Enhanced Crankback Signaling in Multi-Domain Optical Networks

M. Esmaeili[1], F. Xu[1], N. Ghani[1], M. Peng[2], Q. Liu[3]
[1]University of New Mexico, [3]Wuhan University, [3]Oak Ridge National Laboratory

Abstract: A novel solution is proposed for joint intra/inter-domain signaling crankback for multi-domain DWDM lightpath setup. The scheme leverages existing intra/inter-domain routing protocol state and provides mechanisms to track signaling failures and limit setup overheads and delays.

1. Introduction

Multi-domain optical *dense wavelength division multiplexing* (DWDM) networking is a key focus area and various inter-domain *routing and wavelength* (RWA) schemes have been proposed. For example, some have used hierarchical link-state routing with topology abstraction to compress domain wavelength/ converter state [1],[2], whereas others have adapted more scalable "optical" distance/path-vector routing approaches [3],[4]. Furthermore, the IETF *path computation element* (PCE) framework [1],[5] has also defined several distributed multi-domain path computation strategies, e.g., *per-domain* and *PCE-based*.

However, neither hierarchical link-state nor distance/path-vector routing can provide fully-accurate DWDM state across domains. Hence it is important to consider distributed "per-domain" crankback/re-try signaling strategies as they have minimal *inter-domain* routing requirements. However, existing multi-domain crankback studies have only focused on IP networks. For example, the exhaustive *"per-domain"* (PD) scheme [5] probes domain egress nodes and upon failure, cranks back to upstream gateways. Results show higher blocking and delay versus pre-determined search strategies. Meanwhile [6] addresses crankback delays and outlines next-hop domain selection using inter-domain round-trip time samples. This yields decent reductions in setup delay but requires use of a specialized coordinates system. Overall, these schemes leave significant room for improvement/adaptation to DWDM settings and, more importantly, require detailed comparisons with competing hierarchical DWDM routing schemes. This is the focus herein.

2. Enhanced Crankback Solution

An enhanced crankback RWA scheme is proposed using standard RSVP-TE signaling (with crankback, RFC 4920). The work addresses realistic settings in which interior *optical cross-connect* (OXC) nodes and PCE entities have full resource visibility of their domain, as provided by link-state routing protocols such as OSPF-TE. However, inter-domain visibility is only available at border OXC nodes (and PCE entities) and is limited to path/distance-vector routing state, as provided by *border/exterior gateway protocols* (BGP/EGP). Furthermore, all interior OXC nodes are all-optical whereas border OXC nodes have full opto-electronic conversion, i.e., reflective of real-world settings where "all-optical" islands are delineated by domain boundary regeneration for full bit-level monitoring [1]. The proposed solution here introduces several key crankback innovations, i.e., 1) dual intra/inter-domain counters to limit delays, 2) crankback failure history tracking and usage to improve success, and 3) intelligent next-hop domain selection.

A. Multi-Domain Notification and Re-Computation

Consider the notation. A multi-domain network is comprised of D domains, with the i-th domain having n^i OXC nodes and b^i border OXC nodes. This network is represented as a set of domain sub-graphs, $\boldsymbol{G}^i(\boldsymbol{V}^i,\boldsymbol{L}^i)$, $1\leq i\leq D$, where $\boldsymbol{V}^i=\{ v^i_1, v^i_2, \ldots \}$ is the set of domain OXC nodes and $\boldsymbol{L}^i=\{ l^{ii}_{jk} \}$ is the set of *intra-domain* links, i.e., l^{ii}_{jk} is the link from v^i_j to v^i_k with free capacity c^{ii}_{jk} ($1\leq i\leq D$, $1\leq j,k\leq n^i$). Meanwhile, the inter-domain link connecting border node v^i_k in domain i with border node v^j_m in domain j is denoted as l^{ij}_{km} and has free capacity c^{ij}_{km}, $1\leq i,j\leq D$, $1\leq k\leq b^i$, $1\leq m\leq b^j$. Finally, all RSVP-TE messages contain a path route tracking vector, $\underline{\boldsymbol{R}}$, a wavelength availability vector, $\underline{\boldsymbol{\lambda}}$, and crankback fields, i.e., including an exclude link vector, $\underline{\boldsymbol{X}}$, to track signaling failure history and *dual* intra/inter-domain crankback counters, h_1 and h_2.

"Per-domain" comptuation is done in a recursive manner starting at the source domain and uses RSVP-TE *PATH* and *PATH_ERR* messaging. Here the source OXC (or domain ingress border OXC) consults its PCE to determine the next-hop domain to the destination domain, i.e., identify next-hop domain and egress border OXC/link in the current domain (detailed in Section 2.B). Upon receiving this information, the source OXC (or domain ingress OXC) uses its local routing database to compute an *explicit route* (ER) to the appropriate egress border OXC. This route is then inserted into the route field, $\underline{\boldsymbol{R}}$, of a downstream *PATH*. This message also contains an "all-ones" wavelength availability ($\underline{\boldsymbol{\lambda}}$ =[1,…1]) to be AND-ed with the available wavelength vectors of the intra-domain ER links to find an all-optical intra-domain path. Since wavelength conversion is done at domain entry, ingress border OXC nodes must save the availability vectors from the previous domain in $\underline{\boldsymbol{R}}$ and generate a new "all-ones" vectors for downstream processing.

978-1-55752-884-1/10/$26.00 ©2010 IEEE

Now crankback operation is only initiated upon *PATH* signaling failure, i.e., wavelength inavailability at a given outbound link. Here for multi-domain settings two main crankback procedures are defined, *notification* and *re-computation*, Figure 1. The former performs upstream notification upon *PATH* failure at an intermediate OXC. Meanwhile the latter performs actual re-routing to select a new lighpath route. Now in general, *PATH* signaling failures can occur at *three* OXC node types, i.e., domain ingress OXC nodes, domain egress OXC nodes, and interior OXC nodes. In the enhanced scheme herein, only the former OXC types perform re-computation whereas the latter two types simply perform notification:

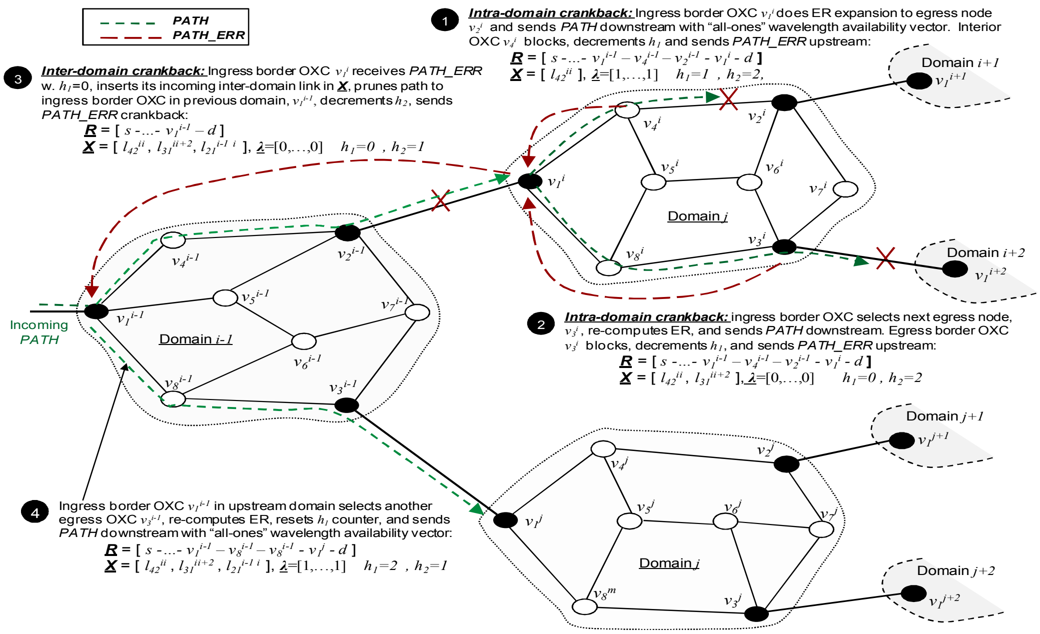

Figure 1: Enhanced intra/inter-domain crankback scheme (H_1=2, H_2=2)

Crankback Notification: Notification is done when there is no available wavelength at an intra-domain link (intra-domain OXC), i.e., i.e., $\underline{\lambda}$=[0,…,0], or there is no available converter at an inter-domain link (egress border OXC). Here the *PATH* message is terminated and a *PATH_ERR* crankback is sent to the domain's ingress border OXC. In this message, the intra-domain counter h_1 is decremented and the failed link noted in $\underline{X}$. In addition the route in $\underline{R}$ is also pruned to remove all intra-domain OXC nodes on the failed ER route up to ingress border node. This overall notification procedure is shown in Figure 1 (for H_1, H_2=2). Namely, when wavelength blocking occurs on link l^{ii}_{42}, i.e., step 1, OXC v_4^i prunes the route vector $\underline{R}$ to the domain ingress node v_1^i, adds the blocked link to $\underline{X}$, and decrements the intra-domain counter h_1. This information is then sent to v_1^i via a *PATH_ERR* (see also step 2 at egress border OXC link v_3^i).

Crankback Re-Computation: Re-computation is done at the intra-/inter-domain levels by ingress border OXC nodes receiving a *PATH_ERR*. If h_1 is not zero, *intra-domain* re-computation selects a new next-hop domain/egress border OXC for ER expansion. The exact sequence of next-hop domains is pre-computed to successively search longer inter-domain routes (see Section 2.B). Crankback history in $\underline{X}$ is also fully leveraged to avoid any failed intra/inter-domain links during ER expansion. Now if a suitable next-hop domain cannot be found and h_2 is non-zero, inter-domain crankback is initiated via a *PATH_ERR* message to the ingress border OXC in the *upstream* domain. Namely, the notifying ingress border OXC inserts its *ingress link* in $\underline{X}$ to further improve failure tracking. Overall, the number of re-attempts is limited to H_1H_2. Now re-computation is also shown in Figure 1. Namely, when ingress OXC v_1^i receives a *PATH_ERR* with h_1=0, it notes ingress $l^{i-1\ i}_{21}$ as failed, prunes the route to the ingress border OXC in the prior domain, v_1^{i-1}, and sends a *PATH_ERR* (step 3). This upstream OXC then re-tries a new egress border OXC, v_3^{i-1} (step 4).

B. Next-Hop Domain Selection

The enhanced crankback scheme leverages existing inter-domain path/distance vector routing state to improve the search process. Namely, PCE entities pre-compute *multi-entry* distance vector tables with up to K next-hop domains/egress links to each destination domain. This computation uses the exterior path/distance vector database to first build a "simple node" graph view [2] of the global network, $\boldsymbol{H}(\boldsymbol{U},\boldsymbol{E})$, i.e., where $\boldsymbol{U}$ is the set of domains $\{\boldsymbol{G}^i\}$ reduced to vertices and $\boldsymbol{E}$ is the set of inter-domain links $\{l^{ij}_{km}\}$, $i \neq j$. An iterative shortest-path algorithm is then run to successively compute/prune multiple routes to all destination domains over $\boldsymbol{H}(\boldsymbol{U},\boldsymbol{E})$ and the respective egress links from the source domain are stored in the table. Note that the number of entries to a destination will be upper-bounded by the minimum of K and the maximum number of links egressing the domain. Hence crankback re-computation (Section 2.A) simply searches these K entries to a destination domain, sequentially driving searches along increasing length domain sequences. Overall, these entry tables will be relatively static if EGP state changes are infrequent.

3. Performance Evaluation

The enhanced multi-domain DWDM crankback scheme is tested by developing detailed models in *OPNET Modeler*^{TM}. Tests are done for a 10-domain topology with 25 inter-domain links as well as the NSFNET topology (nodes replaced by domains) with 16 domains/25 inter-domain links, i.e., average of 2.5 and 1.56 links/domain, respectively. Each domain has 15 nodes and 16 wavelengths/link. Lightpath requests are randomly generated between domains/nodes and each run comprises of 250,000 requests with exponential holding times (mean 600 sec). Also, K=5 next-hop domain entries are computed in the multi-entry distance vector table, although the number searched is limited by H_2. Finally, the enhanced crankback scheme is also compared against more complex hierarchical inter-domain DWDM link-state routing solutions that use simple node and full-mesh topology abstractions [11].

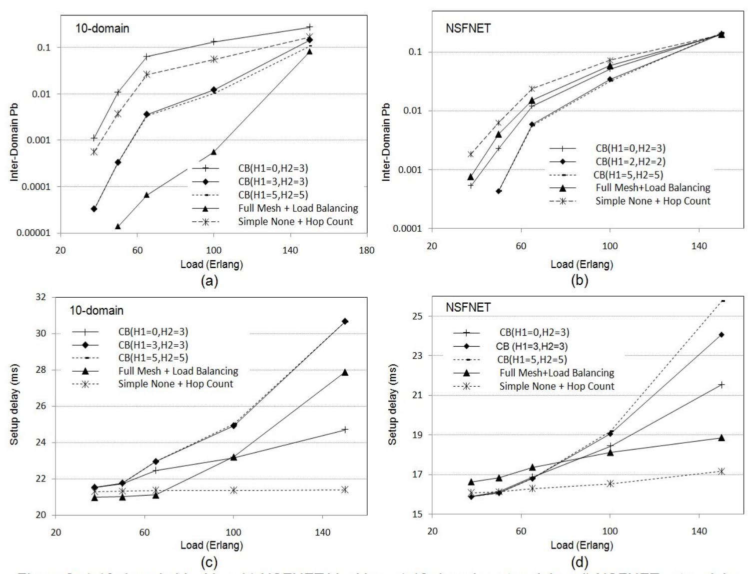

Figure 2 a) 10-domain blocking, b) NSFNET blocking, c) 10-domain setup delay, d) NSFNET setup delay

Inter-domain lightpath blocking is first measured in Figures 2a (10-domain) and 2b (NSFNET) for varying crankback levels. Foremost, the results indicate that joint intra/inter-domain crankback with moderate counter values yields the best performance, i.e., blocking reduction levels off after H_1,H_2=3. Moreover, inter-domain-only crankback (H_1=0) is not effective and yields notably higher blocking. More importantly, the enhanced crankback scheme outperforms hierarchical DWDM routing with simple node abstraction in all cases and even outperforms advanced full-mesh abstraction for the NSFNET topology, i.e., lower inter-domain connectivity. These gains also come with much lower control plane overheads as crankback overheads are over an order magnitude lower than hierarchical routing overheads at mid-to-high loads (not shown). Next, inter-domain setup delays are plotted in Figures 2c (10-domain) and 2d (NSFNET), assuming 1 ms backbone link delays and 0.05 ms OXC message processing delays. Here the enhanced crankback scheme yields increased lightpath setup delays, particularly at high loads, and this is most notable in the NSFNET topology which has lower inter-domain connectivity, i.e., almost 30% higher setup delays. Nevertheless, these values are generally acceptable for long-standing circuit-switched demands.

Overall these results show that moderate levels of intra/inter-domain crankback driven by distance/path-vector state achieve a good tradeoff between provisioning complexity and blocking for inter-domain RWA.

4. References

[1] N. Ghani, *et al*, “Control Plane Design in Multidomain /Multilayer Optical Networks,” *IEEE Communications Magazine,* Vol. 46, No. 6, June 2008, pp. 78-87.

[2] Q. Liu, *et al*, “Hierarchical Inter-Domain Routing and Lightpath Provisioning in Optical Networks”, *OSA Journal of Optical Networking,* Vol. 5, No. 10, October 2006, pp. 764-774.

[3] M. Francisco, *et al*, “End-to-End Signaling and Routing for Optical IP Networks,” *IEEE ICC 2002*, New York City, NY, June 2002.

[4] M. Yannuzi, *et al*, “Toward a New Route Control Model for Multidomain Optical Networks,” *IEEE Communications Magazine,* Vol. 46, No. 6, June 2008, pp. 104-111.

[5] S. Dasgupta, J. C. de Oliveira, J. P. Vasseur, “Path-Computation-Element-Based Architecture for Interdomain MPLS/GMPLS Traffic Engineering: Overview and Performance,” *IEEE Network*, Vol. 21, No. 4, July/August 2007, pp. 38-45.

[6] C. Pelssner, O. Bonaventure, “Path Selection Techniques to Establish Constrained Interdomain MPLS LSPs”, *Proceedings of IFIP International Networking Conference*, Coimbra, Portugal, May 2006.