

Evaluation of a Combined Intra- and Inter-domain Constraint-based Routing Model for Optical Networks*

Marcelo Yannuzzi¹, Eva Marin Tordera¹, Sergio Sánchez-López¹,
Xavier Masip-Bruin¹, and Jesús Felipe Lobo Poyo²

¹ Advanced Network Architecture Lab, Technical University of Catalonia (UPC), Spain
{yannuzzi, eva, sergio, xmasip}@ac.upc.edu

² Telefónica I+D, Spain
jflp@tid.es

Abstract. This paper describes and evaluates the performance of a combined intra- and inter-domain routing model for optical networks. The model is supported by a set of network components called Inter-Domain Routing Agents (IDRAs). Each domain or Autonomous System allocates one or more of these agents, which are the ones responsible for computing and finding constrained inter-domain light-paths. Each IDRA exchanges with its peers two kinds of information, i.e, reachability information and network state information. In order to provide a highly scalable routing model, the network state information exchanged among the IDRAs is highly aggregated. In this particular work we introduce two opaque additive metrics which are used by the IDRAs to assess: i) the wavelength occupancy across domains; and ii) the resilience capabilities of downstream domains. Aggregating information at the domain boundaries aids in terms of overall scalability, but clearly, the frequency of this information exchange in order to keep the opaque metrics updated, needs to be tightly controlled. Our simulation results show that the opaque-aggregation scheme we are proposing in this paper is robust to the frequency of these updates.

Keywords. Inter-domain, constraint-based routing, optical networks

1. Introduction

The Border Gateway Protocol (BGP) is the de facto standard inter-domain routing protocol in the Internet. Its current release is version 4 (BGP-4), which was devised by Rekhter et al. in [1]. This version of the protocol was designed and implemented according to two fundamental principles. The first one was that it needed to be extremely scalable, and second, it needed to be capable of being configured by each domain in an independent manner, so as to reflect the different routing policies and the distributed nature of a large-scale routing system. The good news is that after eleven years, more than 21.000 domains or Autonomous Systems (ASes) in the Internet and nearly 180.000 entries in the routing tables of the BGP routers in the default-free zone, the protocol still

* This work was partially funded by the MCyT (Spanish Ministry of Science and Technology) under contract TEC2005-08051-C03-01, the European Commission through the projects IST-NOBEL under contract FP6-506760 and IST-ePhoton/One under contract FP6-001933.

scales [2]. The bad news is that the protocol is unable to handle most of the service requirements of the modern and the future Internet.

Among these requirements we can find, constrained-based routing, fault-tolerance and resilience, Quality of Resilience (QoR), traffic load balancing, multi-path routing, empowered Traffic Engineering (TE) capabilities, and much more. The main reason for this is that for scalability reasons, BGP-4 was devised as a routing protocol whose responsibility is basically the exchange of reachability information. Therefore, the only network “state” information exchanged between two BGP peers is that related to reachability issues, such as a route is up, or a route is down, or a shorter AS-path exists toward a certain destination prefix. Unfortunately, all the modern and future service requirements mentioned before demand the exchange of other kind of network “state” information. Particularly, the future optical-based networks will demand enriched information exchanges between the domains so as to be able to cope with these service requirements.

An option could be to extend BGP to the optical networks, so some researchers have proposed to adopt OBG (Optical Border Gateway Protocol) [3]. Even though this seems a highly scalable approach, it suffers from the inability to cope with the previous requirements. This is mainly because this inability is rooted in part on the design decisions taken while developing BGP, in part in its implementation decisions, and also in the routing paradigm itself. The future optical networks offer a clean path to devise novel routing paradigms and routing protocols, and we should take advantage of this. From our perspective, a promising line of work is the development of architectures which decouple the complexity of computing routes and signalling the enriched network “state” information from the devices forwarding traffic [4].

As a first step in this direction and from the years of lessons learned in BGP, we proposed in [5] a combined intra- and inter-domain routing model to support constrained-based routing at the inter-domain level in optical networks. The model is supported by a new set of network components called Inter-Domain Routing Agents (IDRAs). Each AS allocates one or more of these agents, which are the ones in charge of computing and finding constrained inter-domain light-paths taken into account highly aggregated and combined intra- and inter-domain network “state” information.

Favourably, the IETF has recently chartered a Working Group (WG) called the Path Computation Element (PCE) [6, 7], and the ideas that this WG is managing are aligned with the architecture and the core ideas presented in [5]. The goal of the WG is basically to develop a PCE, i.e. a device capable of computing G/MPLS TE-Label Switched Paths (LSPs) across domains subject to multiple constraints.

In [5] we mainly proposed the routing skeleton. Our goals in this paper are, first, to define a particular aggregation strategy so as to be able to compute paths constrained to:

- Certain connection call requirements (i.e. maximum likelihood of wavelength availability without wavelength conversion capabilities)
- Maximum resilience between the source and destination nodes

To accomplish this goal we define two opaque metrics, one related to the wavelength occupancy across the different domains, and the other related to the resilience capabilities of the those domains. We call them cost metric, and the resilience metric respectively. They are the outcome of an additive aggregation processes which extends from one domain to another.

Our second goal is to assess the performance of the combined routing model. We have contrasted its performance against a Shortest AS-Path Least Loaded (SPLL) inter-domain routing algorithm, under the assumption that it is able to fulfil the same set of constraints. Our simulation results show that the combined routing model performs much better than the SPLL routing algorithm. The results also show that the aggregation model is robust when we relax the frequency of the updates of the opaque metrics.

The rest of the paper is organized as follows. In Section 2 we briefly present the main concepts of the IDRA-based routing model. In Section 3 we present the cost metric computations and the strategy that we follow to aggregate this additive cost across the different domains. In Section 4 we introduce the resilience metric as well as the aggregation strategy that we follow. In Section 5 we present our evaluation results, and finally, Section 6 concludes the paper.

2. IDRA-based Routing Model

The combined routing model is based on the introduction of specialized agents called IDRAs. These devices are the glue between the intra- and the inter-domain routing schemes. These agents are able to enrich the network advertisements with combined intra- and inter-domain routing information targeting any peering agent, which blurs the current gap between the intra- and inter-domain routing protocols. These new kind of combined advertisements will allow any upstream domain to choose the next downstream domain for any given destination not only based on the inter-domain state of the network, but also based on the availability of intra-domain network resources of the different domains within the alternative paths to reach that destination.

As depicted in Fig. 1 the IDRAs exchange between them a vector containing two kinds of information. On the one hand, they exchange Network Reachability Information this is represented as NRI in the figure. On the other hand, they exchange network “state” information which comprises both the opaque cost metric C , associated with the wavelength occupancy, and the opaque metric R , associated with the resilience capabilities of the domains along the path. The opaqueness of these metrics is two-fold. First, because such selection supplies a highly scalable routing model, given that the information exchanged between the IDRAs belonging to different domains is highly synthesized. And second, because it allows providing network state information without disclosing intra-domain resource state information, something which is mandatory to gain acceptance among the Network Service Providers (NSPs).

In this sense, the metrics C and R associated with any destination known (given by NRI) are advertised as composite metrics by the IDRAs, wherein one part of them depends on the state of intra-domain resources, and the other on the state of inter-domain resources. It should become clear that the combined intra- and inter-domain routing protocol runs between the IDRAs, and each AS participating in our combined routing model has at least one IDRA. For resiliency and scalability reasons, clusters of IDRAs are allowed on each domain. The IDRAs may be directly connected by physical connections or logically connected, and those connections could be point-to-point or point-to-multipoint.

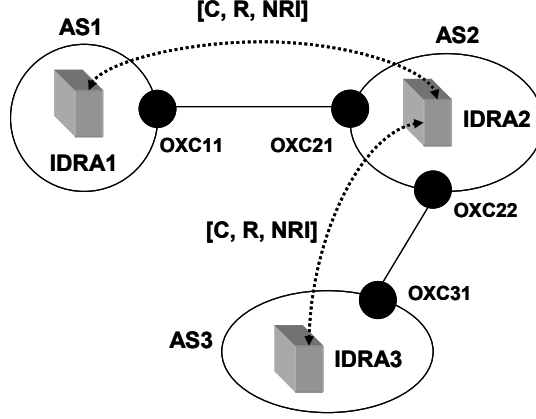


Fig. 1. IDRA-based routing model and information exchange.

3. Cost Metric Computations and Aggregation Scheme

The IDRAs are responsible for carrying inter-domain routing information, and deciding within each domain which path is the best among the ones available to reach any known destination. This decision is affected not only by the state of intra- and inter-domain resources within the traversed domains in those available paths, but also by the state of local intra-domain resources within the source domain.

3.1 Intra-domain cost computation

In terms of the cost metric, each domain will locally compute a cost C_{ID} , which is the intra-domain cost between any pair of nodes belonging to the domain. In particular, an IDRA within a transit domain needs to compute the cost C_{ID} between any pair of border nodes or Optical Cross Connects (OXCs). Formally, let the directed graph $G = (V, E)$ represent the optical network for a particular domain, wherein V is the set of OXCs and E is a set of links (i.e., fibres) connecting the OXCs. We denote by N and M the number of nodes and links respectively, i.e., $N=|V|$ and $M=|E|$. Let v_i represent a particular OXC, and (v_i, v_{i+1}) a particular link. Let $P(v_r, v_q, \lambda_j)$ represent an intra-domain light-path between the nodes v_r and v_q using wavelength λ_j . Assuming that each link (v_i, v_{i+1}) supports K different wavelengths from $\{\lambda_0, \dots, \lambda_{K-1}\}$, an IDRA computes the intra-domain cost C_{ID} in two steps. First it computes the effective number of available wavelengths of type λ_j for all the possible paths between the OXCs v_r and v_q as follows:

$$W(v_r, v_q, \lambda_j) = \max \left\{ \min \left(W((v_i, v_{i+1}), \lambda_j) \quad \forall (v_i, v_{i+1}) \in P(v_r, v_q) \right) \quad \forall P(v_r, v_q, \lambda_j) \right\} \quad (1)$$

The Equation (1) can be easily interpreted by means of the example in Fig. 2. Let us

5 Evaluation of a Combined Intra- and Inter-domain Constraint-based Routing Model for Optical Networks

consider the transit AS AS3 in the figure, and let $j = 0$. Thus, for this example the effective number of available wavelengths λ_0 between the nodes $OXC31$ and $OXC33$ is $W(OXC31, OXC33, \lambda_0) = 4$. This is due to the fact that from the two possible paths between the border nodes $OXC31$ (ingress node) and $OXC33$ (egress node), the path that goes through $OXC32$ has a minimum $\lambda_0 = 1$, whereas the one that goes through $OXC35$ has a minimum $\lambda_0 = 4$. Therefore, the maximum between both of them is 4. Once the IDRA has computed $W(v_r, v_q, \lambda_j) \forall (v_r, v_q)$, and $\forall \lambda_j$, it is able to associate an intra-domain cost C_{IA} to reach v_q from v_r using the wavelength λ_j . This is computed as follows:

$$C_{IA}(v_r, v_q, \lambda_j) = \begin{cases} MAX_INT & \text{if } W(v_r, v_q, \lambda_j) = 0 \\ \frac{1}{W(v_r, v_q, \lambda_j)} & \text{if } W(v_r, v_q, \lambda_j) \neq 0 \end{cases} \quad (2)$$

The number MAX_INT is the biggest integer number that can be represented by an IDRA, reflecting the lack of internal resources to handle connections between the nodes v_r and v_q for the particular wavelength λ_j . If this is the case, an IDRA will remove from the NRI field of its advertisements all the paths whose destinations made transit over the sub-path (v_r, v_q) for λ_j .

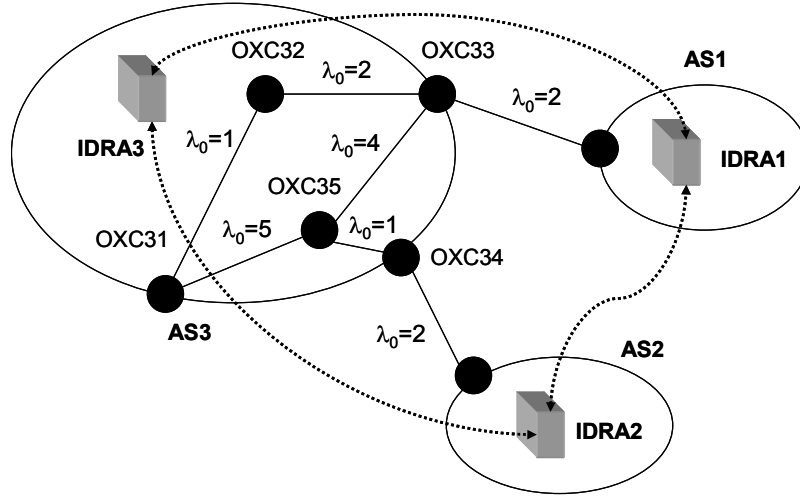


Fig. 2. An example showing how to compute the effective number of available wavelengths on an intra-domain path.

3.2 Combined cost computation

Once a domain has computed the cost C_{IA} , it needs two additional pieces of information in order to compute the end-to-end cost between any source-destination pair. These two pieces of information are: i) the aggregated inter-domain cost received from downstream peering IDRAs; ii) the local inter-domain cost associated with the wavelength occupancy on its own inter-domain links. Each IDRA knows which wavelengths are actually being used on its inter-domain links. Thus, for any given source s in its own domain and a given destination d in an external domain, an IDRA will compute the following cost associated with each wavelength:

$$C(s, d, \lambda_j) = H \left[C_{IA}(s, \bar{d}, \lambda_j) + C_{ID}(\bar{d}, \bar{d}_+, \lambda_j) + C_{ID}^{adv}(\bar{d}_+, d, \lambda_j) \right] \forall \lambda_j, j = 0, \dots, K, \forall \bar{d}_+ \quad (3)$$

- H : is the number of hops¹
- \bar{d} : is the border OXC within the domain toward the destination node d
- \bar{d}_+ : is a border OXC belonging to the next downstream domain in the path
- $C_{ID}(\bar{d}, \bar{d}_+, \lambda_j)$: is the cost associated with the inter-domain link between (\bar{d}, \bar{d}_+)
- $C_{ID}^{adv}(\bar{d}_+, d, \lambda_j)$: is the aggregated cost to reach node d as advertised by node \bar{d}_+

The rationale of (3) is that each IDRA computes a combined intra- and inter-domain cost based on its local visibility of resources, i.e., by computing C_{IA} and C_{ID} , and the visibility of its neighbors by means of the aggregate C_{ID}^{adv} . Moreover, for the same overall cost an IDRA prefers shorter AS-paths than longer AS-paths. This is simply because longer AS-paths increase the blocking probability of the connection requests. And reciprocally, for the same AS-path length, an IDRA prefers lower costs than higher costs. Indeed, the path and wavelength selected by an IDRA when the constraint is cost is the one that minimizes (3).

It is worth noting that an IDRA will never receive $C_{ID}^{adv} = MAX_INT$, since the IDRA issuing the advertisement appropriately removes the route from the NRI field.

Each IDRA determines $C_{ID}(\bar{d}, \bar{d}_+, \lambda_j)$ by computing the following:

$$C_{ID}(\bar{d}, \bar{d}_+, \lambda_j) = \begin{cases} MAX_INT & \text{if } W(\bar{d}, \bar{d}_+, \lambda_j) = 0 \\ \frac{1}{W(\bar{d}, \bar{d}_+, \lambda_j)} & \text{if } W(\bar{d}, \bar{d}_+, \lambda_j) \neq 0 \end{cases} \quad (4)$$

Where $W(\bar{d}, \bar{d}_+, \lambda_j)$ is the total number of available wavelengths λ_j taking into account

¹ Each intra-domain sub-path counts as only one hop

all the fibers connecting the border nodes (\vec{d}, \vec{d}_+) . The example in Fig. 3 should help to clarify the overall computation of costs by an IDRA. Suppose a light-path needs to be established between the source node S on AS1 and the destination node D in AS4. The IDRA in AS1 knows two different paths to reach the destination D . At the AS level the paths can be denoted by $P_1 = \{AS2, AS4\}$ and $P_2 = \{AS2, AS3, AS4\}$. In such a case BGP or OBGp typically prefer the shortest AS-path, so P_1 would be the one selected. However, from the point of view of AS2 the inter-domain link connecting the nodes $(\vec{d}, \vec{d}_+) = (OXC22, OXC41)$ shows that there is only one wavelength available by the time AS1 is going to start the connection request.²

The computation of the routes yields, $P_1 = 11$ and $P_2 = 10.092$. Thus, even though the path through AS3 implies one extra AS-hop the blocking possibilities are much lower when the connection setup avoids the link $(OXC22, OXC41)$. Clearly, the path P_1 would have been the one selected under a small increment in any of the costs of the path P_2 . By this assertion what we want is to stress the trade-off between the unscalable approach of having extremely detailed and constantly updated information (which is what is actually needed to avoid the problematic link), and a scalable approach based on highly synthesized information in the form of an aggregated and additive cost. In section 5 we will show the robustness of this approach in terms of the frequency of the updates between the IDRA's.

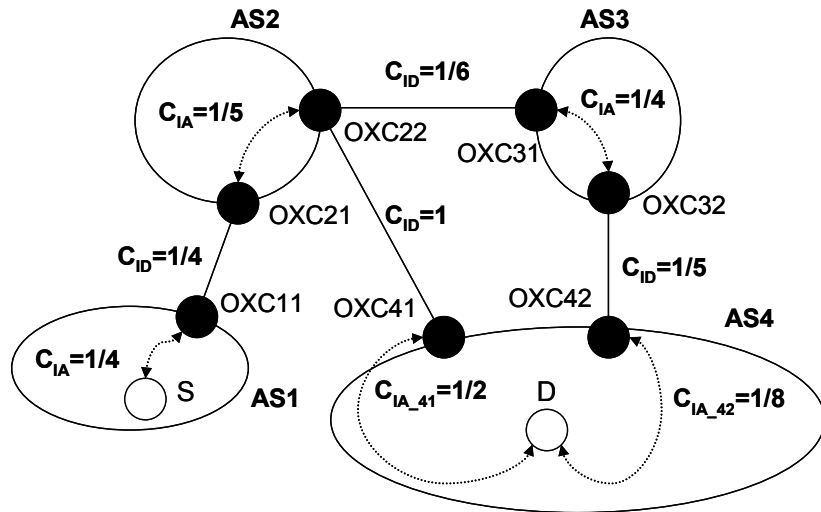


Fig. 3. Selecting the best path using the combined intra- and inter-domain cost.

² We assume that the figure depicts the lowest costs along each of the two different paths by the time of the call request. For the sake of simplicity, we assume the same wavelength minimizes the cost both in P_1 and P_2 (but in practice the wavelengths will be probably different)

3.3 Aggregated Cost advertisements

The cost to reach any given destination known by an IDRA is advertised as follows:

$$C_{ID}^{adv}(b, d, \lambda_j) = \left[C_{IA}(b, \bar{d}, \lambda_j) + C_{ID}(\bar{d}, \bar{d}_+, \lambda_j) + C_{ID}^{adv}(\bar{d}_+, d, \lambda_j) \right] \forall \lambda_j, j = 0, \dots, K, \forall \bar{d}_+ \quad (5)$$

In which b is the border OXC in the IDRA's domain which is actually issuing the cost advertisement in (5).

In the next section we propose a similar approach but for another constraint, that is, resilience across domain boundaries.

4. Resilience Computation and Aggregation Scheme

The idea is that the IDRAs could become capable of computing and exchanging an opaque parameter that we call Resilience parameter (R), which could represent the concept of a risk of a domain. We propose a straightforward mechanism which consists of a reduced amount of TE information to be carried in the IDRA's protocol. We will make use of this opaque metric so as to roughly assess the protection and restoration capabilities of a domain. Indeed, this metric can be used as one of the multiple constraints in a constrained path selection problem.

When an IDRA is asked to compute a path with certain protection requirements, it first consults the resilience information stored in its TE database, which was basically build-up by the R field in the state advertisements received from its peers. Next, the IDRA needs to find the resilience capabilities of its own domain and decide if it is able to locally cope with the connection demands. Finally, it needs to choose the appropriate combination between the intra- and inter-domain resilience capabilities.

Let $R_{IA}(v_r, v_q)$ denote the number of available disjoint intra-domain paths for working and protection between the border OXCs v_r and v_q in a particular domain. We propose to use the following resilience metric:

$$R_p(s, d) = \frac{1}{H} \sum_{k=1}^{k=H} \hat{w}_k R_{IA,k}(p, (s, d)) \quad \forall p / p = (v_0 = s, \dots, v_n = d) \quad (6)$$

$$R(s, d) = \max_p \{ R_p(s, d) \} \quad \forall p / p = (v_0 = s, \dots, v_n = d)$$

Let $R_p(s, d)$ denote the mean of all the $R_{IA,k}(v_r, v_q)$ for a particular path p . As before, H represents the number of hops, and \hat{w}_k is the weight associated with the k^{th} domain. This weight can be used to distinguish different resilience techniques or just to assign

priorities to a set of particular domains. Now we can define the resilience $R(s,d)$ between distant domains as the maximum of $R_p(s,d)$.

5. Simulation Results

The simulations presented in this paper were performed under the following requirements. Always choose the path with the lowest cost C , and in case than two or more paths exhibit the same cost, choose among these the one with the highest resilience metric R .

We have used a multi-domain topology consisting of five domains with multiple connections between them such as in [8]. Each domain has at least two disjoint paths to reach any given node inside it. We have conducted a number of trials using different updating policies ($N=5, N=10, N=20, N=200$) between the IDRAs. Fig. (4) shows the comparison between the Shortest As-Path Least Loaded and our Cost-Resilience routing proposal. From the figure two different things become clear. First, our proposal performs always much better than the SPLL algorithm, in some cases even one order of magnitude better. And second, that the proposal is indeed robust to the triggering policy utilized for the updates, since there are no significant variations in the blocking ratio as the inaccuracy of the data on the IDRAs databases increases.

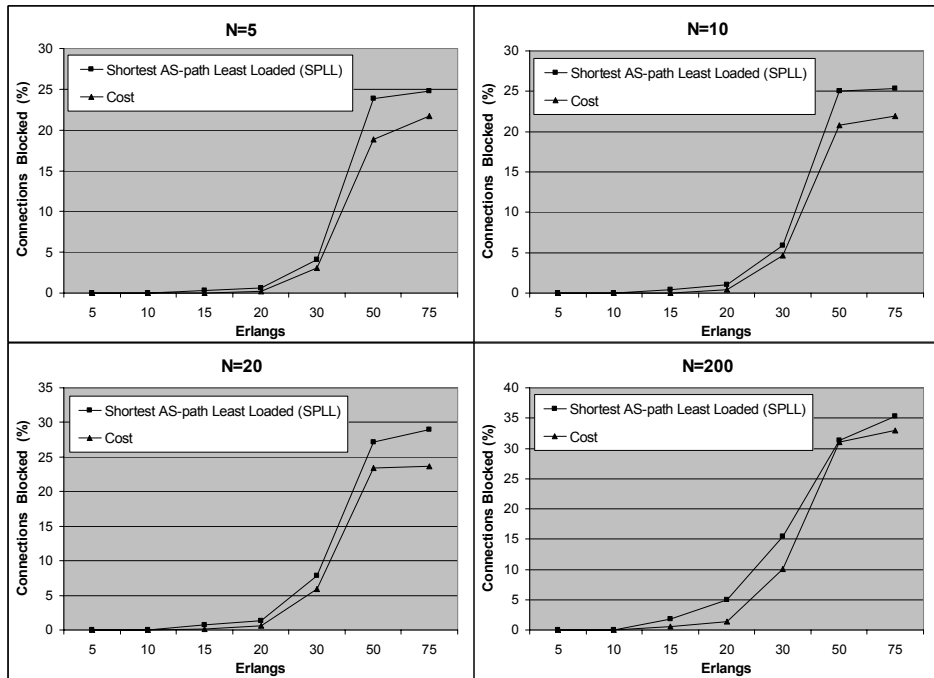


Fig. 4. Connections Blocked Ratio (in %) for our combined intra- and inter-domain routing model and for the SPLL algorithm

6. Conclusions and Future Work

In this paper we introduced and evaluated two opaque metrics and an aggregation scheme to straightforwardly synthesize network state information regarding the wavelength occupancy and resilience capabilities of the domains along the Inetrnet. The metrics as well as the aggregation scheme we are proposing in this work, become now part of a combined intra- and inter-domain QoS framework that we devised in a previous work.

We have assed the performance of the routing scheme and we have found that it always performs much better than an extended version of the Shortest AS-Path Least Loaded routing algorithm, under the same set of constraints, and the same traffic demands. Moreover, we have empirically obtained enough evidence supporting the robustness of the solution given that large variations in the frequency of the updates do not introduce tangible differences in the connections blocking ratio.

Our future work will mainly consist of investigating the feasibility of introducing a stochastic model for the wavelength occupancy across multiple domains and based on this model analyze the possibilities of forecasting and filtering the wavelength occupancy along inter-domain light-paths.

References

- [1] Y. Rekhter, T. Li “A Border Gateway Protocol 4 (BGP-4),” Internet Engineering Task Force, Request for Comments 1771, March 1995.
- [2] CIDR report, July 2005: <http://www.cidr-report.org/>.
- [3] L. Wang, H. Zhang, X. Zheng, et al. “A Novel OBGp-based Mechanism for Lightpath Establishment in WDM Mesh Networks,” in Proceedings of European Conference on Optical Communication, (Academic, Rimini, Italy, 2003), We4.P.136, vol. 3, pp. 828-829
- [4] N. Feamster, H. Balakrishnan, J. Rexford, A. Shaikh, and J. Merwe, “The case for separating routing from routers,” in Proc. ACM SIGCOMM FDNA Workshop, 2004.
- [5] M. Yannuzzi, S. Sánchez-López, X. Masip-Bruin, J. Solé-Pareta and J. Domingo-Pascual, “A Combined Intra-Domain and Inter-Domain QoS Routing Model for Optical Networks,” IFIP/IEEE ONDM 2005, Milan, Italy, 2005 .
- [6] A. Farrel, J. P. Vasseur, J. Ash: Path Computation Element (PCE) Architecture Internet draft, draft-ietf-pce-architecture-00.txt, work in progress, March 2005.
- [7] J. Ash and J. L. Le Roux: PCE Communication Protocol Generic Requirements Internet draft, draft-ietf-pce-comm-protocol-gen-reqs-00.txt, work in progress, May 2005.
- [8] E. Marin-Tordera, X. Masip-Bruin, S. Sánchez-López, J. Solé-Pareta, J. Domingo-Pascual, “A Hierarchical Routing Approach for Optical Transport Networks,” to appear in the Computer Networks Journal, Elsevier.