Multi-Failure Post-Fault Restoration in Multidomain DWDM Networks

F. Xu[1], M. Peng[2], A. Rayes[3], N. Ghani[1], A. Gumaste[4]

[1]*University of New Mexico,* [2]*Wuhan University,* [3]*Cisco Systems,* [4]*IIT Bombay*

Abstract: This paper proposes an adapted crankback-enabled post-fault restoration scheme against large-area failure events in multi-domain DWDM optical network. This novel solution provides recovery and resource efficiency without increasing routing overheads.

I. Introduction

Multi-domain network survivability in optical *dense wavelength division multiplexing* (DWDM) networks has drawn notable attention recently and a range of protection strategies have been proposed [1]. For example, simpler "per-domain" strategies have been tabled in [2] to combine "localized" per-domain protection schemes with reliable multi-domain interconnection techniques. Others have also developed more advanced "domain diverse" strategies [3],[4] to achieve increased working/protection path separation. However, these algorithms rely upon resource *pre-provisioning* and are only designed to handle single failures. As such, they are not effective against large multi-failure events, particularly those yielding highly-correlated (near-simultaneous) failures, e.g., those resulting from large power outages, natural disasters, *weapons of mass destruction* (WMD) attacks, etc.

In light of the above, there is a growing need to develop distributed post-fault restoration solutions for multi-domain/multi-failure recovery using crankback [5]. Although these schemes will yield longer recovery (100's ms to seconds range), they offer a viable last-stop alternative in case of failure of pre-provisioned paths, i.e., under correlated failure scenarios. Now an effective crankback *routing and wavelength assignment* (RWA) algorithm has been proposed for multi-domain optical networks in [6] for *working-mode* only operation. Hence this paper extends this framework and develops two novel multi-domain lightpath restoration strategies for multi-failure recovery.

II. Dynamic Multi-Domain Lightpath Restoration With Signaling Crankback

A multi-domain restoration solution with enhanced signaling crankback is presented. This approach assumes realistic *generalized multi-protocol label switching* (GMPLS) settings in which *optical cross-connect* (OXC) nodes and domain *path computation element* (PCE) entities have complete knowledge of domain-internal resources, e.g., via *open shortest path first-traffic engineering* (OSPF-TE) routing. Meanwhile, only border OXC nodes and PCE entities have some inter-domain visibility, e.g., as provided by *border gateway protocol* (BGP) path vector state. The framework also assumes "all-optical" domains with full conversion at border OXC nodes. This is a very valid assumption in large (inter-carrier) settings as it minimizes inter-domain physical impairments concerns and enables bit-level service monitoring at domain boundaries [6]. Meanwhile, all setup signaling is done using crankback extensions for RSVP-TE [5]. Overall, this solution introduces some key innovations, i.e., 1) end-to-end and intermediate restoration, and 2) adapted crankback with failure *and* link congestion tracking. Consider the details.

A. Enhanced Crankback Scheme for Working-Mode Provisioning

Related notations are provided first. A multi-domain optical network is comprised of D domains, with the i-th domain having n^i OXC nodes and b^i border OXC nodes, $1 \le i \le D$. This network is represented as a set of domain sub-graphs, $\boldsymbol{G}^i(\boldsymbol{V}^i,\boldsymbol{L}^i)$, where $\boldsymbol{V}^i = \{v^i_1, v^i_2, \ldots\}$ is the set of OXC nodes in domain i and $\boldsymbol{L}^i = \{l^{ii}_{jk}\}$ is the set of *intra-domain* links ($1 \le i \le D$, $1 \le j,k \le n^i$). Meanwhile, the *inter-domain* link connecting border OXC node v^i_k in domain i with v^j_m in domain j with free capacity w^{ij}_{km} wavelengths is denoted as l^{ij}_{km}, $1 \le i,j \le D$, $1 \le k \le b^i$, $1 \le m \le b^j$. Each OXC node v^i_j also maintains a list of traversing connections, denoted as $\underline{\boldsymbol{A}}^i_j$ (where each entry consists of a route vector) and a congestion tracking table, $\underline{\boldsymbol{T}}^i_j$ (where each entry consists of an egress link id and expiry time). Finally the RSVP-TE messages store a route vector, $\underline{\boldsymbol{R}}$, a wavelength availability vector, $\underline{\boldsymbol{\lambda}}$, an exclude link vector, $\underline{\boldsymbol{X}}$ (to track signaling failures), and dual intra/inter-domain crankback counters, h_1 and h_2 (initial values set to $h_1=H_1$ and $h_2=H_2$).

In the proposed multi-domain RWA framework, crankback signaling is used in conjunction with distributed "per-domain" lightpath computation, i.e., with limited inter-domain state. Namely, when a request arrives, the source domain PCE selects a "next-hop" domain and initiates intra-domain *explicit route* (ER) RWA expansion towards the specific egress border OXC/link to that domain. Specifically, this step uses *fixed alternate routing* (FAR) with *most-used* wavelength selection over the local domain topology. This process is then repeated by downstream domains until the destination is reached. To improve efficiency and accuracy here, the intelligent next-hop domain selection scheme in [6] is re-used. Namely, all domain PCE's pre-compute multi-entry *distance vector tables* to store up to K next-hop domains/egress links to each destination domain, i.e., done by extracting path-vector connectivity state from BGP tables, see [6]. Note that the route vector, $\underline{\boldsymbol{R}}$, for expanded-route record and the wavelength availability vector, $\underline{\boldsymbol{\lambda}}$, for wavelength selection/conversion, are continuously updated along the route.

Now given that wavelength deficiencies can arise during lightpath setup, various crankback procedures can be initiated. Along these lines, the work herein leverages the joint intra/inter-domain crankback scheme in [6] which limits the number of intra- or inter-domain crankback attempts to H_1 and H_2, respectively, and implements two key operations, i.e., crankback *notification* and *re-computation*. Specifically, if signaling resource failure occurs, i.e., link with w^{ij}_{km}=0 encountered, upstream notification is sent using a *PATH_ERR* message to the current domain's ingress border OXC, i.e., *notification*. Here the intra-domain crankback counter h_1 is decremented and the failed link is noted in the exclude link vector, $\underline{\boldsymbol{X}}$, to avoid re-consideration. Meanwhile, upon receiving a *PATH_ERR* notification, the ingress border OXC (the source OXC if in source domain) performs *re-computation*. Namely, if the intra-domain h_1 counter is not zero, another egress border OXC is chosen for ER expansion towards it, i.e., intra-domain crankback. Otherwise, more aggressive inter-domain crankback is done by sending a *PATH_ERR* message to the ingress border OXC in the upstream domain (and h_2 decremented). The lightpath request will be considered ultimately failed if both counters expire.

To further improve setup success, link congestion is also tracked here, extending upon [6]. Namely, each border OXC v^i_j maintains a congestion tracking table, $\underline{\boldsymbol{T}}^i_j$, to record the congested egress links that have no free wavelengths, i.e., w^{ij}_{km}=0. This tracking compliments exclude link vectors which only track information for a specific signaling instance (and cannot benefit subsequent setup attempts). Here, if an egress OXC detects an outgoing inter-domain link with no free wavelengths, it loops through all active connections traversing the link, $\underline{\boldsymbol{A}}^i_j$, to find the one with earliest stop-time. This value is then inserted/piggybacked in an appropriate signaling message (i.e., *PATH_ERR*, *RESV_ERR*, or *RESV* depending upon the stage of setup process) to notify the upstream ingress OXC in the domain. This OXC then updates the congestion tracking table entry for the link, $\underline{\boldsymbol{T}}^i_j$, and does not select it until the stop-time expires. Hence link status history is effectively extracted from both working and restoration mode operation.

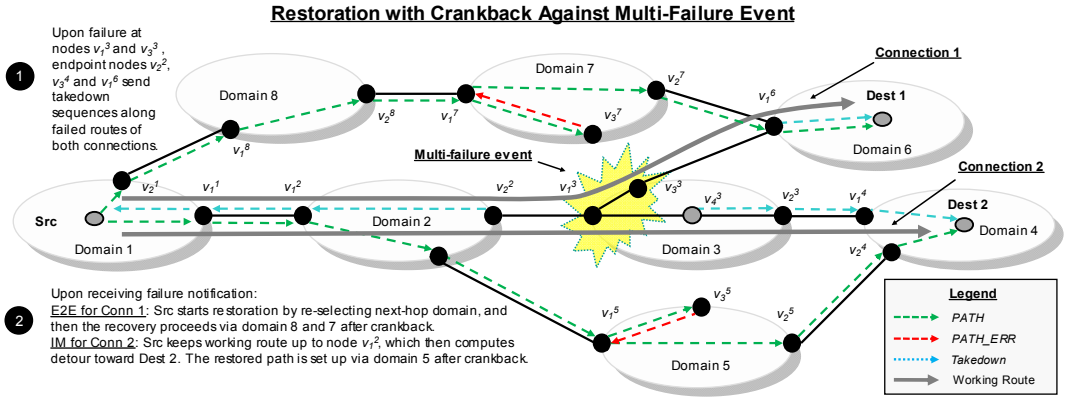

Figure 1: Dynamic end-to-end and intermediate restoration schemes with signaling crankback

B. Multi-Domain Post-Fault Signaling Restoration

Consider the further extension of the above crankback framework for post-fault restoration under extreme conditions with multiple link and/or node failures. Here it is assumed that failed OXC nodes will bring down all of their links, and the working OXC endpoints on these failed links will quickly discover these failures via rapid lower layer mechanisms (out of scope herein). Hence each failure-detecting end-point OXC node, i.e., v^i_j, simply loops through its active connection list, $\underline{\boldsymbol{A}}^i_j$, to search for any lightpaths traversing the failed link. All such lightpaths are then deleted from $\underline{\boldsymbol{A}}^i_j$ and appropriate restoration procedures are initiated. Specifically, a detecting OXC node first sends out resource takedown messages along the failed upstream or downstream path segments to free wavelengths. Next, the detecting OXC on the upstream side of the failed link also notifies the source OXC of the connection via a *PATH_ERR* message with the failed link noted in the exclude link vector, $\underline{\boldsymbol{X}}$. This notification also sets an appropriate restoration flag to indicate connection failures. Upon receiving failure notifications from a downstream OXC, the source OXC node initiates post-fault crankback setup to re-compute backup lightpaths. Now since multiple crankback attempts can be triggered immediately after a failure, source OXC's use random back-off intervals before initiating their crankbacks. In terms of re-computation, two restoration strategies are proposed:

End-to-end (E2E): The source OXC re-selects the next-hop domain (using current distance vector tables) and sends out a new *PATH* message towards the destination OXC node. Here the route vector $\underline{\boldsymbol{R}}$ is cleared, crankback counters h_1 and h_2 are reset, and the failed links from received *PATH_ERR* message are noted in exclude link list, $\underline{\boldsymbol{X}}$. The request is then processed using regular crankback procedures as specified for working-mode in Section II.A.

Intermediate (IM): The source OXC preserves the original path up to the ingress node in the domain preceding the failures, i.e., by removing all nodes downstream of this node to the destination. A new *PATH* message is sent containing this modified route and crankback is then initiated from the downstream node.

Both E2E and IM restoration are shown in Fig. 1. Namely, connection 1 uses E2E restoration, where upon receiving notification from detecting OXC node v^2_2, the source OXC finds domain 8 as the next-hop domain. The restoration concludes successfully via domain 7 after one intra-domain crankback. Meanwhile, connection 2 uses IM restoration, which traces back along the original route to ingress OXC v^2_1 and then initiates re-routing via domain 5.

III. Performance Evaluation

The restoration scheme is tested using detailed *OPNET Modeler*^{TM} models over a modified NSFNET topology in which each node is replaced by a domain with 7-10 nodes each, i.e., 16 domains, 25 inter-domain links. Link sizes are fixed to 16 wavelengths (intra/inter-domain) and lightpath requests are randomly generated between domains/ intra-domain nodes, with exponential holding times (mean 600 sec). In addition, K=5 next-hop domain entries are computed in the multi-entry distance vector tables, although the number searched are limited by the H_1 or H_2 values. Overall, the scheme is tested for two different *failure scenarios* (FS); namely, single-border-node failures (FS1) and simultaneous three-node failure regions comprising of one border node and two intra-domain nodes (FS2). To perform these tests, the network is first brought into steady state by generating 500,000 requests, and then the multi-failure event is triggered and subsequent recovery performance is measured (further averaged over 10 random seeds).

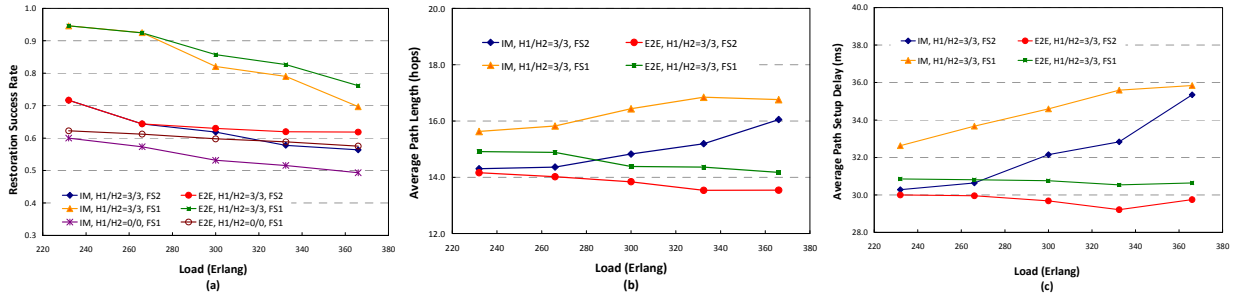

Figure 2: (a) restoration success rates; (b) average restored path lengths; (c) average restored path setup delay (ms)

Post-fault restoration is first tested by measuring the recovery rates for the different failures. Now earlier results for working-mode operation in [6] indicate that intra-/inter-domain crankback counters values of H_1=3/H_2=3 tend to give the lowest blocking. Hence these values are adopted here and different crankback scenarios tested, i.e., joint intra/inter-domain (H_1=3/H_2=3) and no crankback (H_1=0/H_2=0, equivalent to K=1). The associated restoration success rates are plotted in Fig. 2a and indicate declining recovery success with increasing failure severities. Namely, on average 80% of the failed connections for FS1 cases can be recovered whereas this level drops to about 65% for FS2 cases. This is expected as larger failure regions induce more failed connections and increased resource contention during crankback. More importantly, E2E restoration consistently outperforms IM restoration, particularly at higher traffic loads. The reason here is that the latter scheme tries to re-route multiple failed connections from domains immediately prior to the failure region, thereby aggravating resource contention on domain egress links. By contrast, the E2E scheme achieves better load distribution as it attempts recovery over a wider set of network domains. The use of crankback also gives notable gains over non-crankback operation, i.e., absolute restoration success rate increases of 20-35%. Meanwhile, the resource efficiency (in terms of average hop counts) and setup delays are also plotted in Figs. 2b and 2c, respectively (assuming 1 ms backbone link delays and 0.05 ms OXC message processing times). Again, E2E consistently outperforms IM restoration for both metrics, yielding about 15-20% lower utilization and delays at low-mid loadings. As such, these results indicate the viability of E2E post-fault lightpath restoration RWA to recover from large-scale multi-failure events.

References

[1] M. Chamania, A. Jukan , "A Survey of Inter-Domain Peering and Provisioning Solutions for the Next Generation Optical Networks," *IEEE Communications Survey & Tutorials*, Vol. 11, No. 1, March 2009, pp. 33-51.

[2] F. Ricciato, U. Monaco, D. Ali, "Distributed Schemes for Diverse Path Computation in Multidomain MPLS Networks," *IEEE Communications Magazine*, June 2005, Vol. 43, No. 6, pp. 138-146.

[3] A. Sprintson, *et al*, "Reliable Routing with QoS Guarantees for Multi-Domain IP/MPLS Networks," *IEEE INFOCOM 2007*, May 2007.

[4] D. Troung, B. Jaumard, "Recent Progress in Dynamic Routing for Shared Protection in Multidomain Networks", *IEEE Communications Magazine*, Vol. 46, No. 6, June 2008, pp. 112-119.

[5] A. Farrel, *et al*, "Crankback Signaling Extensions for MPLS and GMPLS RSVP-TE", *IETF Request RFC 4920*, July 2007.

[6] M. Esmaeili, F. Xu, N. Ghani, M. Peng, Q. Liu, "Enhanced Crankback Signaling in Multi-Domain Optical Networks", *IEEE/OSA Optic Fiber Communications Conference (OFC) 2010*, San Diego, CA, March 2010.