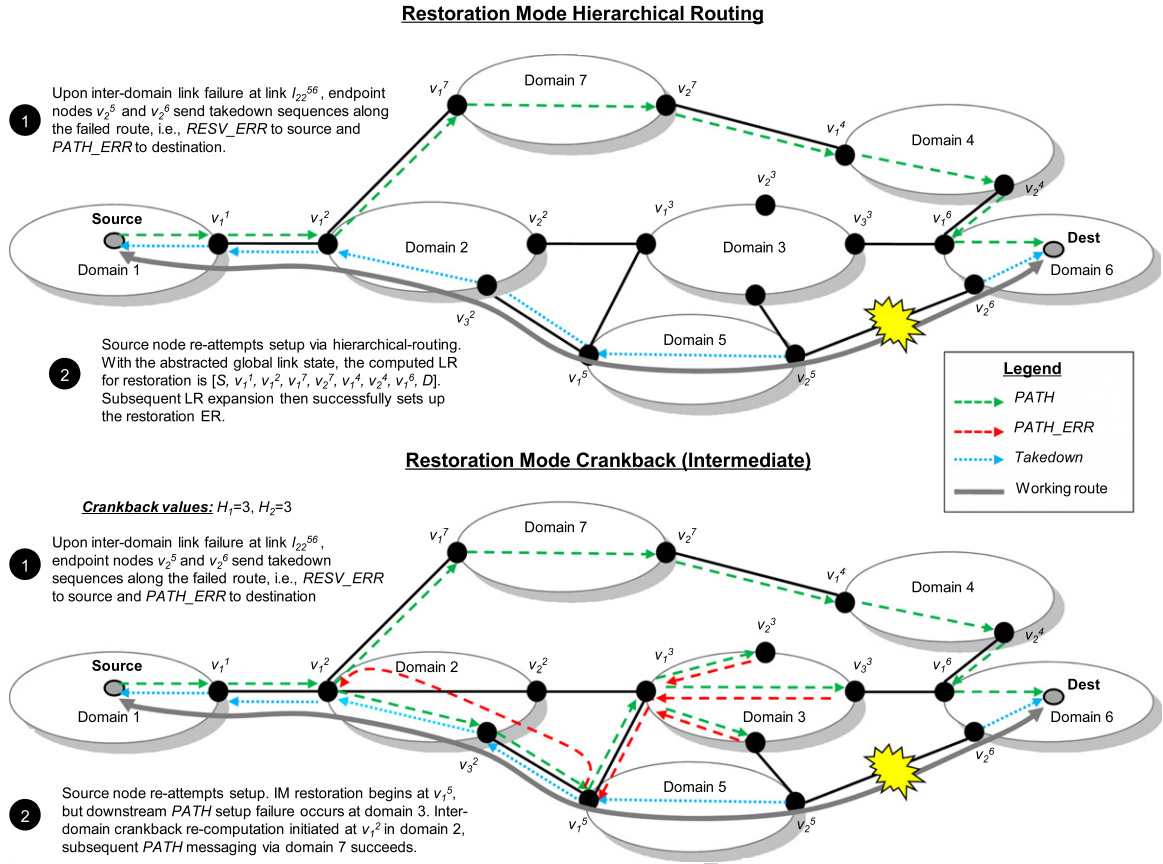
Now consider the extension of this multi-domain crankback strategy for post-fault restoration, as initially studied in [13]. Again, it is assumed that failure end-point nodes can quickly detect faults and issue notification and takedown sequences (Section 3.1). Upon receiving these notifications, source nodes can pursue one of two different restoration approaches:

**End-to-end (E2E).** The source node simply queries its PCE for another next-hop domain and sends a new *PATH* setup sequence towards the destination, i.e., clear route vector  $\mathbf{R}$ , reset crankback counters  $h_1$  and  $h_2$ , and note failed links to outgoing exclude route list,  $\mathbf{X}$ . This request is then processed as per the regular working-mode setup procedures.

**Intermediate (IM).** The source node tries to preserve as much of the original path as possible, i.e., up to the domain preceding the failure. Hence the re-issued *PATH* message only prunes the failed connection route up to this domain and then initiates crankback processing from there, i.e. using same the sequence as the above-detailed working-mode setup.

To further improve restoration performance, a twofold congested link status tracking feature is added. Namely, in addition to using the exclude link vector,  $\mathbf{X}$ , i.e., to avoid re-selecting the resource-deficient links in the subsequent





**Fig. 2.** Restoration processes with hierarchical routing and signaling crankback (intermediate).

crankback re-tries, each border node  $v_j^i$  also maintains a status tracking table,  $T_j^i$ , to record the “congested” egress links for that domain. Specifically, when an egress border node detects that an outgoing inter-domain link has no remaining bandwidth, it checks to see which of its active connections in set  $A_j^i$  use this link. It then selects a connection from this subset with the earliest stop-time and inserts this time value into the corresponding *PATH\_ERR*, *RESV\_ERR* or *RESV* message (depending upon the stage of the setup process) to notify its upstream domain ingress border node. The ingress border node (receiving this information) then updates its entries in  $T_j^i$  and prevents all further attempts to that particular congested egress link until the noted stop-time passes by. Overall, this scheme leverages information from both working connection setups and restoration procedures to improve crankback efficiency.

Finally, a random back-off timer is also introduced to intentionally delay the restoration of each connection. Specifically, since multiple working connections can be affected by a single link failure, i.e., during high load intervals and/or on bottleneck inter-domain links, race conditions can easily arise if multiple failed connections attempt re-routing at the same time. To mitigate the effects of resource contention, random back-off timers are used to stagger crankback restoration procedures. As an aside, note that the associated back-off intervals are generally much

shorter than inter-domain routing update timescales, and hence may not favor restoration with hierarchical routing, i.e., restored LRs will still likely be computed with “dated” topology information. This issue is addressed further in Section 3.3.

The case of IM post-fault crankback restoration is shown in the lower part of Fig. 2, where the section of the original path from the source node to  $v_1^5$  is preserved at first. After  $H_1 = 3$  failed intra-domain crankback attempts in domain 3, inter-domain crankback is performed and finally the restored path succeeds via domains 2, 7 and 4 to the destination node.

### 3.3. Restoration in multi-failure scenarios

One of the most important advantages of post-fault restoration strategies over pre-provisioned protection is its ability to recover from multiple random link/node failures. Namely, restoration schemes (can be coupled with either hierarchical routing or signaling crankback) can be extended to handle multi-failure scenarios. For example, consider a multi-failure event that causes widespread link/node failures across a whole region. A source node (or intermediate upstream node) of each affected connection will receive an appropriate notification message from a downstream node (at the edge of the failure region) notifying it about the failure occurrence. This source

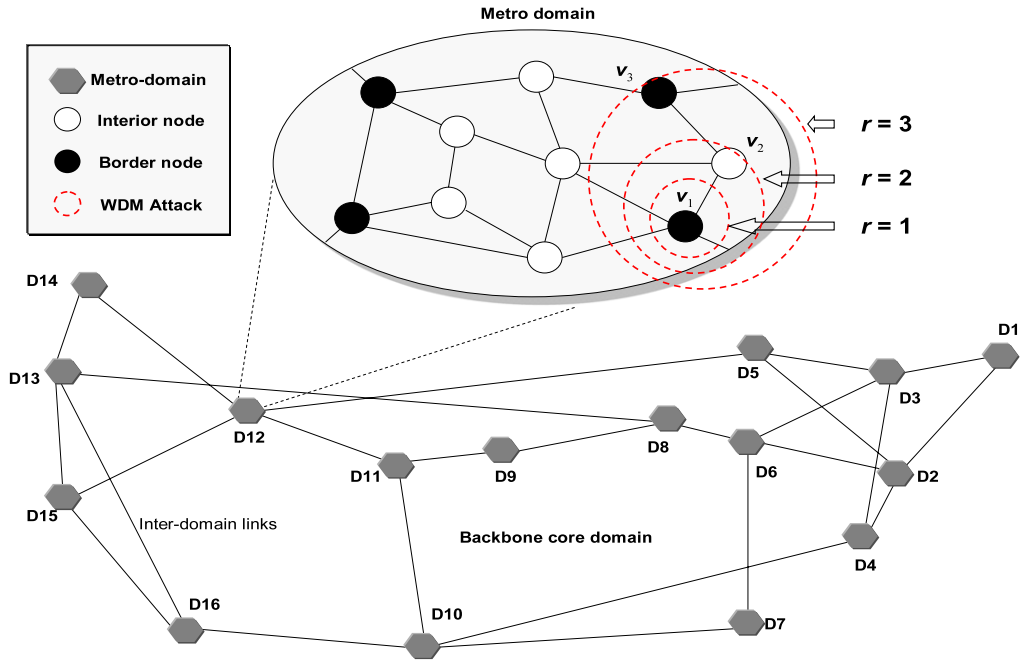


Fig. 3. Modified 16-domain NSFNET topology for single-/multi-failure tests.

(intermediate) node can then initiate appropriate restoration signaling procedures. Specifically, both E2E and IM restoration schemes can be applied here.

However, it is noted here that very large failure regions can seriously compromise the success rate of restoration with hierarchical routing, i.e., as the updates for many failed links may not make it to the appropriate intra-/inter-domain routing databases in time. Hence in order to provide the PCE with most accurate state information on failed links, the routing update/dissemination delays across the network are taken into account when selecting the back-off timer intervals. Specifically, computation of restored LR sequences is not started until the back-off timer expires. Although this approach can lessen the need to rapidly disseminate link failure statuses prior to path re-computation, recovery is delayed and race conditions can still arise in the case of simultaneous restoration attempts after the back-off timer interval.

#### 4. Performance evaluation

The multi-domain post-fault restoration schemes are tested using specially developed models in *OPNET Modeler*<sup>TM</sup>. All tests are done using the network shown in Fig. 3, reflecting a modified NSFNET topology, i.e., replacing original nodes with complete domains (16 domains, 25 bi-directional inter-domain links). In this topology, all intra- and inter-domain link rates are fixed at 10 Gbps and the individual domain sizes average between 7–10 nodes and correspond to metropolitan areas. Overall, the NSFNET topology is chosen as it is representative of a nationwide backbone and widely used in many studies. Moreover, all connection requests are generated between random nodes in random domains with mean holding times of 600 s (exponential) and variable inter-arrival times (as per

desired load). The actual request sizes are further varied uniformly between 200 Mbps–1 Gbps in increments of 200 Mbps, i.e., to model fractional Ethernet demands. In addition,  $K = 5$  next-hop domain entries are computed in the distance vector table for signaling crankback, although the number searched is limited by  $H_1$  or  $H_2$  values.

Meanwhile, to simulate various fault scenarios, both single link failures and multi-failures are considered for performance evaluation. Here, single link failures are limited to inter-domain links with exponential mean inter-arrival times of 12,000 s, i.e., after every 20 connections on average. Furthermore, each simulation run is averaged over 500,000 connections. Meanwhile, the multi-failure events are specified as a series of simultaneous node/link failures. To further incorporate different failure severities, a failure region is introduced as a circular area denoted by the radius,  $r$ , as shown in Fig. 3. Here,  $r = 1$  corresponds to the basic single-node failure case, and the failure region expands as  $r$  increases. Using the above failure region definition, the performance of the recovery schemes is gauged by measuring post-fault recovery success rates after a single multi-failure event. Specifically, the network is first brought to steady state operation for a given input load, i.e., connection inter-arrival time, and then a large-scale event is triggered to fail a cluster of nodes/links. This procedure is repeated 10 times at each input load point and the performance metric results averaged. Typically, steady-state operation is achieved after establishing about 50,000 connections. Note that only transit connections on failed nodes are considered for restoration, i.e., as failed source/destination connections cannot be recovered.

Post-fault restoration is first evaluated by comparing recovery success rates for hierarchical routing (simple-node/full-mesh abstraction) and signaling crankback (IM/E2E recovery) against single link faults. For crankback

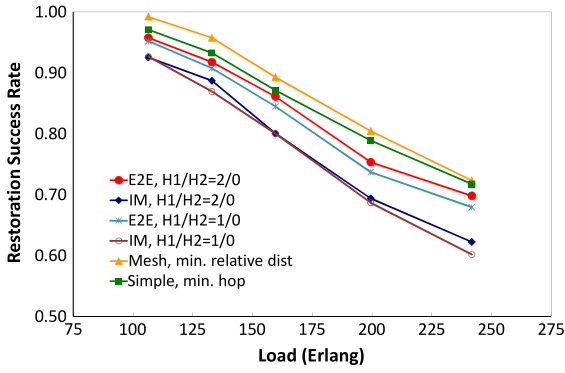


Fig. 4. Single-link-failure restoration success rate (intra-domain crankback only, i.e.,  $H_1 > 0$  and  $H_2 = 0$ ).

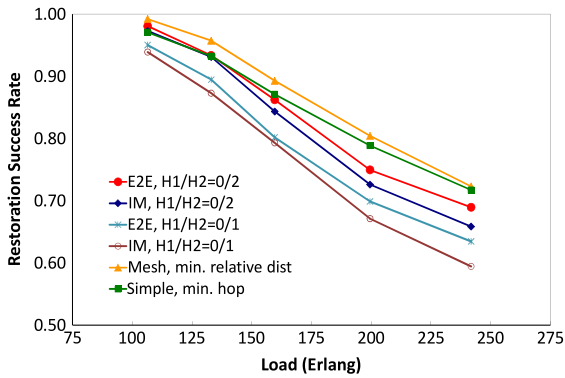


Fig. 5. Single-link-failure restoration success rate (inter-domain crankback only, i.e.,  $H_1 = 0$  and  $H_2 > 0$ ).

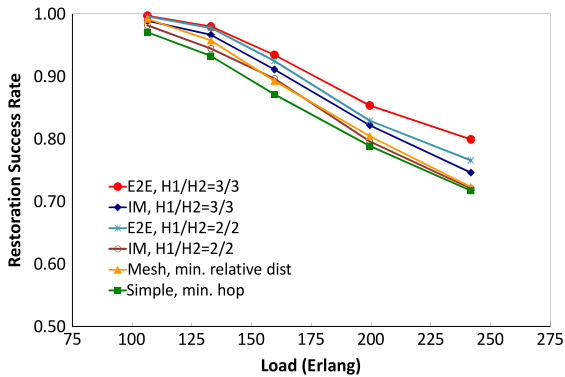


Fig. 6. Single-link-failure restoration success rate (joint crankback, i.e.,  $H_1 > 0$  and  $H_2 > 0$ ).

restoration, different combinations of intra-/inter-domain crankback counters are tested for comparison purposes. Specifically, Fig. 4 shows the restoration success rates with *intra-domain* crankback only, i.e.,  $H_1 > 0$  and  $H_2 = 0$ , whereas the results for *inter-domain* only crankback, i.e.,  $H_1 = 0$  and  $H_2 > 0$ , are plotted in Fig. 5. These results clearly show that intra- or inter-domain only crankback yields lower recovery than both forms of hierarchical routing (without any crankback). Meanwhile, the results for joint intra/inter-domain crankback are plotted in Fig. 6 for crankback counter values of  $H_1 = H_2 = 2$  and 3. Both of

these settings give notably improved performance by allowing multiple re-tries within a domain before regressing to another upstream domain. In fact, the recovery rates surpass those for the “non-crankback” hierarchical routing strategies in almost all cases. Carefully note that further increasing the crankback counters is not always beneficial and may lead to reduced effectiveness, i.e., as resource efficiencies will decline with the selection of excessively long routes.

The results in Figs. 4–6 also show that E2E crankback consistently outperforms IM crankback, as it tends to distribute attempts across the entire network (and not concentrate them to domains immediately prior to the failed links). In addition, for hierarchical routing the full-mesh abstraction approach (coupled with minimum relative distance path computation) always gives better recovery than simple node abstraction, as it provides more accurate “global” state and better load distribution. Overall, these results show very good multi-domain recovery, i.e., over 70% at high traffic loads.

Generally, post-fault restoration may affect the performance of regular working-mode operation as resource usages may increase after failure events. To gauge this effect, Fig. 7 plots the *bandwidth blocking rate* (BBR) for working-only connections for all schemes (the joint crankback counter setting  $H_1 = H_2 = 3$  is used here as it shows the best recovery in the previous runs). These results show that signaling crankback strategies yield blocking regions that lie in between simple-node (upper bound) and full-mesh (lower bound) topology abstractions. As a result, it can be stated that improved signaling restoration comes at the expense of slightly increased blocking of regular working connection requests (versus full-mesh hierarchical routing). Furthermore, the average path lengths across recovered routes and restoration delays are also plotted in Figs. 8 and 9. In particular, restoration delays assume a failure detection time of 1 ms, link propagation delays of 1 ms, and node processing times of 0.1 ms. As expected, the hierarchical routing strategies yield notably shorter paths and recovery times, albeit these gains come at the expense of sizable increased in inter-domain routing overheads due to topology abstraction and update dissemination. In addition, IM restoration always gives longer delays and path lengths than the E2E scheme (by up to 25%), i.e., as it generally encounters more resource contention and hence triggers more crankback re-tries.

Finally, the signaling crankback and hierarchical routing strategies are evaluated for post-fault restoration under multi-failure scenarios. According to the multi-failure model illustrated in Fig. 3, varying radii of  $r = 1$  and  $r = 3$  are chosen to simulate different fault severities. The respective recovery success rates are shown in Figs. 10 and 11 (averaged over 10 independent multi-failure events). These findings show that crankback generally outperforms hierarchical routing with simple-node abstraction (plus minimum hop path computation). However, the more advanced hierarchical routing scheme with full-mesh abstraction gives the best overall recovery, with the most improvement in the case of highly destructive  $r = 3$  failure scenarios. Indeed, this scheme benefits from the improved level of global visibility afforded by increased back-off timer intervals. However, recovery

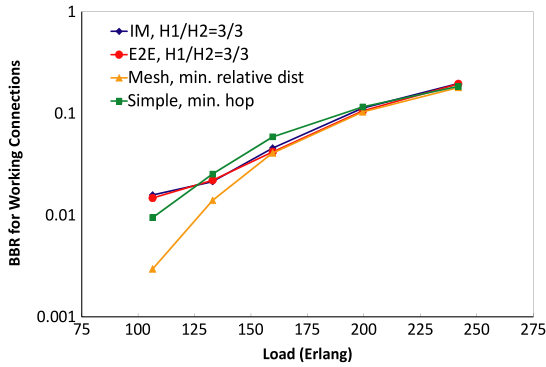


Fig. 7. Blocking rate for working-mode connection only.

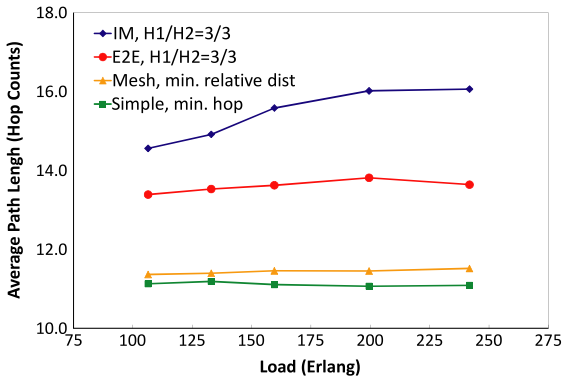


Fig. 8. Average path length of restored routes.

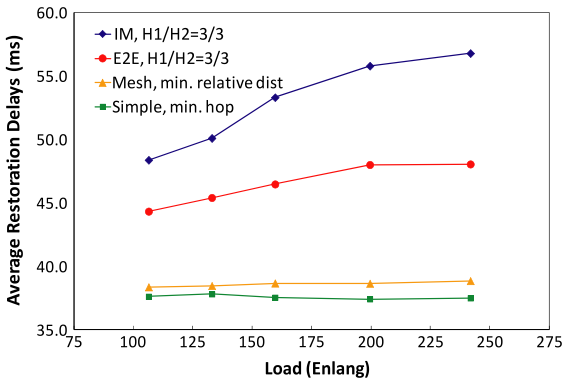


Fig. 9. Average restoration delay (ms).

times can be longer and the increased hierarchical routing overheads (for abstracted domain state) also cannot be ignored.

## 5. Conclusions

This paper proposes two enhanced solutions for post-fault restoration in multi-domain networks, based upon hierarchical routing and signaling crankback, respectively. These schemes extend upon existing strategies for working-mode provisioning by adapting them for post-failure recovery purposes. Overall, both of the schemes show good recovery levels when encountering single and

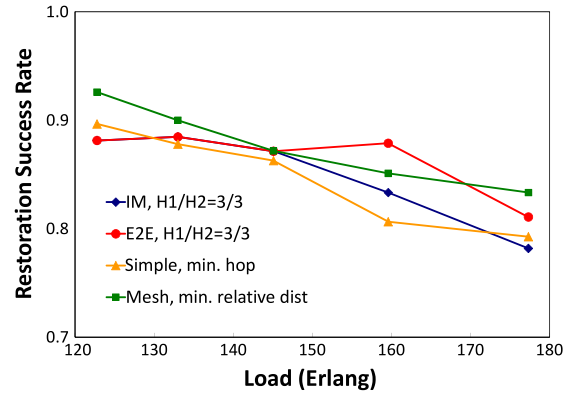


Fig. 10. Multi-failure restoration success rate ( $r = 1$ ).

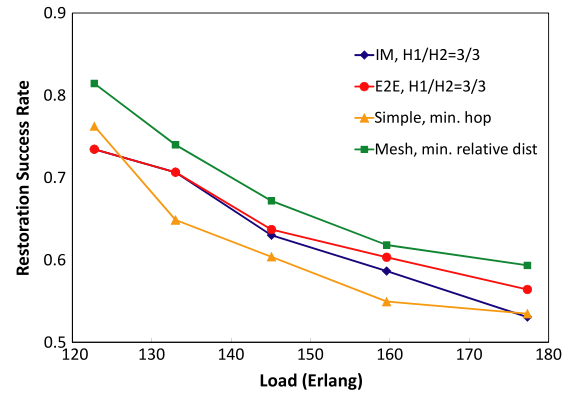


Fig. 11. Multi-failure restoration success rate ( $r = 3$ ).

multiple link failures. However, the hierarchical routing strategy (with advanced full-mesh abstraction) tends to give the best overall performance, particularly when coupled with appropriate offset back-off intervals. By contrast, the signaling crankback schemes can also achieve competitive results without the sizeable increase in inter-domain routing overheads. Future research efforts will look at pre-provisioned protection strategies under various failure scenarios and consider their integration with restoration techniques, i.e., to establish a comprehensive tiered survivability framework. Theoretical rigorous analysis on multi-domain provisioning and survivability will also be a direction of study.

## Acknowledgments

This work has been supported in part by the Defense Threat Reduction Agency (DTRA) and National Science Foundation (NSF) under Award CNS-0806637. The authors are very grateful for this support.

## References

- [1] P. Cholda, et al. A survey of resilience differentiation frameworks in communication networks, in: IEEE Communications Surveys & Tutorials, 4th Quarter, 2007.



- [2] G. Bernstein, B. Rajagopalan, D. Saha, *Optical network control-architecture, protocols and standards*, Addison-Wesley Longman, Boston, MA, USA, 2003.
- [3] N. Ghani, et al., Control plane design in multidomain/multilayer optical networks, *IEEE Communications Magazine* 46 (6) (2008) 78–87.
- [4] D. Staessens, et al., Enabling high availability over multiple optical networks, *IEEE Communications Magazine* 46 (6) (2008) 120–126.
- [5] T. Takeda, et al. Diverse path setup schemes in multi-domain optical networks, in: *IEEE BROADNETS 2008*, London, England, Sep. 2008.
- [6] T. Takeda, et al. Analysis of inter-domain label switched path (LSP) recovery, in: *IETF RFC 5298*, Aug. 2008.
- [7] S. Dasgupta, J. De Oliveira, J. Vasseur, Path-computation-element-based architecture for interdomain MPLS/GMPLS traffic engineering: overview and performance, *IEEE Network* 21 (4) (2007) 38–45.
- [8] A. Sprintson, et al. Reliable routing with QoS guarantees for multi-domain IP/MPLS networks, in: *IEEE INFOCOM 2007*, Alaska, May 2007.
- [9] D. Troung, B. Jaumard, Recent progress in dynamic routing for shared protection in multidomain networks, *IEEE Communications Magazine* 46 (6) (2008) 112–119.
- [10] J. Szigeti, R. Romeral, T. Cinkler, D. Larrabeiti, *P*-cycle protection in multi-domain optical networks, *Photonic Network Communications* 17 (1) (2009) 35–47.
- [11] L. Guo, et al., Local and global hamiltonian cycle protection algorithm based on abstracted virtual topology in fault-tolerant multi-domain optical networks, *IEEE Transactions on Communications* 58 (3) (2010) 851–859.
- [12] J. Wang, L. Sahasrabudhe, B. Mukherjee, Path vs. subpath vs. link restoration for fault management in IP-over-WDM networks: performance comparisons using GMPLS control signaling, *IEEE Communications Magazine* 40 (11) (2002) 80–87.
- [13] F. Xu, et al., Multi-domain restoration with Crankback in IP/MPLS networks, *Optical Switching and Networking* 8 (1) (2011) 68–78.
- [14] J. Suurballe, R. Tarjan, A quick method for finding shortest pairs of disjoint paths, *Networks* 14 (1984) 325–336.
- [15] J. Zhang, et al. A comprehensive study on backup re-provisioning to remedy the effect of multiple-link failures in WDM mesh networks, in: *IEEE ICC 2004*, Paris, June 2004.
- [16] H. Choi, S. Subramaniam, On double-link failure recovery in WDM optical networks, in: *IEEE INFOCOM 2002*, New York City, NY, June 2002.
- [17] D. Schupke, R. Prinz, Performance of path protection and rerouting for WDM networks subject to dual failures, in: *IEEE/OSA OFC 2003*, Atlanta, GA, March 2003.
- [18] S. Kini, et al. Fast recovery from dual link failures in IP networks, in: *IEEE INFOCOM 2009*, Brazil, April 2009.
- [19] S. Stefanakos, Reliable routings in networks with generalized link failure events, *IEEE/ACM Transactions on Networking* 16 (6) (2008) 1331–1339.
- [20] H. Lee, E. Modiano, Diverse routing in networks with probabilistic failures, in: *IEEE GLOBECOM 2009*, Honolulu, HI, Dec. 2009.
- [21] Q. Liu, et al. Distributed grooming in multi-domain IP/MPLS-DWDM networks, in: *IEEE GLOBECOM 2009*, Hawaii, Nov. 2009.
- [22] F. Xu, et al. Enhanced signaling crankback for multi-domain traffic engineering, in: *IEEE ICC 2010*, Cape Town, South Africa, May 2010.
- [23] R. Alnuweri, et al., Performance of new link state advertisement mechanisms in routing protocols with traffic engineering extensions, *IEEE Communications Magazine* 42 (5) (2004) 151–162.
- [24] D. Cavendish, H. Ohta, H. Rakotoranto, Operation, administration, and maintenance in MPLS networks, *IEEE Communications Magazine* 42 (10) (2004) 91–99.
- [25] A. Farrel, et al. Signaling crankback extensions for MPLS and GMPLS RSVP-TE, in: *IETF Request RFC 4920*, July 2007.