Enhanced Crankback for Lightpath Setup in Multi-Domain Optical Networks

M. Esmaeili, F. Xu, N. Ghani, C. Xie, M. Peng, and Q. Liu

***Abstract*—This paper proposes a novel crankback scheme for routing and wavelength assignment across domains. The scheme leverages existing routing state and provides mechanisms to track signaling setup failures and limit overheads and delays.**

***Index Terms*—Multi-domain optical networks, crankback.**

I. Introduction

MULTI-DOMAIN *dense wavelength division multiplexing* (DWDM) networking is key focus area and various solutions have been proposed for inter-domain lightpath *routing and wavelength assignment* (RWA), see [1],[2]. For example, some have used hierarchical link-state routing (topology abstraction) to reduce domain wavelength/converter state [2] and achieve skeleton path computation. Others have also developed optical distance/path-vector solutions [3],[4]. However, neither of these approaches provides fully-accurate state across multi domains and hence it is important to consider distributed *"per-domain"* crankback strategies.

Now several studies have looked at inter-domain crankback. However, most of these efforts have only treated bandwidth provisioning IP *multi-protocol label switching* (MPLS) networks and not DWDM networks. For example, [5] defines a basic *"per-domain"* (PD) MPLS scheme which probes egress domain nodes for routes and upon failure, notifies upstream border nodes. Results show higher blocking and delays, particularly when compared to PCE strategies utilizing pre-determined inter-domain routes. Meanwhile [6] studies crankback delays and outlines next-hop domain selection using inter-domain round-trip times. This scheme yields decent reduction in setup delays but mandates a special coordinates system. Overall, these schemes leave significant room for improvement to multi-domain DWDM settings.

This letter addresses these crucial concerns and is organized as follows. Section II presents an overview of the enhanced RWA crankback scheme. Detailed simulation results are then presented in Section III along with conclusions in Section IV.

II. Enhanced Crankback Solution

To date, the *Internet Engineering Task Force* (IETF) has defined some key standards for distributed optical network control as part of its *generalized multi-protocol label switching* (GMPLS) framework. These additions include traffic engineering (TE) extensions for *open shortest path first* routing (OSPF-TE) and *resource reservation signaling* (RSVP-TE). In particular, the latest RSVP-TE enhancements also provide crankback support (RFC 4920). The IETF has also formalized a *path computation element* (PCE) framework [1],[5] which defines domain-level computation entities with access to routing databases. Using this GMPLS framework, an enhanced multi-domain crankback RWA solution is proposed for realistic settings. Namely, interior *optical cross-connect* (OXC) nodes and PCE entities are assumed to have full domain visibility (via OSPF-TE) but inter-domain visibility is limited to border OXC nodes, as per inter-area or inter-*autonomous system* (AS) routing protocols. Also, all interior OXC nodes are all-optical whereas border OXC nodes support full opto-electronic conversion on their inter-domain links, i.e., *"all-optical"* islands with regeneration and monitoring [2]. Overall, several key innovations are introduced, i.e., 1) dual intra/inter-domain counters to limit delays, 2) crankback failure tracking, and 3) intelligent next-hop domain selection.

A. Inter-Domain Notification and Re-Computation

As per 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 given as a set of domain sub-graphs, $\mathbf{G}^i = (\mathbf{V}^i, \mathbf{L}^i)$, where $\mathbf{V}^i = [\mathbf{V}^i_1, \mathbf{V}^i_2, ...]$ is the set of nodes and $L^i = [l^{ii}_{jk}]$ is the set of intra-domain links in domain i, i.e., l^{ii}_{jk} is the link from v^i_j to v^i_k $(1 \leq i \leq D, 1 \leq j, k \leq n^i)$ Meanwhile, the inter-domain link between the *k*-th (border) node v^i_k in domain *i* with the *m*-th (border) node v^j_m in domain *j* is denoted as l^{ij}_{km}, $1 \leq i, j \leq D$, $1 \leq k \leq b^i$, $1 \leq m \leq b^j$. Also, RSVP-TE messages contain an *explicit route object* (ERO) to record the path vector, $\underline{\boldsymbol{R}}$, and an *exclude route object* (XRO) to track the list of congested (crankback) links, $\underline{\boldsymbol{X}}$. For simplicity, it is also assumed that all links use the same wavelength grid and hence the label set objects which track (sub-channel) wavelength usages can be represented by binary availability vectors, $\underline{\boldsymbol{\lambda}}$. Dual intra/inter-domain countdown counters, h_1 and h_2, are also supported and these are initialized to the respective maximum number of intra/inter-domain crankback attempts, i.e., h_1=H_1 and h_2=H_2.

Now in the proposed scheme,"per-domain" RWA is done in a recursive manner starting at the source domain using RSVP-TE *PATH* and *PATH-ERR* messages. Here the source OXC (or domain ingress border OXC) queries its PCE to determine the next-hop domain to the destination domain. Upon receiving this request, the PCE identifies the next-hop domain and egress border OXC/link in the current domain (see Section II.B) and uses its local routing database to expand an *explicit route* (ER) to the selected egress OXC. Here the intra-domain lightpath route is selected as the minimum-hop feasible route, i.e., at least one free wavelength. Upon receiving this information, the source OXC (domain ingress OXC) sends a downstream *PATH* with the expanded route in the ERO. This message

Manuscript received November 19, 2009. The associate editor coordinating the review of this letter and approving it for publication was G. Lazarou.

This work was supported by the NSF and DOE.

M. Esmaeili, F. Xu, N. Ghani, and C. Xie are with the ECE Department, University of New Mexico, USA (e-mail: nghani@ece.unm.edu).

M. Peng is with the CS Department of Wuhan University, Wuhan, China.

Q. Liu is with the Oak Ridge National Laboratory, TN, USA.

Digital Object Identifier 10.1109/LCOMM.2010.05.092269

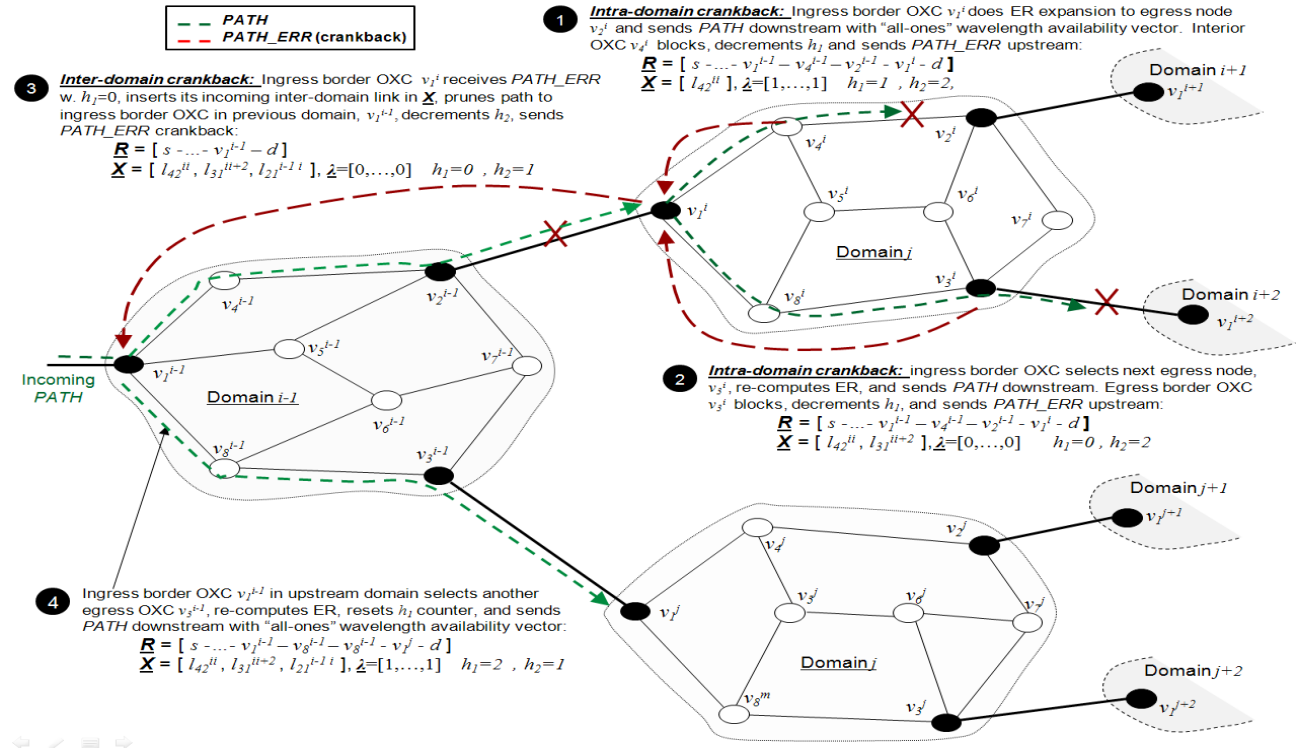

Fig. 1. Enhanced signaling crankback for inter-domain RWA (H1, H2).

also contains the crankback counters (h_1,h_2) and an "all-ones" wavelength availability vector ($\underline{\lambda}$=[1,...1]) to be AND-ed with the available wavelength vectors of the intra-domain ER links, i.e., to find an "all-optical" intra-domain path. Now since wavelength conversion is done at domain entry, an ingress border OXC node must also save the availability vector from the previous domain in $\underline{\boldsymbol{R}}$ and and then generate a new "all-ones" $\underline{\boldsymbol{\lambda}}$ vector for downstream *PATH* processing. Note that wavelength selection for lightpath segments is done during the upstream *RESV* phase using *most-used* (MU) selection as it tends to give lower blocking [2]. Although this requires per-wavelength link-state within a domain, this is generally feasible for mid-sized domains.

Now crankback is only initiated when there is a *PATH* signaling "failure", i.e., wavelength and/or converter inavailability at a given outbound link. Here, two main crankback procedures are defined for multi-domain settings, i.e., *notification* and *re-computation*, Fig. 1. The former performs upstream notification upon *PATH* failure at an intermediate OXC. Meanwhile the latter performs re-routing to select a new lightpath route. Now *PATH* signaling failures can occur at three OXC node types, i.e., domain ingress OXC nodes, domain egress OXC nodes, and interior OXC nodes. Herein, only the former OXC types perform re-computation whereas the latter two types only perform notification.

Crankback Notification: Notification is done when there is no available wavelength at an intra-domain link (intra-domain OXC), i.e.,$\underline{\boldsymbol{\lambda}}$=[0,...0], or there is no available wavelength or converter at an inter-domain link (egress border OXC). Here the *PATH* message is terminated and an upstream *PATH-ERR* crankback is sent to the domain's ingress border OXC. In this message, the intra-domain crankback counter h_1 is decremented and the failed link noted in $\underline{\boldsymbol{X}}$. Also, the route in $\underline{\boldsymbol{R}}$ is pruned to remove all intra-domain OXC nodes on the failed ER route up to ingress border node. This notification procedure is shown in Fig. 1 for H_1,H_2=2. Namely, when wavelength blocking occurs on link l_{42}^{ii} (step 1, Fig. 1), OXC v_4^i prunes the route vector $\underline{\boldsymbol{R}}$ to the domain ingress node v_1^i, adds the blocked link to $\underline{\boldsymbol{X}}$,and decrements h_1.This information is then sent to v_1^i via a *PATH-ERR*. The case of subsequent wavelength blocking at an egress border OXC link is also shown, i.e., at link $l_{31}^{i\ i+2}$ at node v_3^i(step 2, Fig. 1).

Crankback Re-Computation: Re-computation is done at the intra/inter-domain levels by an ingress border OXC node receiving a *PATH-ERR*. Here, if h_1 is not zero, *intra-domain* re-computation is done to select a new next-hop domain/egress border OXC for ER expansion. Now the exact sequence of next-hop domains is pre-computed to successively search longer inter-domain routes (see Section II.B). Crankback history in $\underline{\boldsymbol{X}}$ is also fully leveraged to avoid any failed intra/inter-domain links during ER expansion, i.e., by pruning appropriate links before expansion. Now if a suitable next-hop domain cannot be found (or h_1=0) and inter-domain h_2 counter is non-zero, inter-domain crankback is initiated by sending an upstream *PATH-ERR* to the *ingress border* OXC in the previous domain. Here the notifying *ingress border* OXC also inserts its *ingress link* in $\underline{\boldsymbol{X}}$ to further improve failure tracking. The *intra-domain* counter is also reset to allow a new set of retries in the upstream domain, i.e., h_1=H_1. Hence the number of crankbacks is limited to H_1H_2. The case of crankback re-computation is also shown in Fig. 1. For example, when ingress OXC v_1^i receives a *PATH-ERR* with h_1=0, it notes ingress $l_{21}^{i-1\ i}$as failed, prunes the route to the ingress border OXC in the prior domain, l_1^{i-1}, and sends a *PATH-ERR* to l_1^{i-1}(step 3, Fig. 1). This upstream OXC then re-tries path expansion to a new egress border OXC, l_3^{i-1}(step 4, Fig. 1).

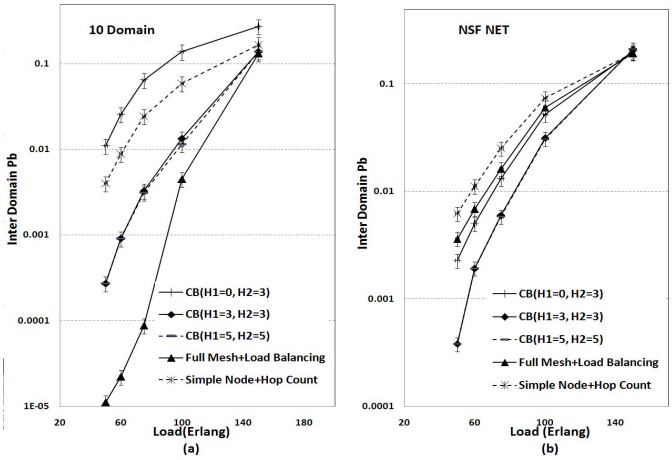

Fig. 2. Inter-domain blocking: a) 10-domain, b) NSFNET.

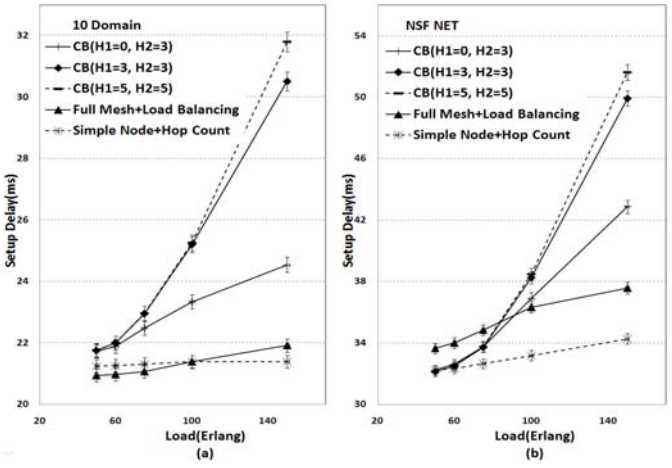

Fig. 3. Inter-domain setup delay: a) 10-domain, b) NSFNET.

B. Next-Hop Domain Selection

As mentioned earlier, the scheme also leverages available inter-domain routing state to help improve the overall crankback search process. Here all PCE entities pre-compute static multi-entry distance vector tables with up to K next-hop domains/egress links to each destination domain. To do this, a "simple node" [2] view of the global network is derived, ***H(U,E)***, where ***U*** is the set of domains$\{\boldsymbol{G}^i\}$ reduced to vertices and ***E*** is the set of inter-domain links $\{l_{km}^{ij}\}$, $i \neq j$. At the inter-area level, this graph can be obtained from hierarchical OSPF-TE databases whereas at the inter-AS level it can be deduced from border gateway protocol (BGP) path state (albeit all links may not be visible due to policy control). Using ***H(U,E)***, an iterative shortest-path algorithm is run to successively compute/prune multiple routes to all destination domains, 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 domain egress links. Hence crankback re-computation (as per Section II.A) simply searches these K entries to a destination domain, sequentially driving searches along increasing length domain sequences.

III. Performance Evaluation

Multi-domain crankback RWA is detailed models in OPNET $Modeler^{TM}$. Runs are done for a 10-domain topology with 25 inter-domain links as well as a modified NSFNET topology (i.e., nodes replaced by domains) with 16 domains/25 inter-domain links, i.e., average of 2.5 and 1.56 links/domain, respectively. In both cases, the domain size is set to 15 nodes and each link has 16 wavelengths. All requests are randomly generated between domains/nodes and have exponential holding times(mean 600 sec). Here a single run generates 500,000 requests and sufficient runs are done to achieve 95% of confidence. Also, K=5 next-hop domain entries are computed in the PCE multi-entry distance vector tables, although the number searched is limited to H_2. Finally, crankback RWA is compared against hierarchical link-state routing schemes which compute abstract topologies (simple node, full mesh) to summarize/propagate domain-level wavelength state, see [2].

Inter-domain lightpath blocking is first plotted in Figs. 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. Conversely, 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 NSFNET, i.e., lower inter-domain connectivity. These gains also come with much lower control plane complexities, as crankback message loads are over an order magnitude lower than hierarchical routing message loads (not shown). Next, inter-domain setup delays are plotted in Figs. 3a (10-domain) and 3b (NSFNET). Here link delays are set to 1 ms in the 10-domain network and to realistic distance-based propagation values in NSFNET (0.05 ms OXC processing). These results show that crankback yields higher setup delays at increased loads. Moreover, these increases are most visible in NSFNET due to its lower inter-domain connectivity, e.g., 30% higher delay for H_1,H_2=3 versus hierarchical routing. Nevertheless, these values are acceptable considering the long-standing nature of lightpaths.

IV. Conclusions

This paper presents a novel crankback solution for inter-domain RWA, generally matching hierarchical routing.

References

[1] M. Chamania and A. Jukan, "A survey of inter-domain peering and provisioning solutions for the next generation optical networks," *IEEE Commun. Surveys & Tutorials*, First Quarter 2009, vol. 11, Feb. 2009.

[2] Q. Liu, *et al.*, "Hierarchical inter-domain routing and lightpath provisioning in optical networks," *OSA JoN*, vol. 5, no. 10, pp. 764-774.

[3] M. Francisco, *et al.*, "End-to-end signaling and routing for optical IP betworks," *IEEE ICC 2002*, New York, June 2002.

[4] M. Yannuzi, *et al.*, "Toward a new route control model for multidomain optical networks," *IEEE Commun. Mag.*, pp. 104-111, June 2008.

[5] S. Dasgupta, *et al.*, "Path-computation-element-based architecture for interdomain MPLS/GMPLS traffic engineering: overview and performance," *IEEE Network*, vol. 21, no. 4, pp. 38-45, July/Aug. 2007.

[6] C. Pelssner and O. Bonaventure, "Path selection techniques to establish constrained interdomain MPLS LSPs," in *Proc. IFIP International Networking Conference*, Coimbra, Portugal, May 2006.