SEVENTH FRAMEWORK PROGRAMME

Project Number: FP7-ICT-2007-1 216863

Project Title: Building the Future Optical Network in Europe (BONE)

CEC Deliverable Number: FP7-ICT-216863/UC3M-UPC/O/PU/D22.3

Contractual Date of Deliverable: 31/12/09

Actual Date of Delivery: 25/12/09

Title of Deliverable: D22.3: Final Report on TP Activities

Workpackage contributing to the Deliverable: WP22 : Topical Project on MPLS, GMPLS and routing

Nature of the Deliverable Make your choice from: R

Dissemination level of Deliverable Make your choice from: PU

Editors: UPC / Salvatore Spadaro
UC3M / David Larrabeiti

Abstract:

This deliverable describes the activity carried out during in the framework of the WP22 “MPLS, GMPLS and routing”. The main results achieved by the different Joint Activities are presented and discussed.

Keyword list: MPLS, GMPLS, routing, signalling, all-optical networks, Path Computation Element (PCE), BGP, multi-domain, QoT.

Clarification:

Nature of the Deliverable

R Report
P Prototype
D Demonstrator
O Other

Dissemination level of Deliverable:

PU Public
PP Restricted to other programme participants (including the Commission Services)
RE Restricted to a group specified by the consortium (including the Commission Services)
CO Confidential, only for members of the consortium (including the Commission Services)

Disclaimer

The information, documentation and figures available in this deliverable, is written by the BONE ("Building the Future Optical Network in Europe*) – project consortium under EC co-financing contract FP7-ICT-216863 and does not necessarily reflect the views of the European Commission*

Table of Contents

1. Executive Summary

This document is the final deliverable of the Work Package 22 "*Topical Project on MPLS, GMPLS and routing*". The overall purpose of WP22 was to research on key issues in the evolution of IP-MPLS multi-service networks to all-optical networks. In line with this overall objective, eleven Joint Activities (JAs) were defined with the participation of eighteen different partners. These JAs mainly dealt with open research topics on MPLS/GMPLS networks, such as path computation (based on Path Computation Element and on BGP extensions), multi-domain/multi-layer optical networking issues (for example multi-domain recovery and Traffic Engineering), Quality of Transmission (QoT)-aware signalling, Routing and Wavelength Assignment (RWA) and resilience issues in GMPLS networks, the extensions of GMPLS control stack to packet switched networks and the optimization of bidirectional signalling.

1.1 Participants

In the framework of the WP22, there have been 18 partners collaborating in the 11 Joint Activities defined. Table 1 show the list of the participating partners with the number of the Joint Activities in which they have been involved.

Partner number	Partner name	Joint Activities
1	IBBT	8
7	UST-IKR	1
8	COM-DTU	2, 6, 9, 10
9	CTTC	1, 3, 6, 8
12	UC3M	2, 4, 5
13	UPC	4, 7, 8, 9
14	UPCT	6
17	Orange Labs	6, 8
19	AIT	11
21	RACTI	6
24	BME	2, 5, 7
27	FUB	4, 5, 6
31	SSSUP	3, 6, 10, 11
32	DEIS	9
38	AGH	1,2, 4, 7, 8, 9
41	KTH	6
43	UNIROMA3	9
47	UEssex	10

Table 1: Work package participants and their joint activities

1.2 Publications

The different Joint Activities achieved several scientific contributions in terms of papers that have been accepted. Specifically, right now 39 papers have been published in the framework of the WP22 activities; 15 of them are joint papers. At the end of the description of each JA, the achieved publications are listed.

1.3 Mobility Actions

Some mobility actions were promoted among the participants partners. Next, a list of the mobilities carried out in the last year is presented:

- In the framework of JA2 "*BGP extensions for inter-domain TE in transport networks*", the following mobility actions were carried out:
  - Anna Manolova (DTU) hosted by UC3M: November 2008.
  - Ricardo Romeral (UC3M) hosted by DTU: September 2009.

- In the framework of JA11 "*Monitoring for GMPLS Control Plane in Optical Networks*", the following mobility action was carried out:
    - Nicola Sambo (SSSUP), hosted by AIT: from 04/01/2009 to 21/01/2009.

1.4 Meetings

Apart from the participation in the project plenary meetings and in the WP11-WP12-WP21-WP22-WP24-WP26 joint meeting held in Bologna on June 8-9, 2009, several face-to-face (conferences) and remote (conference-calls) meetings were held among the participants of the JAs in order to define objectives, to discuss the methodology/results and, finally, to prepare joint papers.

1.5 Summary of Joint Activities

The Joint Activities (JAs) defined within the WP22 were the following (the partner responsible for each JA is highlighted in boldface). A detailed description and the main achievements of the JAs are reported in the following Section 2.

JA number	JA title	Involved Partners
1	Scalability of Path Computation Elements (PCE)	**CTTC**, UST-IKR, AGH
2	BGP extensions for inter-domain TE in transport networks	**COM-DTU**, UC3M, BME, AGH
3	QoT-Aware GMPLS Control Plane	**SSSUP**, CTTC, Orange Labs
4	MPLS-ASON/GMPLS Interconnection	**UPC**, UC3M, AGH, FUB
5	Scalability issues in G/MPLS-based VPLS network design	**UC3M**, FUB, BME
6	GMPLS-based RWA algorithms for optical protection/restoration	**CTTC**, SSSUP, UPCT, FUB, RACTI, AIT, KTH, COM-DTU
7	Resilience Issues in the GMPLS-enabled Control Plane	**UPC**, AGH, BME
8	Multi-domain provisioning/recovery within GMPLS all-optical networks	**CTTC**, UPC, Orange Labs, IBBT, AGH
9	GMPLS-based control plane for optical packet-based technologies	**UPC**, DEIS, UNIROMA3, AGH, COM
10	Bidirectional service signalling in GMPLS networks	**SSSUP**, DTU Fotonik, UESSEX
11	Monitoring for GMPLS Control Plane in Optical Networks	**SSSUP**, AIT

Table 2: Summary of the Joint Activities

2. Joint Activities description

This Section reports the description and the main results achieved by the different Joint Activities.

2.1 Scalability of Path Computation Elements (PCE)

The IETF Path Computation Element (PCE) working group has defined the architecture and a communication protocol (PCEP) so Path Computation Clients (PCCs) may request the computation of an explicitly routed path given a set of constraints. Such an architecture is motivated by the complexity of path computation in large, multi-domain, multi-region, or multi-layer networks, and that of advanced (e.g. protection-enabled) algorithms and heuristics, which may eventually require dedicated computational resources and cooperation between network domains.

The joint activity "Scalability of Path Computation Elements (PCE)" deals with the application of PCE in transport networks in general, starting with wavelength switched optical networks (WSON) with a GMPLS control plane and ending up covering more complex scenarios such as transparent optical networks with shared path protection, OSNR-aware translucent networks, and multi-layer networks.

In the GMPLS architecture, label switched routers (i.e., optical connection controller, OCC, in WSON) have full topology visibility within their domain boundaries and limited visibility of the other domains, usually as aggregated information (e.g., reachability).

Consequently, in traditional source routing approaches, a source OCC is not able to compute, autonomously, an end-to-end inter-area path with the same control and degree of TE as for an intra-area path. In this context, two methods are applicable for inter-domain path computation, the per-domain path computation method and the path computation element (PCE)-based path computation method. The latter method assumes that a domain chain (succession of transit TE domains from source to destination) is known in advance. The method relies on dedicated PCEs, which collaboratively compute an inter-domain optimum path along the given domain chain. Each PCE is responsible for the path computation within its domain.

The methodology for the work in the JA is clearly biased towards the study by means of simulations and experimental research (prototypes), with final interoperability-focused experimentation and assessment. In this line, CTTC IETF PCE and UST-IKR Dragon Network Aware Resource Broker (NARB) implementations have played a fundamental role in the activity, ranging from feasibility and "proof-of-concept" deployments to in-depth scalability studies of these PCE implementations.

The list of expected results defined for the joint activity covers several aspects: from actual PCE implementation(s) by the different partners, to scientific publications covering the evaluation of implemented path computing algorithms with respect to scalability parameters along with the experimental validation of PCE-based extensions, including a PCE-to-PCE gateway for interoperability.

From a macroscopic point of view, the first year (Y1) was clearly focused on specific developments, implementing the software applications and focusing on basic functionality and infrastructure and first applications. On the contrary, the second year (Y2), although still open for new additions, has clearly focused on interworking and inter-operability, including testbed interconnection and experimental activities spanning several domains.

For clarity, the Joint Activity itself has been divided into specific tasks, each one targeting one concrete problem or research item. The set of tasks is presented below, roughly corresponding to the sequential work done within the JA.

2.1.1 Task I: PCE for WSON with SPP and DRAGON NARB for Ethernet switched

The first task covered the deployment of CTTC PCE implementation and IKR SDRAGON NARB system. This covered the actual development along with preliminary tests and validations, along with first actual applications such as the study if the PCE in WSON with enabled Shared Path Protection.

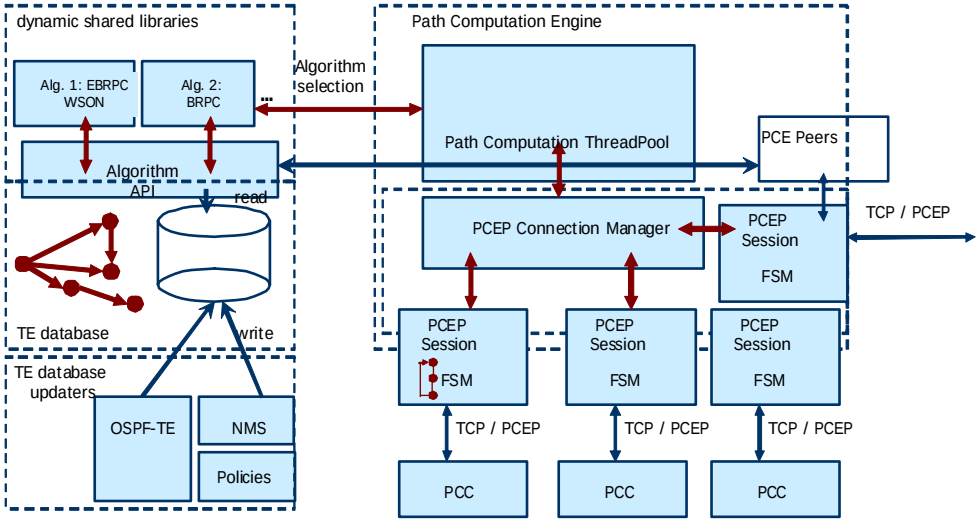


Figure 1: CTTC high level architecture

The Figure 1 shows a high level overview of CTTC PCE. The PCE implementation is a multi-threaded process, with an asynchronous design. The main thread focuses on the management of PCC clients, by means of a PCEP connection manager. The other threads are either worker threads for path computation forming a pool of resources or dedicated threads that build the Traffic Engineering Database (TED) by means of, for example, passive OSPF-TE listening of opaque TLVs.

This task led to publications [Cas08a] and [Cas08b], showing the application of the PCE in WSON with Shared Path Protection.

2.1.2 Task II: Enhanced BRPC with WCC

All-optical networks raise well-known challenges such as the wavelength continuity constraint (WCC). The WCC is hard to address in a multi-area scenario when provisioning an end-to-end lightpath owing to network topology hiding requirements and the limited exchange of information between areas.

In such architecture, the approach named backwards recursive path computation (BRPC), also under standardization at the IETF, aims at overcoming the limitations of the per-domain mechanism. However, although BRPC does provide end-to-end shortest paths, it fails to take into account the WCC, which is the main motivation for this work. We extend the BRPC algorithm and the companion PCE protocol in order to address the end-to-end WCC efficiently. We perform a quantitative comparative analysis of the different approaches,

experimentally showing the improvements of the conceived solution, which has been evaluated in a GMPLS-controlled network of the ADRENALINE testbed.

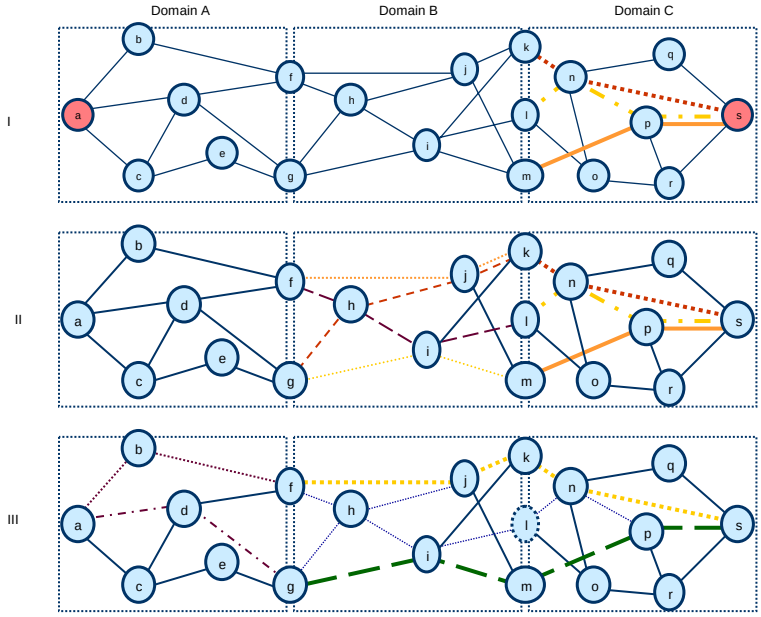


Figure 2: BRPC overview

As shown in Figure 2 the BRPC mechanisms involves the en-to-end path computation by computing a set of optimum paths starting from the destination domain, backwards towards the source domain. In the process, each PCE computes a set of optimum paths from its own domain entry points towards the destination, using the previously received paths from the downstream PCE, thus insuring end-to-end optimum (shortest) paths. In the Figure, PCE within domain C computes optimum paths from k to sm l to s and m to s. These paths are used by PCE in domain B to compute optimum paths from f to s and from g to s.

We focused on the mechanisms for lightpath provisioning in multiple-area all-optical networks. We implemented, deployed, and experimentally validated three different path computation algorithms: the “per domain”, the PCE-based BRPC, and our proposed PCE-based EBRPC approach that efficiently addresses the WCC. Topology confidentiality is one major requirement from carriers, and providing enough information in order to ensure the wavelength continuity constraint does not necessarily mean that topology confidentiality cannot be preserved. The proposed extension does convey information on the number and identification of the wavelengths that are available end to end, but does not explicitly disclose information on link bandwidth, and it does not preclude the (optional) use of some policing at the domain edges. We quantitatively evaluated the key performance indicators such as the blocking probability and the setup delay.

As expected, the per-domain method shows on average the smallest path setup delay, providing robustness in front of very high traffic dynamics. However, it also shows the highest blocking probability because it is constrained to a given entry and exit boundary nodes and thus it is unable to find the shortest feasible end-to-end path and because it has limited visibility to take into account the WCC during the ERO expansion. We have also shown that the well known BRPC, conceived to allow the computation of an end-to-end (shortest) path in the presence of multiple ABR nodes in the network, fails to capture the

WCC constraint present in all-optical transparent networks, since the information regarding wavelength availability is lost in the VSPT processing.

The devised extended BRPC, motivated by the specific requirements of WSON under WCC, minimizes blocking due to WCC while still computing optimal end-to-end paths, outperforming the other two without significantly impacting scalability (in terms of additional control plane required bandwidth or latency).

Both PCE-based approaches come at the cost of a higher path setup delay due to the increased path computation latency, and at the cost of additional path computation entities and control plane extensions. It is noteworthy that the proposed extension to BRPC is relatively negligible in terms of path computation (the PCE executes the path computation algorithm in a few milliseconds) and only extends the PCEP Reply message marginally, having no noticeable impact on a dedicated control network of 100 Mbit/s Ethernet based control channels.

The main result of this task is the joint publication [Cas09a].

2.1.3 Task III: OSNR-aware Multi-Domain Path Computation

In this task, we present a multi-domain PCE architecture including an OSNR-aware algorithm, successfully deployed between geographically separated, to dynamically provision end-to-end lightpaths in multi-domain GMPLS-controlled translucent WSON networks. We also demonstrate its feasibility in the experimental field trial, highlighting the path computation latency, while assessing its overall feasibility.

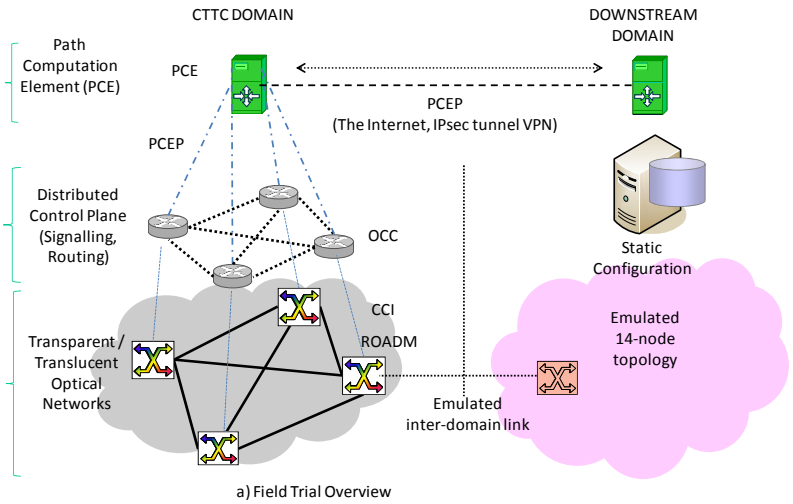


Figure 3: PCE-based path computation in OSNR-aware Multi-domain networks, field trial

As shown in Figure 3, both PCEs were connected using an IPsec tunnel. End-to-end path computations were requested and computed successfully, starting from any node within the upstream domain and ending in any node within the downstream domain (which, for practical reasons was emulated, contrary to the real testbed in the upstream domain).

The field trial was presented during ECOC 2009 in Vienna, see [Cas09b].

2.1.4 Task IV: Multi-layer routing

The main goal of this part of activity is to propose and evaluate several routing strategies in a multilayer, single domain network. Here, the proposed routing strategies are based on centralized Path Computation Element (PCE). The PCE gathers information about the state of the network; it processes the information and selects a path for a given request. The arriving requests are not known in advance, i.e., the fully dynamic scenario is assumed. The PCE's task is to find a path for a request in a way which allows guaranteeing transport of data with given throughput and delay. It is assumed that requests may be categorized into three categories: with high, medium and low required throughput. The aim of the PCE is to find a route for a request in given network. The physical topology of the reference network is shown in Figure 4. It is assumed that in the physical layer of the reference network each link consists of two fibres, one in each direction. Each fibre may be used to establish a lightpath. The number of wavelengths per link is constant in the network. It is assumed that no wavelength converters are used. The usage of Optical Transport Network (OTN) and the circuit-switching paradigm is considered.

On the top of the optical layer, the electric layer is placed. At the electric layer, the most promising technology seems to be Next Generation Ethernet (NG Ethernet). The IEEE 802.1ad standard allows using two tags devoted to VLANs. The IEEE 802.1ah additionally enhances the Ethernet by introduction MAC-in-MAC functionality and the possibility to further differentiate offered services by I-SID, i.e., the service identifier. The IEEE 802.1ag is devoted to OAM. One of the most important and revolutionary issue of NG Ethernet is possibility to disable address learning and forwarding frames with unknown address. The proposed IEEE 802.1Qay standard enables to explicitly select path for given data. The work on NG Ethernet is still ongoing; however, some new standards are expected soon.

For the work on scalability of PCE the OMNeT++ was chosen and used along with the Boost Graph Library. The proper simulation environment was build and tested. Subsequently, we analyzed some routing strategies, which are based on proposed heuristics and evaluated through simulation studies.

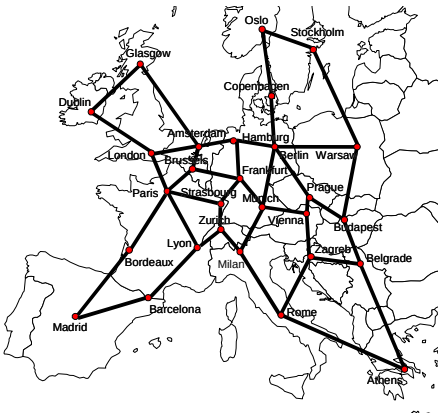


Figure 4: The reference network used in the study

It is assumed that the PCE, for a given request, first tries to select path from source to destination without any transit nodes. If there is no direct path from source to destination or there is no enough capacity on the existing direct paths, then the path with one transit node is chosen. If there is no path with one transit node or there is no enough free capacity on any candidate path from source to destination with one transit node, then request for new lightpath

for given pair source-destination is sent to the optical layer. If such optical path is established then the request is routed through such route. Otherwise, the request is blocked.

As it can be noticed (Figure 5, Figure 6 and Figure 7), the blocking probability is highly dependent on the demanded granularity. The notation shown in the Figures i.e., 0A_0B_0C denotes percent of requests for low (A), medium (B) and high granularity (C). For example, the 050_030_020 indicates that low granularity demands account for half requests, medium granularity requests account for 30% of all requests, and high granularity requests account for 20% of all demands.

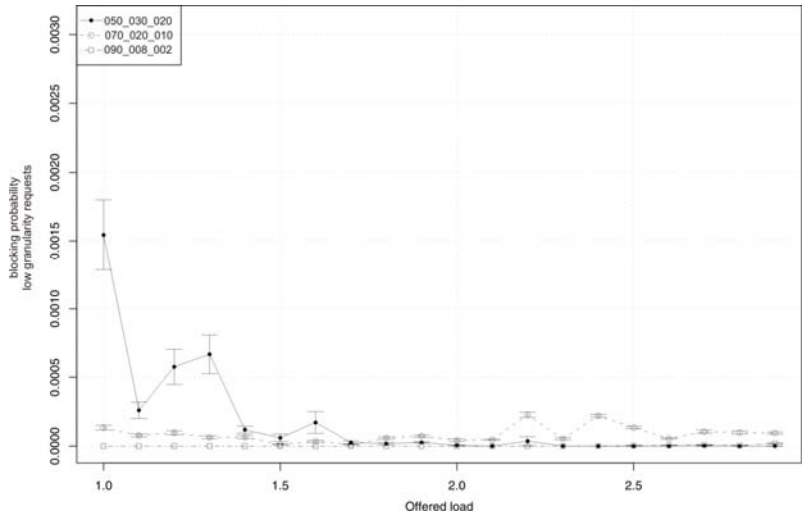


Figure 5: The blocking probability for low granularity requests

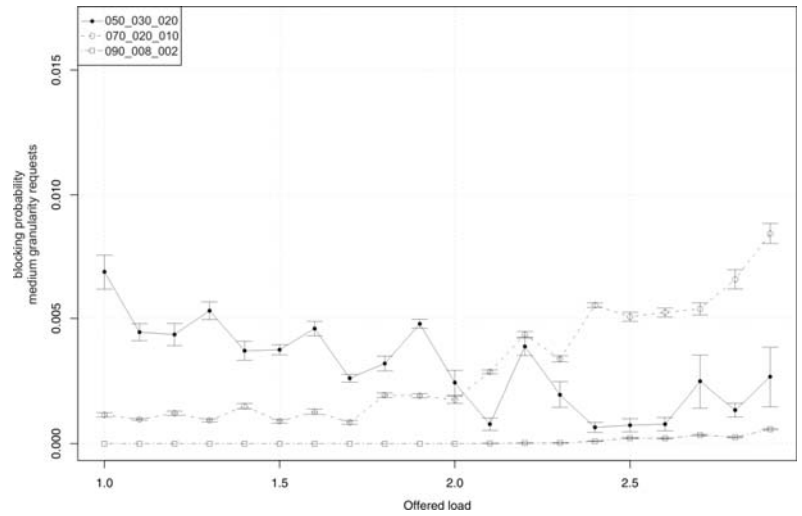


Figure 6: The blocking probability for medium granularity requests

For the whole range of offered load, the blocking probability of low and medium granularity requests is low. On the contrary, the blocking probability of high granularity requests for broad range of the offered load is unacceptable.

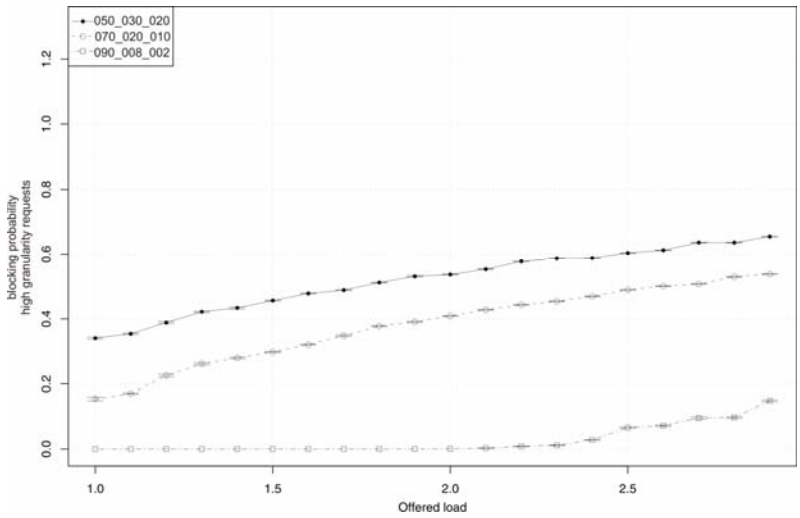


Figure 7: The blocking probability for high granularity requests

In the next step, we plan to analyze other strategies to keep the blocking probability of high granularity demands on low level without deterioration of blocking probability low and medium granularity requests.

2.1.5 Task V: Experimental Validation and Assessment of Multi-domain and Multi-Layer path computation with PCE-NARB interworking

The last task deals with the specific realization of interoperability tests, deployment of a multi-layer scenario for the Interworking of Path computation architectures and protocols including translation of request/response messages PCE to NARB and NARB to PCE. The task has spanned different activities:

- The definition of testbed connectivity requirements, deployment (IPSec, VLAN, etc) along with generic testbed verification and maintenance.
- The actual realization of tests, obtaining experimental results for the proposed scenario.

In short, we setup a multi-domain multi-layer testbed covering three different networks at two distinct locations in Europe. The testbed includes two Ethernet switched client networks, which are interconnected by a wavelength switched server network. Each of these networks runs a GMPLS based control plane and implements a path computation entity, either following the IETF PCE proposal or the DRAGON NARB.

Since their respective communications protocols are not compatible, we propose and develop an application layer gateway, enabling inter-domain path calculation.

Our contributions are the three-fold: first, we provide a comparison of both communication protocols; second, we present the architecture and working principles of the designed NARB/PCE Gateway, specifying the available features and constraints of our implementation.

Third, we validate, for the first time, the PCE/NARB connectivity while evaluating the performance of a path computation request in terms of request response time in the multi-domain and multi-layer testbed.

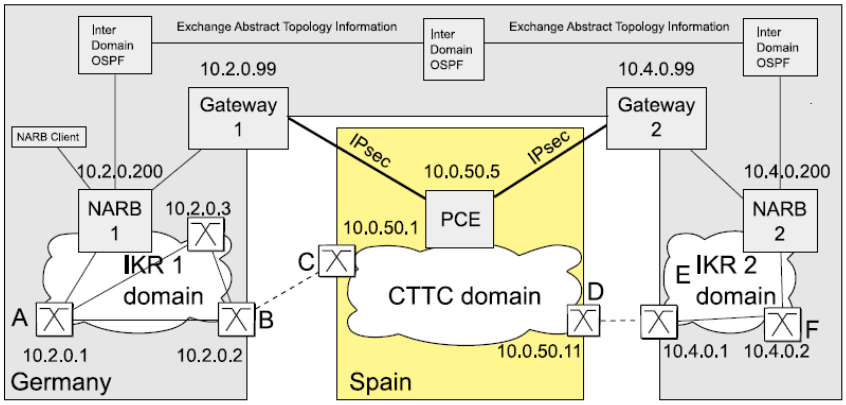


Figure 8: IKR-CTTC interoperability test

This work has been submitted to Tridentcom 2010 as a joint paper, see [Gun09].

2.1.6 Conclusions

In this joint activity, driven by requirements of complex path computation in multi-layer, multi-domain or protection enabled networks, we have worked on a set of tasks covering the design, implementation and assessment of PCE (or NARB DRAGON equivalents), serving path computation requests such as the computation of working / protecting path pairs or paths with OSNR-constraints. Additionally, we have started preliminary work on the scalability of the PCE. In particular, the last steps were focused on defining analytical models, which also left some questions open.

The particular problem domains of the application of Path Computation Elements in multi-layer and multi-domain networks has proved to be of particular interest, opening a wide range of specific issues, with plenty of room for new architectures and extensions.

We believe that the Joint Activity has met the original goals that were set when it was defined. It has successfully combined development, experimentation and model validation on a key subject and hot topic. We have covered all the targeted scenarios and results have been published in relevant conferences and journals.

2.1.7 List of publications

- R. Casellas, R. Martínez, R. Muñoz, "*Design, implementation and validation within ADRENALINE® testbed of a Path Computation Element for Wavelength Switched Optical Networks*", in Proc. 4th International Conference on IP over Optical (iPOP2008), Tokyo (Japan), June 2008.
- R. Casellas, R. Muñoz, R. Martínez, "*A Path Computation Element for Shared Path Protection in GMPLS-enabled Wavelength Switched Optical Networks*", in Proc. 34th European Conference on Optical Communications, ECOC2008, Brussels, (Belgium) September 20-25 2008.
- R. Casellas, R. Martínez, R. Muñoz, S. Gunreben, "*Enhanced BRPC for multi-domain PCE-based path computation in Wavelength Switched Optical Networks under Wavelength Continuity Constraint*", Journal of Optical Communications and Networking (JOCN) Vol. 1, No. 2 pp. A180-A193, ISSN: 1943-062, July 2009.

- R. Casellas, R. Muñoz, R. Martínez, "*Experimental Field-Trial of Multi-domain PCE-based Path Computation for OSNR-aware GMPLS enabled translucent WSON*", in Proc. of 35th European Conference on Optical Communications, ECOC2009, September 2009.
- S. Gunreben, R. Casellas, R. Martínez, R. Muñoz, J. Scharf, "*Experimental Validation and Assessment of Multi-domain and Multi-layer Path Computation*", submitted for publication to the 6th conference on Testbeds and Research Infrastructures for the Development of Networks and Communities (Tridentcom 2010) 18-20 May 2010 - Berlin, Germany.

2.2 BGP extensions for inter-domain TE in transport networks

2.2.1 Objective

The objective of this JA is to investigate theoretically and via simulations the efficiency of different extension of the BGP protocol in support of multi-domain Label Switched Path establishment with QoS and /or resilience provisioning. The activity focuses on identifying possible BGP extensions for support of interconnected ASON/GMPLS networks, as well as elaborates on policy-exchange mechanisms which support inter-domain TE and QoS provisioning across transport networks.

2.2.2 Description

The JA is divided in two main parts. The first part focuses on the modification of the BGP for achieving different TE capabilities such as increased survivability and/or QoS support. This part is related to the dynamics of the network operation and thus the efficiency of the proposed algorithms and BGP modifications are to be shown experimentally via simulations. The second part of the activity focuses on different political aspects of TE-oriented interconnections between transport networks. Different export policies as well as their effect on the QoS provisioning are to be investigated. This part of the activity focuses on algorithms facilitating the design of networks and thus the used methodology will be applying and designing mathematical models, focused on the optimization of different aspects of the design and operation of the multi-domain networks.

2.2.3 Current results

The following specific topics have been covered by the JA:

1. BGP modifications for end-to-end TE support and multi-path dissemination for survivability support (COM DTU).
2. BGP modifications for AS disjoint path dissemination for survivability support (US3M, COM DTU).
3. Genetic algorithms for efficient inter-domain traffic distribution (AGH)
4. Managing inaccurate advertisements by penalty methods in multi-domain networks (BME).

2.2.4 BGP modifications for TE support (COM DTU)

The implementation of the BGP modifications has been done in an event driven simulator (OPNET), which is flexible and can be used for evaluation of different BGP modifications in real network environments (i.e. dynamic not static environments). The BGP modeler is complemented with an RSVP-TE model which combined with the implemented BGP engine can be used for evaluation of the real effect of different modifications on the operation of a multi-domain GMPLS network and the efficiency of the routing protocol itself.

A novel extension to the BGP protocol has been suggested, implemented and tested. The main objective is automatic TE support across multiple dynamic GMPLS domains. Two main BGP enhancements have been proposed: disseminating end-to-end TE information per path (referred to as *TE attribute*) and disseminating additional path attribute, specifying the border nodes along the path (referred to as *Border_node sequence*). Furthermore, the suggestion for using the BGP protocol only as a dissemination protocol and not as a path selection one has been evaluated. Implementing all suggested modifications is referred to as ***Enhanced BGP***. Under this scheme, multiple paths per destination, each augmented with an end-to-end TE attribute, are distributed in the multi-domain environment. Figure 9 illustrates the operation of the ***Enhanced BGP*** protocol. Instead of performing Path selection, border nodes B1 and D1 only disseminate further all policy-compliant paths to their neighbours (nodes A1 and A2). Thus, after path dissemination the source node obtains multiple paths towards the advertised destination, each with an end-to-end TE metric, and builds a Virtual Topology representing the multi-domain connectivity with respect to the disseminated destination. At time of LSP request the source node chooses the path with the best TE metric (which is periodically updated), and uses the provided Border_node sequence as a loose hop ERO for the RSVP-TE PATH message. The design of the Enhanced BGP aims to support load balancing, reduced LSP blocking and survivability support by providing multiple paths per destination.

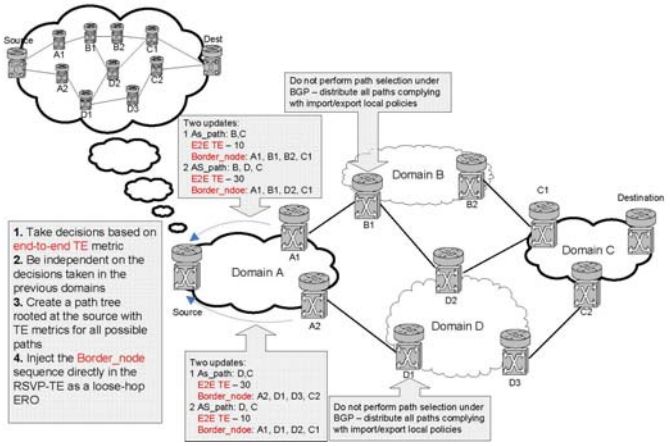


Figure 9: Enhanced BGP protocol operation

Three different strategies for TE support can be applied having implemented the stated extensions:

1. Using the end-to-end TE metric as a first decision criterion under a standard BGP path selection (***BGP TE case 1***);
2. Using the end-to-end TE metric as a first tie-breaking criterion under a standard BGP path selection (***BGP TE case 2***);
3. Using the end-to-end TE metric as the only decision criterion without implementing the traditional BGP path selection (***Enhanced BGP***).

For the last case due to scalability considerations it is important that suitable export policies are applied in the multi-domain network. Several such policies have been designed and evaluated.

The results from applying the stated strategies in a dynamic network with wavelength continuity constrained can be seen on Figure 10, which illustrates the blocking probability of LSP requests with varying load in a multi-domain mesh network. ***BGP TE case 3*** refers to the case where the Multi-Exit-Discriminator is used in the BGP path selection with the "Always compare" policy [Rek06]. Normalized load of 2 indicates the maximum input load per node (since all source nodes in the network have two outgoing links). From the figure it can be seen that the proposed ***Enhanced BGP*** scheme, which includes all modifications, performs the best. An interesting result is that ***BGP TE case 1*** performs the worst which is due to the greediness of the approach and the inherent BGP drawback of path dependency.

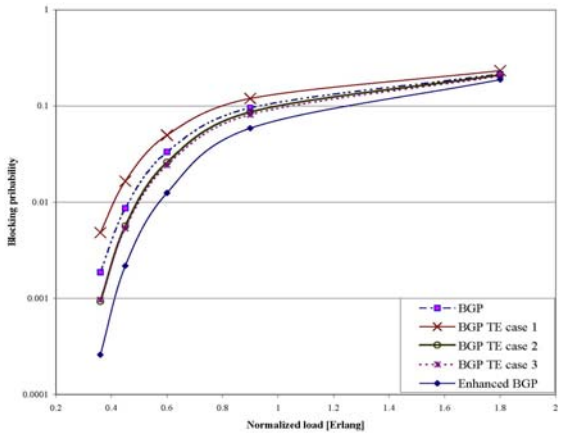


Figure 10: Blocking probability vs. Normalized Load

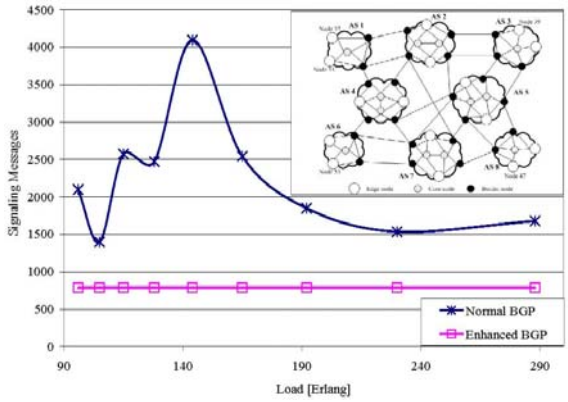


Figure 11: Signalling messages vs. Load

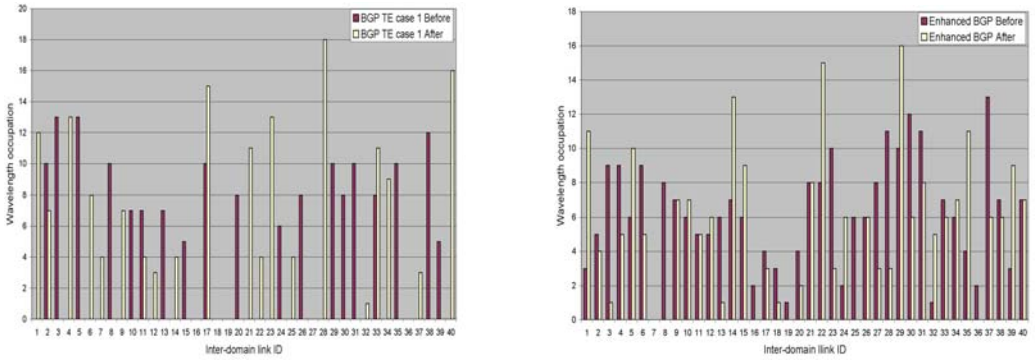


Figure 12: Wavelength occupations before and after re-convergence for BGP TE case 1 and Enhanced BGP schemes

The efficiency of the modified BGP protocol was also evaluated in terms of needed overhead for re-convergence of the protocol in order to accommodate any changes in the TE state of the disseminated paths. Figure 11 illustrates the result. The stable operation of the ***Enhanced BGP*** is evident, whereas the normal BGP operation has unpredictable overhead. Next, Figure 12 illustrates how the ***Enhanced BGP*** copes with one of the biggest BGP drawbacks[1] – the path dependency, which leads to poor utilization of the inter-domain links. Using ***BGP TE case 1*** strategy many of the inter-domain resources are used either only before or only after re-convergence, whereas the ***Enhanced BGP*** utilizes them continuously. The blocking probabilities for the illustrated simulation runs are 0.048 for ***BGP TE case 1*** and 0.014 for the ***Enhanced BGP*** respectively. The performance of the ***Enhanced BGP*** scheme was also compared to the performance of the ***BGP TE case 1*** (referred to as ***BGP-TE*** in Figure 13) in a network with wavelength conversion (WC) capabilities. Only the ***BGP TE case 1*** scheme is allowed to use wavelength conversion. The goal is to see if the cost of implementing the modified BGP protocol can be comparable to the cost of deploying wavelength converters, which are considered very expensive components. It can be clearly seen that for low to medium network loads the ***Enhanced BGP*** without WC outperforms the ***BGP-TE*** even with 10 WC per node. At medium to high network loads the performance of the ***Enhanced BGP*** is comparable to that of the ***BGP-TE*** with WC. If the cost of implementing the proposed BGP modifications is lower than the cost of WCs then the scheme is clearly the better choice for enhanced network performance.

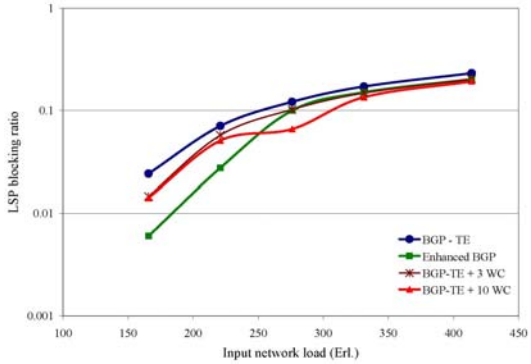


Figure 13: LSP blocking ratio vs. Input network load for Enhanced BGP without WC and BGP-TE with WC

[1] Drawback when BGP is applied in optical networks. For further explanation please refer to the publications and the references therein.

The novel extension of BGP proposed here can be used as a complementary routing protocol to a more advanced path computation framework such as the PCE architecture. PCE provides optimal path computation given an AS-path, whereas BGP provides the most suitable AS-path at the time of connection request.

From the presented results it can be seen that with some modifications the BGP protocol can be a viable routing protocol for the next generation multi-domain transport networks. It has the potential to support TE and QoS provisioning as well as to be used for survivability support.

2.2.5 BGP modifications for AS disjoint path dissemination (UC3M)

In multi-AS scenarios it is not possible to obtain the complete graph of the network without flooding the network with a lot of sensitive information. This is unacceptable because of the strong privacy protection policies between ASes and because the scalability of the network would be aggravated. The traditional approach for survivability support in multi-domain networks is to compute disjoint paths within the same set of ASes [Sta08] (see Figure 14 a)). In this JA, a novel solution for disjoint path selection, based on BGP has been studied (see Figure 14 b)). In particular, the developed mechanism obtains two AS-disjoint paths to each destination.

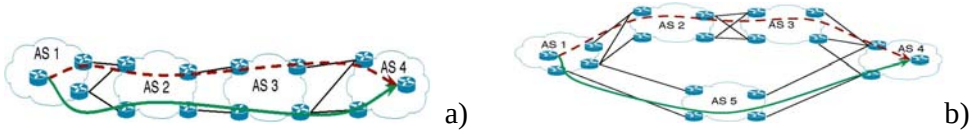


Figure 14: a) Disjoint paths within the same AS sequence, b) Disjoint paths on AS level

Our proposed mechanism is a concurrent modified BGP decision procedure, which obtains a disjoint AS PATH to the primary AS PATH (path chosen under the standard BGP operation), referred to as secondary AS PATH. This secondary path can be used for resilience purposes, load balancing or routing of subsequent LSP requests during BGP protocol re-convergence.

Performance enhancement under link failures

In our work we focus on three performance aspects. First we analyze the benefit of having two disjoint paths per destination with respect to the loss of traffic. Since the BGP protocol re-convergence takes significant time, this can result in high loss of traffic on existing connections and thus degraded network performance. Then, we focus on applying connection restoration for the affected LSPs. Utilizing the disjoint paths we can restore the affected connections at the time of failure without being affected by the BGP re-convergence delay or route oscillations. The third aspect we analyze is the performance of the dynamic multi-domain network in terms of blocking of future connection request. During the BGP re-convergence some nodes loose visibility of destinations or loops are created, which increases the LSP connection request blocking. In case of a link failure it is paramount to inform the proper network elements in order to minimize the impact of the failure through proper failure notification. In a multi-domain scenario there is still no consensus whether a failure should be signalled all the way to the head-end or if it shall be handled locally. In order to evaluate this we use the proposed extensions of the BGP protocol and we analyze the blocking ratio of connection requests after an inter-domain link failure using the following notification strategies:

- No notification: In this case we leave the BGP protocol to re-converge without notifying anybody of the failure. All LSP requests, which cannot be routed due to lack of visibility or routing loops in this period, are dropped.
- Local notification: In this case we notify only the border nodes of the domains that detect the failure. The border nodes will route the upcoming LSP requests using the backup paths obtained by the proposed BGP modification. If a routing loop occurs (in case a domain uses its upstream neighbour for the backup path) the requests are dropped at the upstream node4. No BGP re- convergence is done.
- Head-end notification: In this case the head-ends of the connections are notified that they must use their corresponding backup paths obtained using the proposed extensions. In this case no routing loops are possible and LSP blocking occurs only due to lack of resources. No BGP re-convergence is done.
- Mixed strategies: Here the LSP requests are routed on the backup paths during the BGP protocol re-convergence (using either the Head-end or the Local notification) and when the BGP protocol converges, the subsequent connection requests are routed on the new primary paths.

Simulation results

In order to evaluate the efficiency of our proposal we have implemented the BGP extensions in the event driven simulator tool OPNET and have tested it performance in two network topologies: a random one (see Figure 15) and the COST 266 Pan-European network. Two different sets of simulations are presented. The outcome of the modified BGP protocol illustrating its ability to provide AS disjoint paths is illustrated on Figure 15. Figure 16 shows the BGP re-convergence time and the amount of lost traffic during BGP re-convergence in case of 13 randomly selected inter-domain link failures in the COST 266 network. As it can be seen, in a Pan-European network BGP re-convergence takes tens of minutes which results in several terabits of lost traffic. Thus, applying survivability techniques for recovering affected connections and redirecting new requests on alternative paths can potentially save a lot of traffic and revenue for network providers.

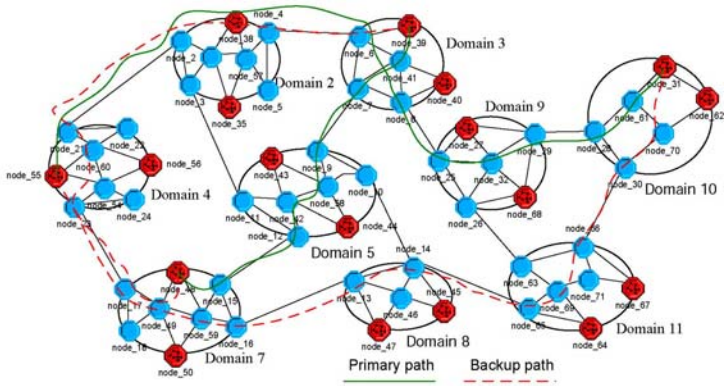


Figure 15: Disjoint paths from node_55 (Domain 4) to node_31 (Domain 10) and from node_48 (Domain 7) to node_39 (Domain 3)

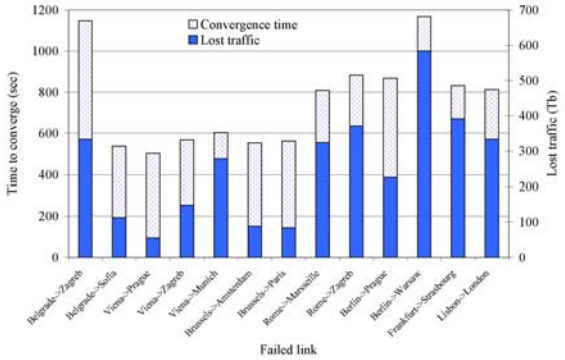


Figure 16: BGP re-convergence time and lost traffic for 13 inter-domain link failures

Figure 17 illustrates the amount of the affected traffic during two different link failure cases and the efficiency of the application of two restoration techniques, namely Local-to-Egress (L2E) and End-to-End (E2E). The most important observation to be made is that the efficiency of the applied restoration techniques is different for the different failure cases. For link Warsaw-Berlin, the E2E technique is more efficient than the L2E, whereas for the link Rome-Zagreb, the L2E technique restores more connections than the E2E technique. This indicates the need for differentiated failure handling in the network since the position of the failed link, together with the overall multi-domain connectivity influence the efficiency of the restoration techniques. Figure 18 illustrates the efficiency of the tested failure-notification mechanisms for reducing the connection blocking of subsequent LSP requests under BGP re-convergence. The failed link is the most-loaded link in the network – Berlin-Warsaw (see Figure 16). Considering the fact that a lot of traffic is affected, applying only local notification and redirecting the big amount of traffic locally, leads to higher blocking of requests since the local backup paths get saturated.

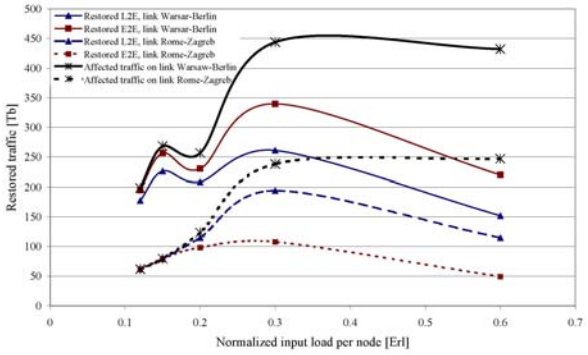


Figure 17: LSP restoration using AS-disjoint paths for two link failures

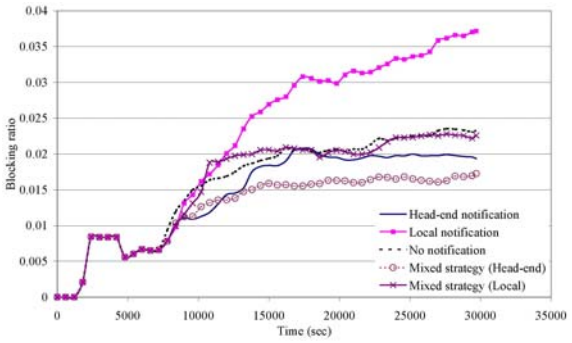


Figure 18: LSP blocking ratio under different failure notification strategies

Furthermore, the dependency of two of the notification strategies from the amount of affected traffic was investigated. For all cases, the network load is the same. The results can be seen on Figure 19. It can be seen that the more loaded the failed link is – the less efficient the L2E notification is. This is due to the fact that the local recovery paths get saturated much faster when there are many affected connections, whereas if E2E notification is applied, a better load balancing is achieved in the network. In the case where the amount of affected connections is smaller, the L2E performs better because it is easier to redirect them locally. In this case the efficiency of the E2E is degraded due to the long end-to-end backup paths for the requests, which increase the blocking probability.

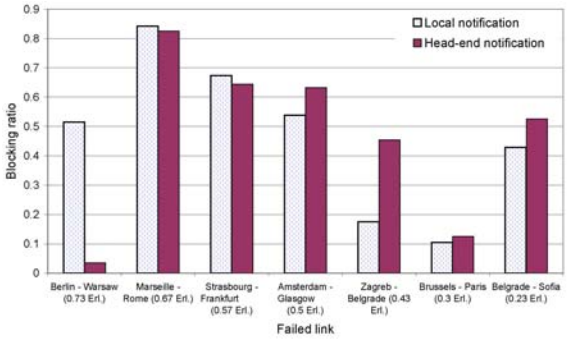


Figure 19: LSP blocking ratio vs. amount of traffic on the failed link for L2E and E2E notifications

Based on the results it is clear that applying AS-disjoint paths in a multi-domain network can bring significant performance enhancements under multi-domain link failure scenarios. Our proposed BGP modification provides such paths which can be used for protection and restoration as well as for redirecting connection requests under BGP protocol re-convergence. Such mechanism can be very beneficial in highly loaded networks with high traffic dynamism (i.e. networks where connection durations are much shorter (couple of hours to minutes) than the typical connections today (is in the order of days and months).

2.2.6 Genetic algorithms for efficient inter-domain traffic distribution (AGH)

Due to the development of Next Generation Networks, leading to a multiservice transport layer with a multi-domain environment, the importance of interconnection issues keeps growing. As the number of possible partners increases, operators face different routing options with regard to service quality and cost. Therefore, the need for developing algorithms

supporting the choice of optimal interconnection routes becomes crucial. Here, genetic algorithm to optimize the utilization of resources is presented and evaluated.

The considered solution is based on the general connection model shown in Figure 20. This connection model is presented from a point of view of an operator which wants to send the traffic to specific network directions/prefixes.

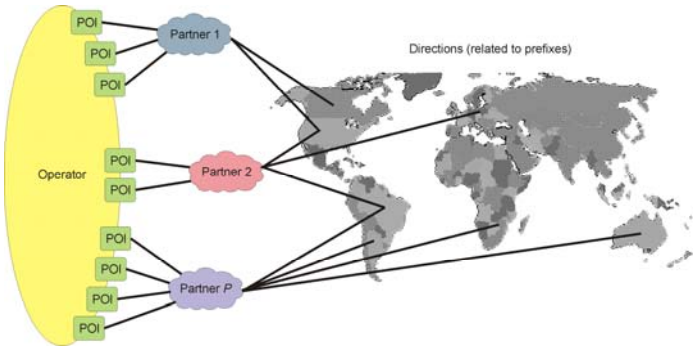


Figure 20: General connection model

It is assumed that the operator has agreements with a certain number of partners which offer the possibility of transiting or terminating the traffic for some directions. The operator usually has many points in its network through which the traffic is exchanged with its partners. These points are called POIs (Point of Interconnection). The parameters related to the POI are the tariffs and capacity.

The operation of the proposed genetic algorithm GEN can be described by the following steps:

- **Step 1**: *Generate initial population*

  The genetic algorithm GEN starts by randomly generating a population with W chromosomes. Each of the chromosomes represents the potential solution for the traffic distribution problem. The chromosome consists of G genes where one gene corresponds to one direction. The gen is composed of the following three parameters: partner ID, interface ID, and aggregated traffic volume which has to be sent to the specified direction.

- **Step 2:** *Generate child chromosomes by crossover process*

  After generating the required number of chromosomes the crossover process is performed. Two parents' chromosomes are selected and a new child chromosome is generated. As results of crossover, we obtained 9W child chromosomes.

- **Step 3:** *Perform mutation*

  Each generated child chromosome is subject to a mutation process with a probability.

- **Step 4:** *Decode each chromosome to obtain its fitness value*

  The fitness function takes into account the cost of sending traffic to the required directions as well as so-called penalty part. The penalty impacts the fitness function if the chromosome contains an infeasible solution. The value of penalty informs about the unfitness of the traffic distribution proposed by the chromosome and equals to the amount of violated capacity over all used partners and tariffs.

- **Step 5:** *Choose the best W chromosomes*

After calculating the fitness value for 10W (W parent and 9W child) chromosomes the best W chromosomes with the smallest fitness value are chosen.

- **Step 6:** *Repeat steps 2-5 until termination criterion is met*

  The termination criterion is once the number of generations N has been reached to avoid long convergence of the algorithm.

To evaluate the performance of the evolutionary methods a simulation scenario has been prepared. In the scenario it has been assumed that there are P=5 interconnection partners and D=40 directions. There is one preferred partner with I=4 interfaces and four non-preferred partners with I=3÷4 interfaces. For the preferred partners the unit costs are uniformly distributed from the 1÷3 interval while for the other partners from the 4÷10 interval. The traffic load is determined in relation to the sum of the partners' limit (100% traffic load corresponds to the sum of limits of all partners equal in the considered scenario to H=8000 units). The preferred partner is able to transit up to 50% of potentially maximum total traffic. The amount of traffic which can be sent through non-preferred partners is assumed to be equal to 1000 traffic units.

In Figure 21 the relation between the transit cost and the number of generations N is presented. The results are showed for 10%, 30%, 50% and 70% traffic load. The mutation probability is equal to p_m= 0.01. As we can observe, at the start of the GEN algorithm the cost is far from the optimal solution because the random population of feasible problem solutions is generated. While increasing the number of generations at the beginning the cost decreases significantly (e.g. from 10334 to 3946 for 70% traffic). After exceeding a certain value (dependent on the traffic load) further increase of the number of generations does not improve the solution.

Apart from the cost, a routing efficiency parameter is used to evaluate the performance of the proposed algorithms. The routing efficiency is defined as the ratio of the lower bound of the traffic cost to the total cost for the traffic. The lower bound cost is obtained by multiplying the cheapest price available to the direction by the total number of minutes to the direction summed over all directions. The introduced parameter informs about the cost loss due to a non-optimal traffic distribution. The routing efficiency for the GEN algorithm is shown in Figure 22. The results indicate that the relation between the routing efficiency and the number of generations is similar to the same relation for the cost. The worst routing efficiency (0.35 for N=1 increasing to 0.5 for N=100) has been obtained for 70% traffic. It comes from the scenario assumptions where there is one preferred partner able to transit traffic up to 50% of total traffic. For the 70% traffic load the traffic has to be distributed to other partners which offer more expensive tariffs.

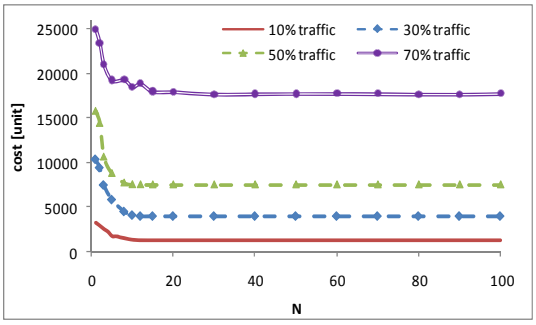


Figure 21: Cost for GEN algorithm

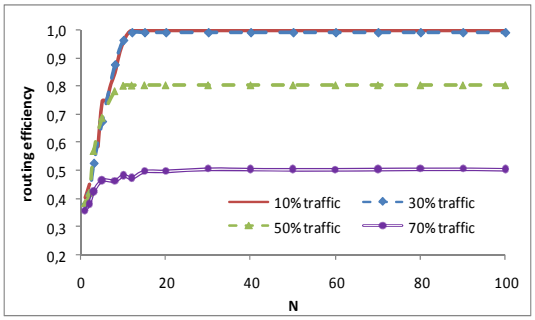


Figure 22: Routing efficiency for GEN algorithm

For the GEN algorithm the number of iterations decided about stopping the algorithm operation. According to the results, the relation between the computation time and the number of generations is almost linear and depends on the traffic load. The computation time for N=100 while analyzing 10% traffic load was less than 1 second while to generate the same number of generations for 90% traffic load required 10 seconds on average. We also checked that the time computation for more complicated scenario with P=10 transit partners offering the possibility of sending traffic to D=200 directions was very long and after 24 hours the algorithm has been stopped.

2.2.7 Managing inaccurate advertisements by penalty methods in multi-domain networks (BME)

The inaccuracy of incoming routing advertisements affects heavily the efficiency of path-selection. A penalty-based method is introduced to manage inaccurate routing advertisements in order to achieve a better efficiency in path-selection. We consider a multi-domain scenario where each domain disseminates routing advertisements pertaining to the aggregated (logical) topology of it. Routing advertisements consist of information for transfer attributes (e.g. bandwidth, latency).

Routing model

We distinguish two kinds of attributes on the basis of their metrics:

- MMA: Maximum Metric Attributes (indicating bottleneck quality, e.g. bandwidth),
- AMA: Additive Metric Attributes (indicating accumulative quality, e.g. latency).

The routing process which is assumed by our model consists of the following steps:

- The source domain collecting the incoming advertisements
- Applying inaccuracy-management → Penalty-adjusted advertisements
- Path-selection on the basis of penalty-adjusted advertisements*
- Sending connection request with the requested MMA values → Selected domains*
- Actual routing by the selected domains*
- The selected domains notify the source domain on the outcome of the transfer

Figure 23 illustrates the steps marked with *.

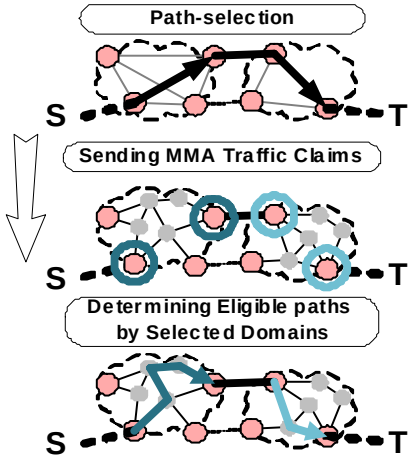


Figure 23: Routing model

Managing the inaccuracy of routing advertisements

The inaccuracy-management method consists of two separate steps in which the (current) incoming advertisements along with previous advertisements and outcome values are considered simultaneously:

- MMA inaccuracy-management,
- AMA inaccuracy-management.

In the first step, edges that are supposed to be not reliable are filtered out. We introduce three different approaches:

- Naïve approach: accepting all MMA advertisements as accurate ones,
- Stochastic approach: drawing (randomizing) the MMA values of the edge on the basis of the statistics of the outcomes at previous transfers,
- Deterministic approach: selecting the Most Reliable Path.

In the second step, penalty-adjusted edge costs are calculated on the basis of the deviations of previous incoming advertisements and the corresponding outcome values. Given the edge-filtered (first step) aggregated view of the network with penalty-adjusted cost values on its edges, the source domain can execute the shortest path selection to an arbitrary target domain.

Measuring the efficiency of path-selections

For measuring the efficiency of path selections an efficiency function is introduced for MMA and AMA separately (both are due to be minimized):

- Unsuccess-ratio (MMA-efficiency),
- Relative Cost-inaccuracy (AMA-efficiency).

Unsuccess-ratio is a quantity that considers the routing attempts that are either blocked for MMA-insufficiency or cancelled for MMA-mismanagement. On the other hand, Relative Cost-inaccuracy is a quantity that expresses the relative deviation of the cost-outcome and the cost of the minimum-cost path (between source and destination).

Results

Numerical results, showing the efficiency of the three inaccuracy management techniques with respect to the MMA efficiency of path selection are presented in Figure 24. The Naïve approach produces the highest unsuccess-ration as it does not provide any inaccuracy management. As the stochastic approach takes inaccuracies into account by applying a

randomization-based management, it results in a lower unusuccess-ratio. The deterministic approach chooses the Most Reliable Path for routing. Since this method relies on routing advertisements that do not vary to a large extent, its performance decreases heavily with the increase of traffic deviation. Figure 25 illustrates the efficiency of three penalty methods: (1) Naïve Approach (no compensation), (2) Exact compensation, and (3) (small) Penalty values use for compensation. As naïve approach does not impose any compensation on the incoming advertisements it produces the highest relative cost-inaccuracy. By applying exact compensation, managing inaccuracies with smaller deviations can be managed very effectively. In case of larger deviations, a more efficient way to manage inaccuracies is imposing penalties on inaccuracies. Penalties take into consideration the empiric deviations of incoming advertisements and the actual (outcome) values.

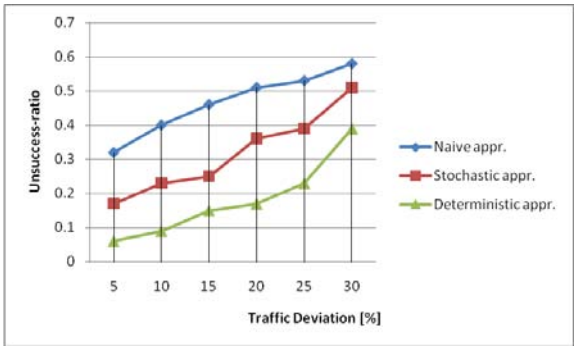


Figure 24: Unsuccess ratio vs. Traffic Deviation [%]

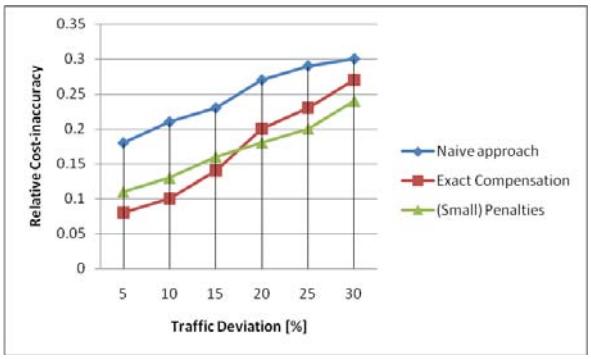


Figure 25: Relative Cost Inaccuracy vs. Traffic Deviation [%]

2.2.8 List of publications

This Joint Activity produced 10 publications, all of them being single partner papers.

- M. Kantor, K. Wajda, "*Inter-domain traffic optimization in resilient Next Generation Network environment*", in Proc. of NOC conference, July 2008, Krems, Austria.
- M. Kantor, P. Cholda, A. Jajszczyk, "*LCR Solution for Interdomain Traffic Distribution*", in Proc. of FITraMEn conference, December 2008, Porto, Portugal.
- M. Kantor, K. Wajda, A. Jajszczyk, "*Evolutionary algorithms for inter-domain traffic optimization*", in Proc. of NOC conference, June 2009, Valladolid, Spain.
- M. Kantor, P. Chołda, A. Jajszczyk, "*Simulated annealing-based algorithms for efficient inter-domain traffic distribution*", in Proc. of PTS conference, September 2009, Lodz, Poland.

- M. Kantor, P. Chołda, A. Jajszczyk, "*Optimized protection schemes for resilient inter-domain traffic distribution*", in Proc. of GLOBECOM conference, December 2009, Honolulu, USA.
- A. Manolova, S. Ruepp, L. Dittmann, "*TE-enhanced Path Selection for QoS Provisioning in Multi-Domain GMPLS Networks*", in Proc. of OFC/NFOEC, March 2009, San Diego, USA.
- A. Manolova, S. Ruepp, J. Buron, L. Dittmann, "*On the Efficiency of BGP-TE Extensions for GMPLS Multi-Domain Routing*", in Proc. of the 13th ONDM, February 2009, Braunschweig, Germany.
- A. Manolova, S. Ruepp, L. Dittmann, "*Performance Comparison of Multi-Domain Routing Schemes in GMPLS Networks with BGP*", in Proc. of 17th Photonics in Switching, September 2009, Pisa, Italy.
- A. Manolova, R. Romeral, S. Ruepp, "*Enhancing network performance under single link failure with AS-disjoint BGP extension*", In Proc. of 4th WSEAS International Conference on Circuits, Systems, Signal and Telecommunications (CISST '10).
- A. Manolova, S. Ruepp, "*Export Policies for Multi-Domain WDM Networks*", In Proc. of OFC/NFOEC'10, March 21-25, 2010, San Diego, California, USA.

2.3 QoT-Aware GMPLS Control Plane

The aim of this JA is the proposal of a QoT-aware GMPLS control plane to provision lightpaths while guaranteeing QoT in dynamic transparent optical networks.

In transparent optical networks, Quality of Transmission (QoT) will have to be ensured before client data is transmitted over a lightpath [Ram99]. In [Pin07], a probe-based IRWA scheme is proposed. This scheme, called Create-and-Wait (CW), uses impairment information to assess the status of the resources reserved for a given connection request both from QoT estimation before/during the reservation of the resources, and from QoT measurement after the reservation. That is, the lightpath set up consists of four phases: first, a lightpath is computed based on shortest path and wavelength availability; second, a chosen QoT performance parameter is estimated using a suitable model; third, the resources of the lightpath are reserved if the estimated QoT is above a predefined threshold; and fourth, the 'real' QoT is measured using probe traffic before client data transmission is allowed. The CW scheme is improved in [Sam08] by simplifying or eliminating the QoT estimation phase. The simplification is based on utilizing an equivalent-length model in the CW while the elimination is based on the utilization of a Probe-based Scheme (PS) only.

In this JA, three probe-based schemes accounting for multi-layer information (i.e., Multi-layer Probe Schemes) have been proposed during 2009. The schemes are based on multi-layer QoT measurements on probe traffic sent along a candidate lightpath to verify QoT before data transmission. Multi-layer QoT measurement brings forth three major advantages: first, the reduction of the monitoring points by combining QoT information from different layers, which allows to take into account QoT for any route in the network even if the ingress/egress nodes are not enabled with physical-layer monitoring capabilities; second, the possibility to combine information of QoT parameters at different layers to perform better QoT measurements, which depend on the characteristics of the probe traffic injected in the established lightpath prior to client data transmission; and third, the possibility to apply different Service Level Agreements (SLAs) to different classes of service (e.g., best effort and premium) which may require different levels of Packet Loss Rate (PLR), i.e. QoT. Moreover, different QoT estimation techniques based on signalling protocol extensions [Mar06] have

been used in two of the proposed schemes. Simulations show that low blocking is obtained for different service classes and fast set up time is also achieved when QoT estimation is performed during signalling.

2.3.1 Multi-layer Probe Schemes

The first scheme, called *Multi-layer Probe Scheme (MPS)*, enhances the PS scheme presented in [Sam08] by considering multi-layer QoT measurements. MPS does not require any additional routing or signalling protocol extensions. Upon lightpath set up from source *s* to destination *d*, *s* sends probe traffic along the lightpath. Then, *d* evaluates QoT by means of measurements (e.g., PLR) on probe traffic. If the measured QoT meets the required SLA threshold, the lightpath is activated and *s* sends data traffic along the lightpath. Otherwise, *s* computes an alternative path and performs another set up attempt. Probe measurements are performed using the Link Management Protocol (LMP) Link Connectivity Verification procedure based on the LMP Test message as in [Sam08].

The second scheme is referred to as *Signaling-based Multi-layer Probe Scheme (S-MPS)*. S-MPS requires signalling protocol extensions to gather information about QoT parameters as in [Cas07, Sal07] for performing QoT estimation before reserving resources and sending probe traffic. To maintain a lightweight control plane, the expected link noise contribution (accounted through optical signal to noise ratio - OSNR) is the only QoT parameter carried by RSVP-TE. The expected noise contribution of each link is recorded in adjacent nodes during WDM system installation or upgrades. Upon a lightpath request from *s* to *d*, *s* computes a path toward *d*. Then, *s* starts signalling by sending an RSVP-TE Path message toward *d* that gathers information related to available wavelengths and to the QoT parameter. In particular, each node traversed by the Path message appends in the Path message the expected OSNR value related to its downstream link. Because the inverse of the OSNR cumulates linearly along the path, upon Path message receipt, *d* calculates the expected OSNR of the path (i.e., $OSNR_P$). Moreover, a fixed margin M is used at *d* to account for other impairments (e.g., polarization mode dispersion - PMD, chromatic dispersion - CD). Thus, *d* subtracts M from $OSNR_P$ and the final computed OSNR (i.e., $OSNR_{PM}=OSNR_P - M$) is used to estimate QoT (e.g., PLR_{PM} as in [Pin08]). If the estimated QoT is unacceptable (e.g., $PLR_{PM} > PLR_{TH}$, where PLR_{TH} defines the SLA threshold), the lightpath is blocked and *d* sends an RSVP-TE PathErr message back to *s*. Otherwise, d sends an RSVP-TE Resv message, then QoT-based Probe measurement and possible successive set up attempts are performed as in MPS.

The third scheme is referred to as *Signaling-based Conditional Multi-layer Probe Scheme (SC-MPS)*. SC-MPS performs an additional, more stringent QoT estimation than the only one in S-MPS. The additional QoT estimation is based on a worst-case margin M' (M'>M). M' is used to compute the estimated $OSNR'_{PM}$ (i.e., $OSNR'_{PM}=OSNR_P - M'$) and the estimated QoT (e.g., $PLR'_{PM}>PLR_{PM}$). If $PLR_{PM}<PLR_{TH}<PLR'_{PM}$ then Probe measurement is performed as in S-MPS. If $PLR_{PM}>PLR_{TH}$, PathErr is sent to *s* as in S-MPS. If $PLR'_{PM}<PLR_{TH}$ data traffic is directly sent into the lightpath without the Probe measurement, thus accelerating the overall lightpath set up process. A one-bit flag inserted in the Resv message informs s on whether the probe-based measurement is required or not. Note that the worst-case margin guarantees acceptable QoT, leading to useless probing. It includes in particular database or measurement errors so when QoT estimation is acceptable, real QoT is also acceptable.

2.3.2 Simulation Results

The performance of the proposed schemes is evaluated by means of a custom C++ event-driven simulator. A Pan-European topology with 32 links, 17 nodes and a network diameter D=5 is considered. Two classes of service, best effort and premium, are considered. It is assumed that each lightpath carries data related to a single class, i.e., a lambda service. Best effort and premium services have the same probability to request a lightpath. Lightpath requests follow a Poisson process and are uniformly distributed among all node pairs. $PLR_{TH1}=10^{-9}$ and $PLR_{TH2}=10^{-2}$ are the PLR threshold considered for premium and best effort class respectively. At the physical layer, Amplifier Spontaneous Emission (ASE), PMD and CD are considered. When S-MPS and SC-MPS are performed, QoT estimation is based on the OSNR parameter included in the Path message and on the margin M considered to account for CD and PMD. Moreover, SC-MPS utilizes worst-case margin M'. In all the schemes, the QoT measurement is simulated considering a more accurate model that computes the OSNR penalty due to the effects of PMD and CD (PenPMD,CD=f(PMD,CD)) [Cas07]. Then, measured PLR is derived from the BER which in turn is derived from the computed OSNR as in [Pin08] because of the unavailability of real physical-layer and network-layer measurement elements in the simulations. To compute the setup delay, it is assumed that QoT measurement requires 10 s. When QoT is not met or not enough resources are available in the network, up to 2 additional set up attempts are triggered along alternative routes. The aim of the performance evaluation is to assess how the proposed schemes affect the blocking probability and the set up time of lightpaths carrying either type of service while assuring the required QoT.

The blocking probability is defined as the ratio between the number of blocked lightpaths and the number of lightpath requests. Lightpaths are blocked when n=3 set up attempts fail due to (i) signalling-based QoT estimation, (ii) probe-based measurement, and (iii) lack of wavelength resources. The set up time is the time elapsed from the lightpath request to the start of the data traffic transmission, i.e. it includes path computation, signalling, and probing (if used).

Figure 26 a) shows the blocking probability versus traffic load experienced by MPS, S-MPS and SC-MPS. Several values of M are considered for S-MPS. M=1.5dB and M'=3dB are considered for SC-MPS. Blocking probability increases with traffic load since longer routes have to be selected to satisfy wavelength continuity constraint but longer routes are more likely to be blocked due unacceptable PLR. Blocking probability increases with M since signalling based QoT estimation is more pessimistic and many paths are rejected during signalling. When M=1, 1.5 dB, S-MPS scheme presents the same blocking probability as MPS since these M values well approximate $Pen_{PMD,CD}$=f(PMD,CD) when a lightpath is critic (e.g., D hops). For M $\geq$ 1.7 dB, the blocking probability experienced by S-MPS is higher than the one experienced by MPS, since higher margins in the QoT estimation exclude the Probe measurement (and the lightpath set up) even on feasible lightpaths. SC-MPS performs as S-MPS with M=1.5 dB since M' does not influence blocking during signalling.

Figure 26 b) shows the average set up time versus traffic load experienced by MPS, S-MPS and SC-MPS considering premium class only and lightpaths traversing D or more hops. The set up time increases with traffic load since more set up attempts (with MPS and S-MPS) are required. S-MPS obtains better performance than MPS in terms of set up time for M=1, 1.5 dB since with S-MPS the blocking during signalling avoids to consume time for measurement on paths that would be rejected.

For M $\geq$ 1.7 dB, the set up time is lower than that in MPS because only lightpaths traversing few hops are established. Moreover, results show that, even by considering premium service

class only and lightpaths traversing D or more hops, significant improvement in the lightpath set up time is achieved with SC-MPS.

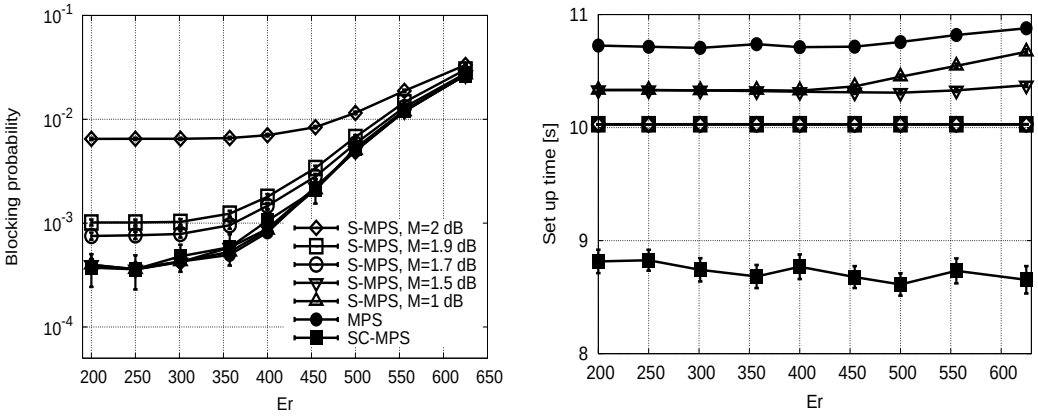


Figure 26: a)Blocking probability vs. traffic load. Best Effort and Premium traffic, b) Average lightpath set up time vs traffic load for Premium traffic lightpaths traversing D or more hops

In Figure 27 the average lightpath set up time is reported for all the service classes and for MPS, S-MPS (M=1.5dB), SC-MPS (M=1.5dB, M'=3dB). Results show that S-MPS obtains better performance than MPS especially for lightpaths related to premium class which traverse D hops or more. SC-MPS obtains the best performance in terms of set up time for lightpaths belonging to both classes and traversing any number of hops, since SC-MPS avoids probing measurements for some paths thanks to the additional QoT estimation.

	All Classes		Best Effort Class		Premium Class	
	All paths	D+ hops	All paths	D+ hops	All paths	D+ hops
MPS	10.06	10.38	10.02	10.03	10.11	10.72
S-MPS	10.03	10.18	10.01	10.02	10.06	10.33
SC-MPS	0.77	4.55	0.05	0.30	1.50	8.81

Figure 27: Average lightpath set up time

2.3.3 Conclusion

Three multi-layer lightpath set up schemes based on probe traffic measurements are proposed for guaranteeing the required QoT in control plane enabled transparent networks (e.g., GMPLS). Multi-layer performance information allows to apply different SLAs to different service classes. Results first show the trade-off due to stringent and loose QoT estimation between blocking probability performance and overall lightpath set up time. Then, results show that the proposed SC-MPS scheme allows the effective lightpath set up for different classes of traffic requiring different levels of QoT.

By exploiting signalling-based QoT estimation, and probe-based measurements only in case of critical QoT estimations, SC-MPS achieves good performance both in terms of lightpath blocking probability and lightpath set up time.

2.3.4 List of publications

- N. Sambo, C. Pinart, E. Le Rouzic, F. Cugini, L. Valcarenghi, P. Castoldi, "*Signaling and Multi-layer Probe-based Schemes for guaranteeing QoT in GMPLS Transparent Networks*", in Proc. of OFC/NFOEC 2009, San Diego - CA (USA), March 2009.

- N. Sambo, N. Andriolli, A. Giorgetti, P. Castoldi, G. Bottari, "*Multiple Path based Regenerator Placement Algorithm in Translucent Optical Networks*", in Proc. of 11th International Conference on Transparent Optical Networks, ICTON 2009, Island of São Miguel, Azores, Portugal, June 2009.
- N. Sambo, N. Andriolli, A. Giorgetti, L. Valcarenghi, I. Cerutti, P. Castoldi, F. Cugini, "*GMPLS-Controlled Dynamic Translucent Optical Networks*", IEEE Network, Vol. 23, No. 3, pp. 34-40, May 2009.
- F. Ramos, A. Giorgetti, F. Cugini, P. Castoldi, J. Crowcroft, I. White, "*Power Excursion Aware Routing in GMPLS-based WSONs*", in Proc. of OFC/NFOEC 2009, San Diego - CA (USA), March 2009.
- F. Cugini, N. Sambo, N. Andriolli, L. Valcarenghi, P. Castoldi, E. Le Rouzic, J. Poirrier, "*Enhancing GMPLS Signaling Protocol for Encompassing Quality of Transmission (QoT)*", Journal of lightwave Technology, October 2008.

2.4 MPLS-ASON/GMPLS Interconnection

2.4.1 Objective

To establish a test-bed for future Inter-Domain research issues. The test-bed should emulates a multi-domain network, two IP/MPLS islands interconnected through an ASON/GMPLS network. This test-bed allows to do research in multi-domain multi-layer protection mechanisms in the future Optical Internet.

2.4.2 Description

The introduction of intelligence in Automatically Switched Optical Networks (ASON) [ITU8080] using a Generalized Multiprotocol Label Switching (GMPLS) control plane [Man04] allows to setup, configure, and release optical connections, in a fast and dynamic way.

In a close future the well-known IP/MPLS network must coexist with the new GMPLS-controlled optical networks and the interaction between them will be inevitable.

One of the future network scenarios will be one or more IP/MPLS networks interconnected through an ASON/GMPLS network [Vel08]. In this scenario interaction between the MPLS and ASON/GMPLS will be needed, especially in signalling protocol, to allow the set-up of end-to-end LSPs with the desired QoS. One of the most common ways to provide QoS in circuit-based networks is by providing LSP protection.

For these reasons, the IETF Common Control and Measurement Plane Working Group (CCAMP) has the standardization of the MPLS-GMPLS interaction as one of their scopes.

To advance in future Optical Core Network research is essential to get the appropriate simulation tools, since research in real optical networks requires high investments. There are two main ways to do this, to simulate the network and to emulate it.

2.4.3 Current results

The implemented test-bed is based on an emulated IP/MPLS test-bed which uses two different tools, *dynagen* and *dynamips*, and the CARISMA network test-bed, an emulated ASON/GMPLS network. Both test-beds have been interconnected via Internet GRE Tunnels.

Using this test-bed research community could advance in new signalling procedures to improve future optical network survivability in a multi-domain and multi-layer scenario.

The CARISMA Network Test-Bed

The CARISMA network test-bed has been implemented to be used as a multi-domain field-trial for the integration and evaluation of the ASON/GMPLS technologies.

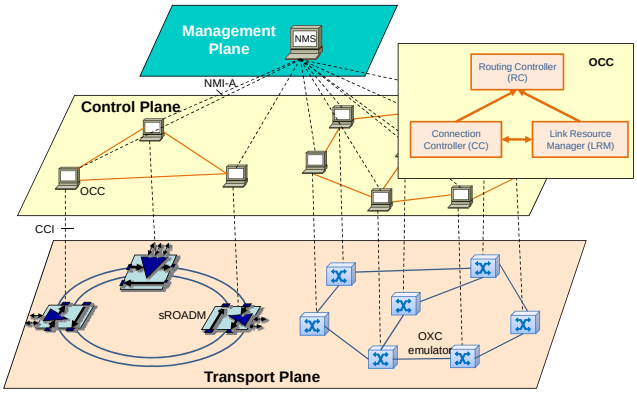


Figure 28: The CARISMA network test-bed

Figure 28 presents the architecture of the CARISMA network test-bed. It contains the following three functional planes: the transport plane, responsible for traffic transport and switching; the control plane, responsible for connection and resource management. It can be either associated with (*in-fiber*) or separated from (*out-of-fiber*) the managed transport network; and the management plane, responsible for management of the whole system (including transport and control planes). It triggers commands to the control plane to set-up and tear-down soft-permanent connections. At the management plane of the ASON/GMPLS CARISMA network test-bed, the Network Management System (NMS) was implemented as a web-based application, easing network management through the Internet.

At the transport plane, two alternative nodes can be used: the physical node and the emulated node.

Dynamips/Dynagen test-bed

Dynamips is a Cisco Emulator created by Christophe Filliot [Dyna]. At first, the project consisted on emulating a Cisco 7200 router but, as the project progressed, it other families such as the 3600, 3700 and 2600 series. It provides support not only for Ethernet interfaces, but also for ATM (Asynchronous Transfer Mode), Serial and PoS (Packet over SONET) interfaces. Dynamips emulates the underlying hardware of the router in order to run a Cisco IOS image. The user can specify the interface modules the emulated router has and the slot in which these modules are placed. The configuration of the router is passed through the command line. Dynamips also has a server mode in which another process can connect and create new router instances. This server mode is used by another application, Dynagen [Dynab], to automatise the creation of emulated network scenarios. Dynagen is a front-end for Dynamips. Dynagen is able to read network description files in a simple syntax. These files describe the routers present in the network, the interfaces they have and how they are connected to other routers by either virtual or physical interfaces. It also provides a console prompt to give basic commands to dynamips, such as starting and stopping a router or even

disconnecting an interface from the router. The emulated IP/MPLS test-bed for this activity can be seen in Figure 29.

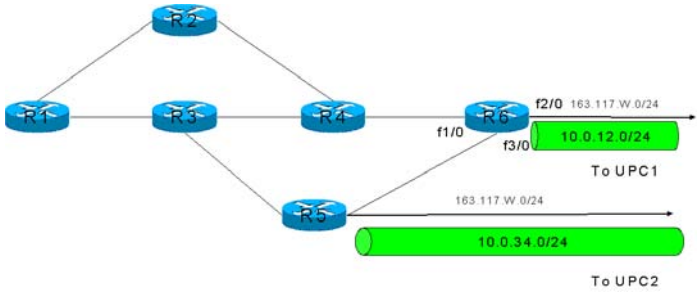


Figure 29: The IP/MPLS emulated network

It consists of six virtual Cisco 7200 routers emulated with the Dynamips Cisco emulator. All routers are connected with virtual Fast Ethernet interfaces except for the link R3-R4, which is a PoS link. The links between the nodes were chosen to build a trap topology. Routers R5 and R6 are directly connected to the Internet by mapping two of their interfaces to one of the wired network cards of the server where the emulation is held. Each one of these two interfaces is given a public IP address.

MPLS-ASON/GMPLS Network interoperation

As we can see in Figure 30, the two routers accessing the Internet have established a GRE tunnel with the network at the UPC. GRE tunnels allow us to encapsulate MPLS frames through the Internet, giving continuity to the MPLS domain we have defined. This way, we can connect several virtual networks hosted in different servers of the Internet. This allows us to simulate large networks in a distributed way. Should several links fail and our network be splitted into two parts, we could connect to the other half via the UPC network. Over this test-bed we can test interoperation between different G/MPLS implementations and different multi-domain G/MPLS protection schemes and multi- layer protection schemes as well. Firstly end-to-end multi-domain LSPs (non-protected) have been established already in the test-bed. After that a whole word of research opportunities is opened.

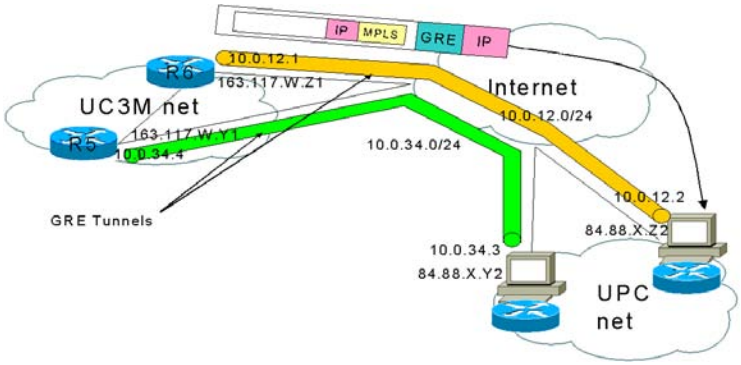


Figure 30: MPLS ASON/GMPLS interconnection test-bed via Internet

2.4.4 Conclusions

In this JA we have presented two emulated network, one of them an ASON/GMPLS network based on the CARISMA network test-bed and the other one an IP/MPLS Network, based in

Dynagen/Dynamips emulated environment. The resulted test-bed is a distributed emulated multi-layer multi-domain network. Over this network different scenarios can be studied in different aspects. Our intention is use it to research in future protection schemes for the future Optical Internet. The test-bed allows studying different inter-domain survivability schemes, doing possible to compare along them and select the best one in each scenario and application. The shown technology test-beds can be larger and simulate a thousand nodes, giving the opportunity to probe the different signalling procedures in a real- like Network. Our main goal is to use the test-bed to probe the proposed mechanism in [Vel08] and resolve the signalling interoperation problems between a MPLS Network and an Optical GMPLS Network.

2.4.5 List of publications

- L. Velasco, R. Romeral, F. Agraz, S. Spadaro, J. Comellas, G. Junyent, D. Larrabeiti "*On the design of MPLS-ASON/GMPLS Interconnection Mechanisms*", in Proc. of VII Workshop in G/MPLS Networks, Vilanova i la Geltrú, Cataluña, Spain, 11-12 March 2008.
- G. Rodríguez de los Santos, R. Romeral, D. Larrabeiti, L. Velasco, F. Agraz, S. Spadaro, "*Emulated MPLS-ASON/GMPLS Inter-connection test-bed*", in Proc. of VIII Workshop in G/MPLS Networks, June 29th 2009, Girona, Spain.

2.5 Scalability issues in G/MPLS-based VPLS network design

2.5.1 Objective

Until recently, most Virtual Private Network (VPN) services could do just with point to point communications and IP multicast sessions were served by VPN service providers (SP) through packet duplication at the source ingress node over a number of LSPs. However, unlike in global Internet communications where IP multicast is disabled by many ISPs (partly due to the lack of cost models for IP multicast traffic exchange) forcing some sort of application layer multicasting (ALM) –e.g. peer-casting–, a bunch of important bandwidth-intensive multicast services prevail in corporate networks. Examples are corporate TV channels, disk replication tools and multi-backup systems. This makes it necessary for both customers and service providers to properly address the transport of multicast traffic when the branches of the organisation are connected by a layer-2 Virtual Private LAN Service (VPLS), where the SP is expected to emulate a shared broadcast network [Agg08] with a given capacity. Indeed, in this case, satisfying a target committed rate in multipoint mode may require an oversubscription of access Provider Edge (PE) nodes that may raise the costs to unacceptable figures. For example, if a bank is using an application based on reliable multicast to quickly copy 20GB of data every week to all its 100,000 branch offices at 100 Mb/s, the service expected by the bank is 100 Mb/s broadcast capability of the VPLS. With edge point-to-point star delivery, the emulation of such service would require a 1 Terabit/s PE backbone link just for this customer. Therefore there is an important scalability issue that could be dealt with multipoint connections.

2.5.2 Description

Regarding multicast, the Internet Engineering Task Force (IETF) group has made notable efforts to provide solutions for multicast MPLS VPN communications

[Agg08][Ros06][Ros07]. Like in the case of Internet, we believe that multipoint many-to-many communications is not actually demanded and the users can simply deal with a few hubs (usually bank central server node and secondary backup nodes). Even with shared tree multicasting the implementation cost scalability issue remains in place and the usage of MPLS multipoint LSP sand even Optical multipoint connections is a must for high-speed services. A clear example is triple-play providers [San06] that deliver TV over IP multicast to their ADSL residential clients, in which usually the last hop is delivered over IP unicast from a multimedia relay at the SP point of presence (PoP). This could be implemented with IP/MPLS over point-to-multipoint (P2MP) label switched paths (LSP) sent from a content delivery root to the relays [Mar07b]. However, in the case of VPLS there is not just a single source but a large number of star or double-star VPLSs demanding an individual tree with a different root. This breaks the intended scalability of MPLS-based VPN implementation: LSPs are not shared by different VPNs. In a MPLS VPN network, unicast packets are forwarded without any state information about the VPNs kept in core routers. That information is only known by the provider edge (PE) routers, which connect sites directly to the VPN. Client data travels from a PE to another PE through core nodes in tunnels, usually Label Switched Paths (LSPs). This technique is optimal, because the associated state information in core routers depends only on the number of PEs, instead of the number of active VPNs. Optimal multicast VPN routing requires at least one distribution tree per source per multicast demand. Like this, the state information in every provider router reaches a triple dimension: multicast source, multicast group and VPN. In this way networks scale poorly. Potentially, this would require unlimited amount of state information at the core routers, because the SP has no control neither over multicast groups within the VPNs, nor over the number of transmitters at each group, nor over the distribution of the receivers. The model proposed in this work is the aggregation of a set of multicast groups on a single shared distribution tree. In this manner, we establish a tunable trade-off between the state and bandwidth optimization.

The work in this JA deals with the concept of aggregation of multicast groups or VPLSs. In this procedure, multiple multicast groups are forced to share a single multicast distribution tree (MDT) a.k.a. aggregation tree. In this way, the number of trees configured within the network is significantly reduced, and consequently the forwarding state information also decreases –because nodes do not need to keep the state information for every multicast tree anymore, but only for every aggregation tree. This idea was proposed for multicast routing at the IP layer in [Fei01]. This improvement is achieved at the expense of bandwidth waste, because aggregation trees may deliver packets to nodes with no associated members. Despite this, the aggregation tree strategy can be much more efficient than unicast-VPNs and one-tree-per-VPN strategies. A methodology to analyze the behaviour and benefits of using shared trees in MPLS-based VPN networks was proposed in [Mar07b]. In the study, an important state vs. bandwidth trade-off analysis was held. The authors introduced a simple model for the estimation of the state-bandwidth gain caused by multicast VPN aggregation in an SP network. Extending the referred methodology, we construct a new simulation model to be applied to real topologies. Major extensions have been added to the framework in order to make it more accurate and closer to real scenarios, like a new model to determine the PoPs, a new mechanism to build aggregation trees and to select Rendezvous Points (RP) –roots of the aggregation trees. We also define a model to determine the distribution of the number of Provider Edge (PE) routers and VPNs. One of the key aspects of our proposal is the VPNs distribution among the established distribution trees, and techniques to solve this issue in a non-random manner have developed. Finally, the aggregation can happen at two layers: the MPLS and the optical GMPLS layer.

How this partition and allocation of VPNs to trees should be made is an open research issue, given the diversity of topologies, traffic and sites of different VPNs. On the other hand, high rate flows may justify the set up of group membership-aware multicast trees to eliminate traffic in nodes not leading to group receivers.

In this joint activity we have studied the impact of intelligent aggregation of VPNs in bandwidth and forwarding state efficiency. This study is not limited to reference topologies but also tries to assess the results in real backbone topologies.

The following figures show several examples of the type of simulations being carried out. Figure 31 illustrates the gain due to intelligent aggregation for different aggregation degrees (0% means no tree aggregation, 100% means a single shared tree) and also the effect of reconfiguration (dynamic re-organisation of trees), for a density of 25% (the probability of a PE belonging to a given VPLS is 0.25).

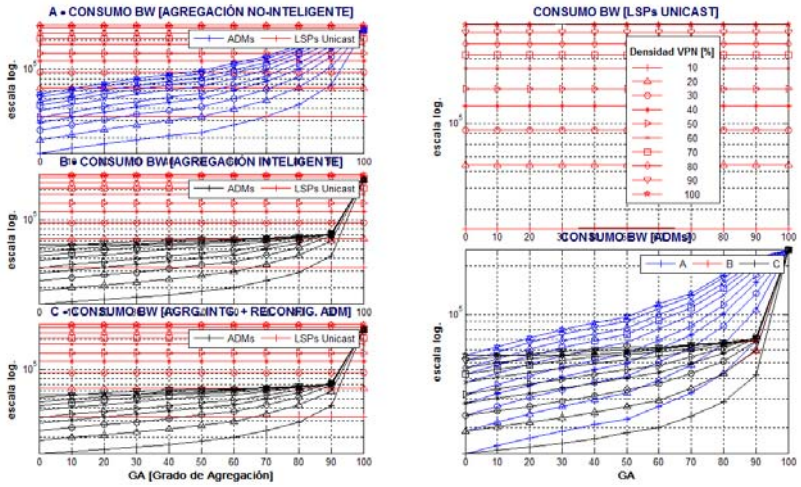


Figure 31: Bandwidth consumption in Tiscali network for 25% of VPLS member density

Figure 32 shows the same parameters for a 75% VPLS density.

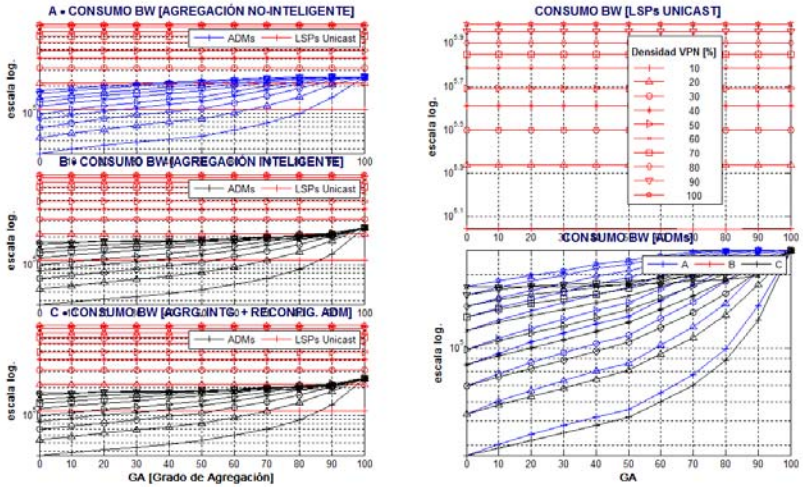


Figure 32: Bandwidth consumption in Tiscali network for 75% of VPLS member density

The results show that intelligent aggregation is very effective for high aggregation degrees and that dynamic reconfiguration adds very little value in this setting.

2.5.3 List of publication

- The JA has not yet published their results in a public conference. The simulation results are available as a technical report until the model is completed with VPLSs of different size distributions. An article to be submitted to Elsevier Communication in 2010 is under preparation.

2.6 GMPLS-based RWA algorithms for optical protection/restoration

This Joint Activity is focused on GMPLS-based protection and restoration recovery schemes for Wavelength-Routed Networks (WRN), taking into account the two major challenges of wavelength-routed networks, namely, the wavelength continuity constraint and the degradation of the optical signal quality. The former arises in all-optical networks without or with limited wavelength converters installed at the node. Indeed, despite being quite technologically mature, wavelength converters are scarce and expensive. The latter is due to the accumulation of physical impairments whilst the signal travels from source towards destination node.

In this context, neither standard GMPLS OSPF-TE nor RSVP-TE protocols convey the required routing and signalling information to deal with efficient distributed RWA algorithms and reservation protocols for protection/restoration. This joint activity aims at:

- Proposing GMPLS-enabled protection and restoration algorithms along with the required extensions to the current GMPLS protocols addressing the identified drawbacks.
- Validating and evaluating the performance of the proposed GMPLS-based protection and restoration mechanisms.

This JA is organized in three specific joint works, namely:

- Shared Path Protection (SPP) in GMPLS networks with limited wavelength conversion (CTTC, SSSUP and DTU).
- Label Preference Schemes for Lightpath Restoration in Distributed GMPLS Networks (SSSUP, DTU and CTTC).
- Network Performance Improvement in Survivable WDM Networks Considering Physical Layer Constraints (AIT, KTH).

2.6.1 Shared Path Protection (SPP) in GMPLS networks with limited wavelength conversion

General objectives and summary of results from Y1

The main objective of this joint sub-activity is to investigate novel GMPLS-enabled Shared-path protection (SPP) path computation algorithms and reservation protocols along with the required extensions to the current GMPLS RSVP-TE protocol for wavelength-routed optical networks with limited wavelength conversion, that is, with scarce wavelength converters (WC). The Shared Path Protection (SPP) scheme is widely accepted as the most capacity-efficient recovery scheme achieving an acceptable recovery time. Such benefits are

accomplished through the so-called backup sharing, i.e., the sharing of resources among the existing protection (or backup) lightpaths (Label Switched Path - LSP in GMPLS context). Specifically, resources along backup LSPs are pre-reserved, but not cross-connected. Therefore, to ensure 100% survivability of the LSPs affected by a single-link failure, shared-reserved resources can be reserved by one (or more) backup LSPs provided that the corresponding working LSPs do not share any link (i.e., no sharing violation). Standard GMPLS OSPF-TE disseminates the Shared risk link group (SRLG) information as a TE link attribute in order to ensure link-disjointness between working and backup paths. SRLG identifies a group of links that share a common risk of failure. Therefore, two paths are link-disjoint as long as they are SRLG-diverse. This work assumes a one-to-one relationship between SRLGs and Links, and thus both can be used indistinctly.

Upon reception of an LSP request, the source node must execute a constraint shortest path first (CSPF) algorithm to find two feasible end-to-end working (wl) and backup (bl) paths, considering as input the topology and network resource state collected in the traffic engineering database (TED) repository. The routing protocol (e.g., OSPF-TE) is responsible for flooding any change occurring in the network state, which permits to update the local TEDs. The requirements that a pair of working and backup LSPs must fulfil are:

- R1) *wl* and *bl* paths for a common connection are SRLG-diverse;
- R2) two (or more) *bl* paths of different connections can share a wavelength channel on a given link as long as their respective *wl* paths are SRLG-diverse (no sharing violation);
- R3) Both working and backup LSPs are subject to the wavelength continuity constraint in each segment between consecutive WCs.

Given these constraints, a CSPF algorithm using standard GMPLS dissemination (Unreserved Bandwidth or UBw and Maximum Bandwidth or MaxBw bandwidth attributes) satisfies R1 only. Thus R1 is fulfilled by computing paths with non-zero link UBw. R2 can be satisfied during the signalling mechanism, since detailed information on shared resources is not flooded by GMPLS routing protocols. R3 cannot be managed by the CSPF, since there is no dissemination of detailed (per-wavelength) information.

Once *wl* and *bl* paths are successfully computed, then they are passed to the RSVP-TE signalling process by means of an Explicit Route Object (ERO) to proceed with the working LSP establishment and the backup LSP reservation. With regard to the wavelength continuity constraint in each segment between consecutive WCs (R3) and wavelength sharing violation (R2), GMPLS relies on the RSVP-TE signalling protocol. Specifically, it relies on two objects of RSVP-TE messages:

- For wavelength sharing violation (R2), the Path message used for establishing backup LSP can include the so termed Primary Path Route Object (PPRO) [IETF RFC 4872]. The aim of the PPRO is to inform each backup LSP hop about the nodes and TE links (SRLGs) being traversed by its associated working LSP. A sharing violation takes place when one or more SRLGs of the working LSP are the same of the SRLGs protected by a shared resource (either a WC or a wavelength on a link)
- For wavelength continuity constraint (R3), the Label Set Object (LS) [IETF RFC 3473] can be included in the Path message at the source node for both the working and protection LSPs. The Label Set allows an upstream node to restrict the set of wavelengths (labels) that a downstream node can select.

Apart from the ERO, PPRO and LS, we propose two different vector objects with the following weights that must be included in the reservation of protection LSP:

1. *Shared wavelength vector* (SW) (also referred to as shared label set in [Mun08]) whose elements indicate the number of links on which the corresponding wavelengths are in shared-reserved status.
2. *Suggested WC vector* (SV) (also referred to as wavelength set metric in [Ji08] or simply suggested vector in [And06]) whose elements indicate the minimum number of WCs that are required in order to use the corresponding wavelengths from the source.

At the destination and each intermediate node, the following strategies for selecting the resources for protection LSPs can be used.

1. *WC Selection*: For the working LSP, the first available WC is selected. For the protection LSP, among the idle and sharable WCs that are available locally at the node and does not incur in a sharing violation, the first WC is selected.
2. *Wavelength Selection*: These strategies apply at both the destination node and any intermediate node that should perform wavelength conversion. Different strategies to be applied to new vector objects have also been proposed for updating object weights and for performing wavelength assignment. Wavelength assignment strategies apply at both the destination node and any intermediate node that should perform wavelength conversion. The main selection criteria that can be devised are:
    - Random (R): a wavelength in the label set is randomly selected;
    - First fit (FF): the first wavelength in the label set is selected;
    - Last fit (LF): the last wavelength in the label set is selected;
    - Maximum-wavelength-sharing (WS): the wavelength in the label set with the highest weight in SW vector is selected. Ties are broken by using LF strategy applied to the restricted label set.
    - Minimum-conversion (C): the wavelength in the label set with the lowest SV weight is selected; Ties are broken by using FF, LF R or WS strategies applied to the restricted label set.

Advances in Y2

During this year, the JA partners worked jointly to finalize the framework and the strategies described in the above section. The implementation of an event-driven simulator supporting the proposed strategies has also been carried out at DTU during this second year. Preliminary results on the performance of the proposed selection strategies have been evaluated on the NSFNET topology, formed by 14 nodes and 22 bidirectional links with 32 wavelengths per link (WL). The number of available WCs per node is fixed to 10. The lightpath-arrival process is Poisson, and the holding time follows a negative exponential distribution, with source-destination uniformly distributed among all (distinct) node pairs. The mean holding time (HT) is set to 30 minutes, and the average inter-arrival time (AT) will be set up according to the normalized input load per node, ranging from 0,2 to 1 Erlang. It is calculated as the HT/AT/WL. The results are obtained by running the OPNET Modeler simulator.

We compare the performance achieved by each of the presented wavelength assignment strategies (R, FF, LF, WS and C) when using the proposed vectors (i.e., SW and SV) for SPP scheme. Two performance key indicators are considered, namely, the blocking probability, and the resource overbuild (RO). In SSP, the blocking probability is the probability that either working LSP (WP) or protection LSP (BP) is blocked. The resource overbuild is a figure of merit specific for recovery schemes, and is defined as the amount of channels consumed by protection LSPs over the amount of channels utilized by working LSPs.

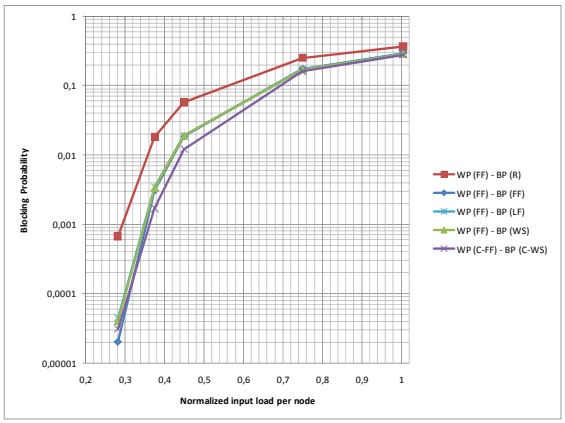


Figure 33: Blocking probability versus load

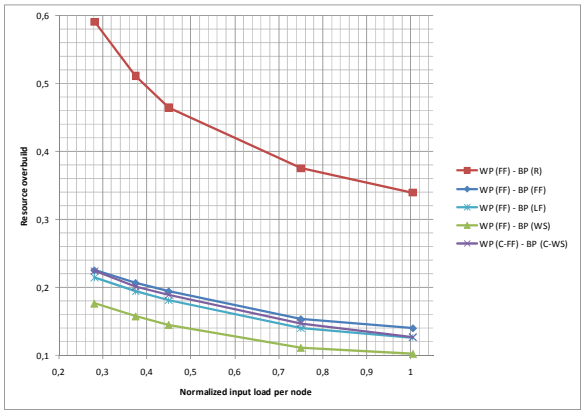


Figure 34: Resource overbuild versus load

The wavelength assignment strategies for the working LSPs is fixed to FF, except when the SV vector is used. In this case, the working LSPs uses C-FF (i.e., C wavelength conversion in which ties are broken using FF). For protection LSPs, all the presented wavelength assignment strategies are tested, that is, R, FF, LF, WS and C-WS. It is worth noting that WS strategy makes use of the proposed SW vector, while C-WS uses both SV and SW vectors. For R, FF and LF, only the standard Label Set is needed.

From Figure 33, it can be observed that the best wavelength assignment strategy in terms of blocking probability is WP (C-FF) – BP (C-WS). The efficient use of the proposed SV vector allows to optimally use the limited number of WCs. The optimal WC usage permits to establish a larger number of LSPs as the blocking due to wavelength continuity constraint is reduced.

Figure 34 shows the resource overbuild versus load. It is desirable to have the resource overbuild as low as possible, i.e., a better sharing of backup resources. From the figure we can observe that WP (FF) – BP (WS) outperforms all the other strategies. The use of the proposed SW vector allows the selection of the wavelength that is in shared state on the largest number of links along the path and, thus, it increases the possibility to efficiently share the backup resources. It is worth noting that R is the worst strategy, with the lowest performance in terms of both blocking probability and restoration overbuild.

2.6.2 Label Preference Schemes for Lightpath Restoration in Distributed GMPLS Networks

To overcome failure consequences in wavelength switched optical networks (WSONs), a prompt restoration of failed lightpaths can be activated to guarantee an uninterrupted service to users.

In WSONs with a distributed control plane based on Generalized Multi-Protocol Label Switching (GMPLS), the ability to promptly recover the disrupted lightpaths is hindered by the multiple restoration attempts that, upon link failure, contend for the residual network resources. Thus, to restore the disrupted lightpaths, the signalling message exchange is triggered, but it can be blocked in the backward direction (i.e., backward blocking) due to resource contentions or in the forward direction (i.e., forward blocking) due to lack of resources. In restoration, backward blocking typically exceeds forward blocking [Sam08b].

In the past, SSSUP and others demonstrated that the utilization of wavelength-converters has been demonstrated to be effective in reducing forward and backward blocking in a lightpath provisioning scenario [And06]. In this joint work, wavelength converters are exploited with the specific aim of reducing wavelength contentions.

The aim of the joint work is the assessment of the benefits of wavelength conversion during the restoration of lightpaths disrupted by a single-link failure in a GMPLS-based WSON. In addition, the use of an intelligent wavelength selection strategy is evaluated in the presence of wavelength converters.

RSVP-TE based Lightpath Restoration

The Resource ReSerVation Protocol with Traffic Engineering extensions (RSVP-TE) is used for both lightpath provisioning and restoration. An RSVP-TE *Path* message is sent towards *d* including the LabelSet object. The LabelSet object is a list of labels (i.e., wavelengths) and it is managed by intermediate nodes so that, at *d*, it contains the wavelengths that are available on the whole restoration route.

Upon reception of the *Path* message, *d* selects the wavelength to be reserved among those contained in the received LabelSet. Two wavelength selection strategies are considered:

- **Random (RND) strategy**: a wavelength is randomly selected within the received LabelSet;
- **SuggestedLabel (SL) strategy**: in this case the *Path* message issued by *s* includes also the SuggestedLabel object carrying, an indication of the preferred wavelength for the restoration lightpath. The SuggestedLabel is set by *s* to $w_r = W - w_p - 1$, where W is the number of wavelengths on each link and w_p ($w_p \in [0,W-1]$) is the wavelength utilized by the disrupted lightpath before the failure. Upon reception of the *Path* message, *d* selects the wavelength indicated in the SuggestedLabel if it is contained in the received LabelSet, otherwise a random selection is performed within the received LabelSet.

In the reservation phase (backward), if a contention occurs, wavelength conversion can be performed: the node randomly selects a new wavelength among those in the LabelSet of the corresponding *Path* message that was received and stored.

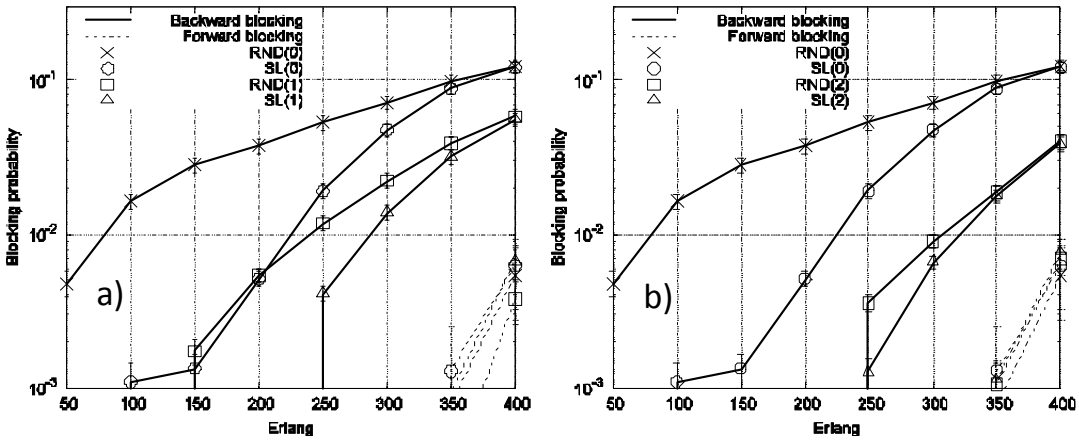


Figure 35: Blocking probability vs. traffic load

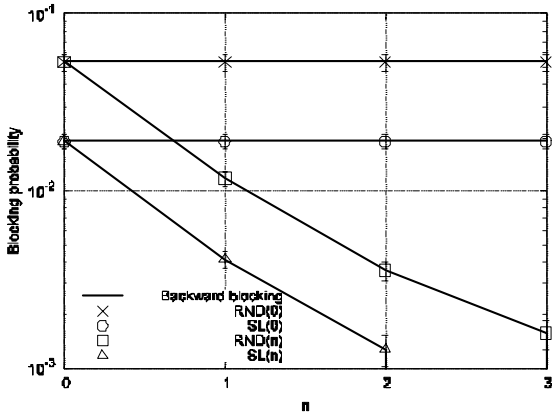


Figure 36: Blocking probability vs. the number n of wavelength converters at 250 Erlang

Performance Evaluation

Restoration performance is evaluated on a Pan European network topology with 27 nodes and 55 bi-directional links, with 40 wavelengths each. It is assumed that each node has the same limited number n of available wavelength converters at the moment of the failure. Wavelength converters are not used during provisioning. Wavelength selection strategies used in restoration are indicated as RND(*n*) and SL(*n*). First-fit wavelength assignment is used during provisioning. Unidirectional lightpath provisioning requests are dynamically generated according to a Poisson process and uniformly distributed among all source-destination pairs. Both inter-arrival and holding time are exponentially distributed with an average of $1/\lambda$=10 s and a varying value of $1/\mu$ respectively. Single-link failures are uniformly and identically distributed on all the links. Each failed link is recovered after that all disrupted lightpaths are restored or blocked. Results are plotted with the confidence intervals at 95% confidence level. Figure 35 shows the blocking probabilities experienced during restoration as a function of the traffic load. The figure confirms that, during restoration, backward blocking dominates even in the presence of wavelength converters. The figure also shows that SL strategy is effective in reducing backward blocking and outperforms RND strategy for a given n. Moreover, it is shown that wavelength conversion permits to strongly reduce the backward blocking. In particular, when n = 1, the utilization of the RND strategy with wavelength conversion is more effective than the utilization of the SL strategy without wavelength

conversion. However, wavelength conversion can be used with great benefits in conjunction with the SL strategy, as shown in Figure 35 (b) where the SL(2) strategy achieves the best performance.

Figure 36 clearly shows that the backward blocking decreases with the increase of n for both RND and SL strategies. For any values of n, the SL strategy guarantees a significant reduction of backward blocking with respect to RND.

In summary, the joint work proved that the use of wavelength converters is effective in avoiding resource contentions among concurrent lightpath restoration attempts in GMPLS-based WSONs. Also, the joint work showed that the use of an intelligent wavelength selection strategy, in conjunction with wavelength converters, further reduces the backward blocking and compensates for the reduced number of available wavelength converters.

2.6.3 Network Performance Improvement in Survivable WDM Networks Considering Physical Layer Constraints

This work focuses on survivable optical networks and studies in detail the network performance improvement that can be achieved when jointly considering network resilience and physical layer constraints. The protection scheme used is path-based shared protection known as backup multiplexing. In the proposed solution routing and wavelength assignment for both primary and protection paths are jointly performed considering their physical performance. Simulations comparing the proposed solution with alternative schemes aiming at maximising sharing of protection resources have shown substantial network performance improvement in terms of blocking probability reduction when jointly addressing resilience and physical layer performance requirements.

Optical networking exploiting wavelength division multiplexing (WDM) is extensively used in existing telecommunications infrastructures and is expected to play a significant role in next generation networks. An important aspect of optical networks particularly in the context of WDM is fault-tolerance, as a single link failure may cause loss of enormous amounts of information. The provision of resilience in WDM optical networks is realized by either proactive protection [Ram99b] or reactive restoration [Dos99]. In addition, traditional routing and wavelength assignment (RWA) algorithms in optical networks make the routing decisions based only on network level parameters such as connectivity and available capacity, without considering the details of the physical layer. When an available path and wavelength are identified, the connection is assumed to be feasible. However, future high speed optical networks are expected to be either fully transparent (signals are transported end-to-end optically) or comprise large domains of transparency. In these networks, the optical signals experience the accumulation of physical impairments through transmission and switching, resulting in some cases in unacceptable signal quality. To address this issue, impairment aware (IA) RWA methods that consider the physical layer impairments have been proposed [Mar07].

Previous work on resilience requirements of traffic requests in WDM networks has revealed that the protection paths are highly susceptible to physical layer impairments as they are commonly longer than the primary paths [Mar08]. This has a direct impact on the overall network performance in terms of blocking probability as a number of protection paths and therefore the corresponding primary paths are blocked due to unacceptable signal quality. To overcome this issue we propose to jointly address resilience and physical layer performance requirements through the design and implementation of a suitable RWA method. The performance of the proposed algorithm is compared with conventional routing approaches

and evaluated through simulations exploring relevant trade-offs. Significant network performance improvement in terms of blocking probability reduction is shown for specific network conditions.

Scenario under Study

The work presented here focuses on proactive protection, i.e. at the time that the primary path is assigned, one or more alternative paths -backup paths- are also identified and the relevant network resources are reserved for protection purposes in case of a failure. The specific protection method applied is path-based and employs shared protection known as backup multiplexing. According to the shared protection scheme and under the single link failure assumption, if two or more primary paths are link-disjoint their protection paths can share the same wavelength channels. The shared path-protection scheme offers improved resource utilization compared to the dedicated-path protection alternative, as introduced in [Han97], while it is still able to offer 100% survivability to a single failure. In addition to taking into consideration the protection requirements of the connection requests, this work jointly performs routing and wavelength assignment for both primary and protection paths considering their physical performance. More precisely, not only the availability of optical bandwidth is considered, before primary connections and their protection paths are established and reserved respectively, but also their quality in terms of bit error rate (BER). The BER of primary and protection paths is calculated through the quality factor Q and compared against a predefined threshold value ($B_{thresh}=10^{-15}$) to decide whether they are of acceptable quality. The analytical model of Q-factor for the performance evaluation of a static unicast IA-RWA has been used to integrate different types of degradations [Li05]. The impairments considered in the Q-factor evaluation include amplified spontaneous emission noise (ASE), cross-phase modulation (XPM) and four-wave mixing (FWM) assuming that they follow a Gaussian distribution. Also, optical filtering and the combined self-phase modulation/group velocity dispersion (SPM/GVD) effects were introduced.

To evaluate the effectiveness of the proposed solution two cases are studied: a) IA-RWA applied for both primary and protection paths and b) IA-RWA used for the primary path and minimum hop routing applied to the protection paths. Simulation results show substantial blocking probability reduction when IA-RWA is used for both primary and protection paths.

Algorithm Specification

Our work has concentrated on solving the online RWA/resilience problem, i.e. traffic requests arrive and get served sequentially without knowledge of future incoming requests. This makes this contribution suitable mostly in the context of traffic engineering. In addition, it is assumed that only a single link failure can occur in the network at any instance of time and re-routing of already established connections is not allowed. The model does not take into consideration any wavelength conversion capability of the network and thus wavelength continuity across any path is a tight constraint in the problem definition.

We assume that all requests have a bandwidth demand of one wavelength unit and for each request a link disjoint backup path is required along with its primary path to provide guaranteed protection. The physical bandwidth of each link (l) can be divided into the following three parts: A_l, B_l, and R_l [Mar08]. Al represents the total amount of reserved bandwidth dedicated to primary paths carried by link l and it is not allowed to be shared. B_l is the total bandwidth occupied by all protection paths on link l and unlike Al it can be shared by protection paths, whose associated primary paths are link disjoint. The residual bandwidth R_l is the difference between the physical bandwidth on link l and the total consumed

bandwidth (A_l + B_l). For any future primary path established on link *l*, R_l is the only available bandwidth that can be used. For setting up a protection path on link *l* for a new primary path a, the available bandwidth S_l consists of two components: the residual bandwidth R_l and the portion of B_l that is able to be shared for carrying this protection path. To identify path costs the relevant link weights are identified for both primary and protection paths. As primary paths do not share bandwidth their cost is the sum of the weight of each link they traverse. In the case of protection paths we give preference to wavelengths that have already been allocated as protection wavelengths by assigning to them a lower weight and therefore reinforce sharing.

The routing and wavelength assignment problems are solved in two separate steps. Routing is implemented based on the Dijkstra's algorithm to compute a primary and a protection path for a given request. The wavelength assignment algorithm selects wavelengths for the primary and protection paths allowing resource sharing between the current request and the already established requests. As explained above, connection requests follow a Poisson arrival process with exponentially distributed time duration. In the initial computation phase a primary lightpath is identified for each request. In this phase impairment aware routing (IAR) is performed by assigning the Q penalty as the link cost and the Dijkstra algorithm is deployed on the weighted graph to calculate the shortest path. If no path is found, the connection is blocked. If at least one path is found, a group of possible wavelengths that can be allocated is identified and the first wavelength is chosen applying the first fit (FF) wavelength assignment algorithm to form the primary lightpath. Furthermore a module that monitors the bit error rate (BER) of the provisioned primary path that checks the path quality is involved and decides whether the path satisfies the quality constraints against the predefined BER threshold. Subsequently, the protection computation phase starts with identifying the portion of the protection bandwidth that can be shared excluding the links that have been already utilized by the primary path. This results in an auxiliary graph representing the current network state. In the case of protection paths two routing algorithms are tested: minimum hop routing reinforcing sharing as described above and IAR. After the link costs are assigned, if no lightpath is found for any wavelength, the connection is blocked due to protection path blocking. In case of discovery of multiple protection lightpaths, the algorithm allocates one wavelength, based on the last fit (LF) wavelength assignment scheme. The LF wavelength assignment algorithm has been applied since it has been shown that when used in conjunction with the FF wavelength assignment algorithm for the primary paths, it maximizes the protection path link reuse [Mar08]. As in the case of primary paths a module that monitors the BER of the selected protection path checks the path quality and decides whether the path satisfies the requirement of the predefined BER threshold.

Performance Study

The results presented in this section, are obtained based on the Pan-European test network defined by COST 239 [Bat00] and NSFNET (Figure 37) and assuming bidirectional fibre links with 16 wavelengths/fibre. Thus, if a link failure occurs the traffic flow in both directions will be disrupted. Particular to the NSFNET topology, regenerators were placed along each link every 600km to avoid unacceptable signal degradation due to physical layer impairments. This is a reasonable assumption to avoid turning entire links unusable due to unacceptably high BER rates. To ensure the validity of results the simulation setup applies an initial transient removal mechanism, whereas sufficient number of experiments is executed with an 80% confidence interval and a statistical error which is less than 20%.

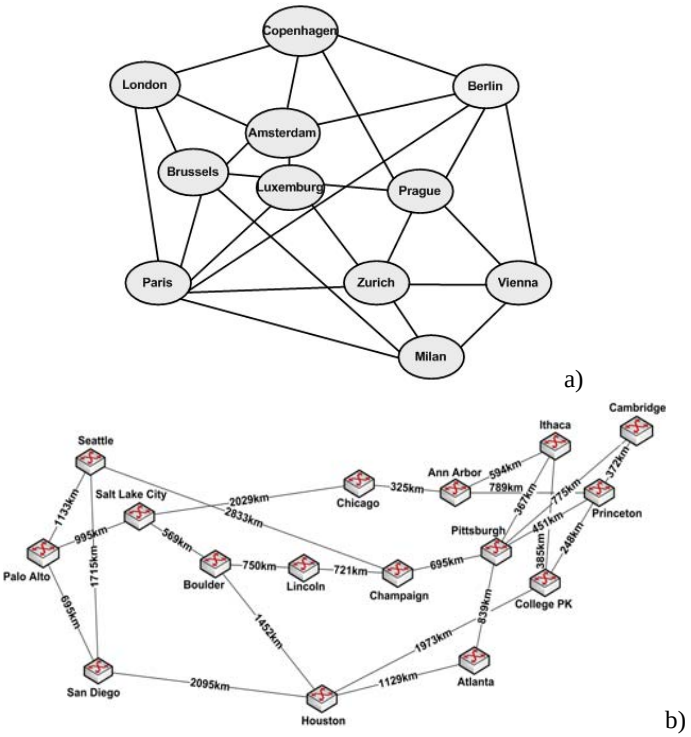
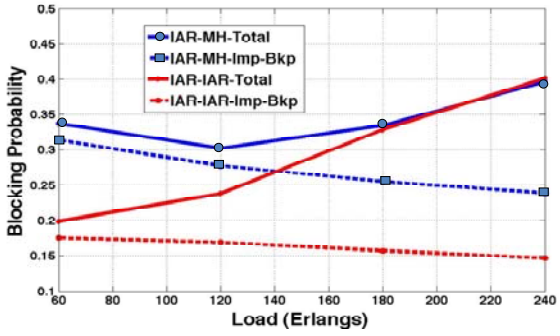


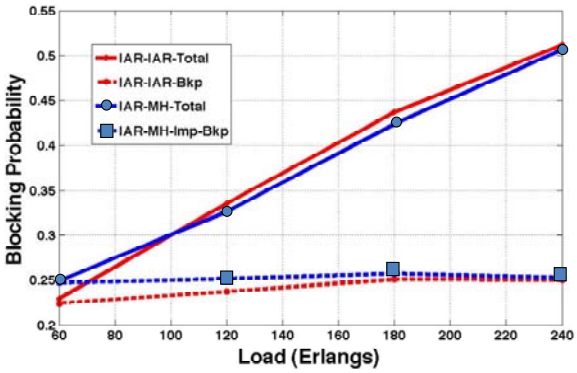
Figure 37: Test network topology defined by (a) COST239and (b) NSFNET

Figure 38 illustrates how the network blocking probability varies with traffic load for both the COST 239 and NSFNET topologies, concentrating on the blocking probability of protection paths and the total blocking probability of both primary and protection paths. The results shown were taken assuming that IAR has been used for the primary paths, while two different routing approaches have been used to discover the protection paths: minimum hop (MH) routing with reinforced wavelength sharing and IAR. These results clearly indicate that it is important to include protection requirements when evaluating network performance since protection capacity allocation gives a significant contribution to the total blocking probability of the network. In addition, Figure 38 demonstrates that even if IAR is used as the routing approach for the primary paths it is important to consider the effect of the physical impairments also in the protection paths. More specifically, in case of minimum hop routing for the protection paths, when BER monitoring is applied to ensure acceptable signal quality, the blocking probability of the protection paths becomes high. This is because a large number of protection paths do not satisfy the BER threshold criterion, which results in blocked connections. It should be noted that in general protection paths are longer than primary paths, exhibiting higher probability to be impaired. This has a significant contribution to the total blocking probability. An alternative approach that can improve the overall network performance is to apply IAR not only to the primary but also to the protection paths. As shown in Figure 38 (a), in case of the COST239 this approach offers blocking probability reduction by 42% for low loading, compared to the MH scheme. The benefit becomes lower for higher loading since in this case there is a smaller reserve of alternative paths that can be exploited; also this is due to the fact that in general MH routing provides the ability to allow a form of load balancing in the network [Mar08]. However, in case of the NSFNET topology, shown in Figure 38 (b), the benefit of the IAR for the protection path is not apparent. This is due to the fact that the NSFNET topology, exhibiting a lower nodal degree on average,

provides fewer possible alternative paths between any source and destination pairs, reducing the effect of IAR in the protection path.



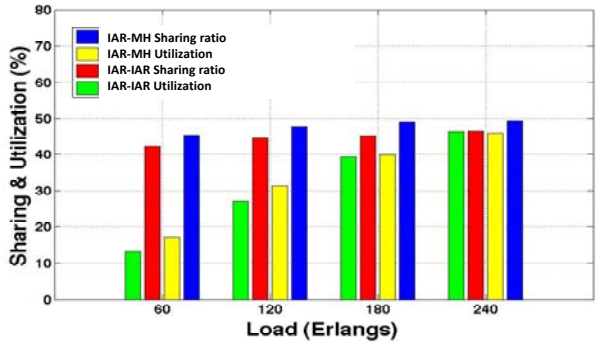
a)



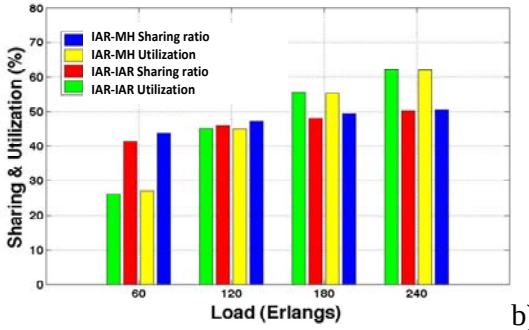
b)

Figure 38: Blocking Probability (total and due to unacceptably high BER at the backup path) for COST239 and (b) NSFNET

Hence, it becomes clear that the use of IAR for the protection paths has a benefit in terms of blocking probability. However, this comes at the expense of network resource sharing. As depicted in Figure 39 (a) in the COST239 case IAR for the protection paths offers blocking probability reduction at the expense of network resource sharing introducing l=less efficient resource utilization. While in the case of NSFNET in which blocking probability was not noticeably reduced the sharing and utilization of network resources also remained at similar levels for the two schemes (MH and IAR).



a)



b)

Figure 39: Sharing ratio and utilization versus load for (a) COST239 topology and (b) NSFNET topology

Conclusions

This work focused on the performance of survivable WDM networks under physical layer constraints. In this context it studied proactive protection and specifically path-based shared protection known as backup multiplexing. In addition to taking into consideration the protection requirements of the connection requests, routing and wavelength assignment for both primary and protection paths were jointly performed considering their physical performance against a predetermined BER threshold. Simulations results have shown substantial network performance improvement in terms of blocking probability reduction when jointly addressing resilience and physical layer performance requirements. This is achieved through suitable selection of routing and wavelength assignments algorithms.

2.6.4 List of publications

- N. Sambo, I. Cerutti, A. Giorgetti, N. Andriolli, P. Castoldi, R. Muñoz, S. Ruepp, R. Casellas, R. Martínez, A.V. Manolova "*Restoration in GMPLS-based Wavelength Switched Optical Networks with Limited Wavelength Converters*", in Proc. of Photonics in Switching (PS), Italy, September 2009.
- A. Tzanakaki, K. Georgakilas, K. Katrinis, L. Wosinska, A. Jirattigalachote and P. Monti, "*Network Performance Improvement in Survivable WDM Networks considering Physical Layer Constraints*", (Invited paper), RONEXT, ICTON 2009, Azores, Portugal, June 2009.
- N. Sambo, I. Cerutti, F. Cugini, A. Giorgetti, L. Valcarenghi, P. Castoldi, "*Segment Restoration Scheme with QoT-guarantees in GMPLS-controlled Translucent Networks*", in Proc. of ECOC 2008, Belgium, September 2008.
- N. Sambo, F. Cugini, N. Andriolli, A. Giorgetti, L. Valcarenghi, P. Castoldi, "*Path Restoration Schemes in GMPLS-controlled Translucent Networks*", in Proc. of Photonics in Switching (PS) 2008, Japan, August 2008.
- R. Martínez, R. Muñoz, R. Casellas, "*Experimental evaluation of the backup sharing aggressiveness for dynamic shared path protection in GMPLS transparent optical networks*", in Proc. of ICTON 2008. Athens (Greece), June 22-26, 2008.
- Demetris Monoyios and Kyriakos Vlachios, "*On the use of genetic algorithms for solving the RWA problem employing the maximum quantity of edge disjoint paths*", in Proc. of ICTON 2008. Athens (Greece), June 22-26, 2008.

- R. Muñoz, R. Casellas, R. Martinez, "*An Experimental Signalling Enhancement to Efficiently Encompass WCC and Backup Sharing in GMPLS-enabled Wavelength-Routed Networks*", in Proc. IEEE ICC 2008, Beijing, China, May 19-23 2008.
- R. Martínez, R. Casellas, R. Muñoz, "*Experimental evaluation of GMPLS enhanced routing for differentiated survivability in all-optical networks*", OSA Journal of Optical Networking. vol. 7, no.5, pp. 496-512, May 2008.

2.7 Resilience Issues in the GMPLS-enabled Control Plane

2.7.1 Control plane resilience: behind reliable connection provisioning (UPC)

The control plane plays an increasingly important role in next-generation transport networks. In fact, not only the signalling functionality is supported on the control plane, but also the routing and management protocols (e.g. the resource discovery protocols seen before) that make connection provisioning to be automated, efficient and cost-effective. Hence, aiming at high network resilience to maximize supported services profitability, a reliable control plane communication becomes essential. Several kinds of failures may appear in the control plane, namely link, node and software failures. Amongst them, control link failures become the most frequent ones.

The vast majority of works on network resilience target at the transport plane. First and foremost, given the nowadays ultra-high transmission rates, milliseconds' failure recovery times may easily lead to terabit data losses. Besides, as the control information has been typically transmitted along with the data traffic (e.g. as in IP or MPLS networks), both control and data planes are equally affected upon failures, which makes no sense to separate both planes resilience. However, in-band control plane configuration is not feasible in all-optical networks, as the end-to-end connections optically bypass all intermediate nodes from source to destination. In view of this, a separation is introduced in GMPLS [Man04] between the control and data planes, so that the control plane can be transmitted on a different wavelength of the same fibre (*in-fibre out-of-band*) or even on a separated network (*out-of-fibre*). Thus, the reliability of the control plane in GMPLS controlled networks becomes no more linked with the one of the data plane. This provides several benefits to network operators, but new challenges are also posed to provide the control plane with the requirements to fulfil emerging services necessities. Among the main benefits, there is an enhanced flexibility in the control deployment or the possibility to design control-plane-driven data plane recovery mechanisms, especially for the out-of-fibre configuration, where the control plane remains alive upon data plane failures. Nonetheless, when the control plane becomes decoupled from the data plane, additional fault detection and recovery mechanisms are required for the former.

Only a few works have so far addressed the resilience of the GMPLS-enabled control plane. Amongst them, [Jaj06] and [Li02] highlighted the reasons of a decoupled control plane in all-optical networks and addressed the new resilience requirements that this would impose. In addition, [Kom08] and [Per07] concluded that the most severe GMPLS protocol disruptions due to message losses (random losses [Kom08] or connectivity outages due to link failures [Per07]) were found in RSVP-TE [Man04]. Comparing the approaches in [Man04] and [Per07], it seems more reasonable to have bursty message losses due to link connectivity outages, rather than random losses due to, e.g., network congestion. In fact, the load in the GMPLS control plane (i.e., RSVP-TE+OSPF-TE+LMP messages [Man04]) should not be very large under normal network operation (connection arrivals in the seconds' or minutes' time scales). In order to evaluate the resilience of a given control plane topology, this work

also focuses on the consequences of the control link failures on a GMPLS-controlled network performance, since these are the most probable ones in transport networks [Gro03]. To this aim, the authors in [Per07] proposed a parameter P_d that stands for the probability that any connection request or tear-down is dropped along the failure recovery time Δt (i.e., forwarded onto the failed control link). Both situations would affect the network Grade of Service (GoS), by either blocking/delaying a connection request, or keeping allocated but not used data plane resources. An analytical P_d formulation in symmetrical ring control planes was presented in [Per07]. As will be reviewed in section III, the final P_d expression depends on the incoming (Poisson) traffic characteristics (λ, μ), Δt, and P_L, which denotes the probability that an incoming connection request/tear-down is supported on the failed control link. Even though ring networks have been extensively deployed over the years, operators are currently moving to deploy meshed network architectures, offering richer connectivity and, thus, enhanced survivability [Gro03]. Therefore, it would be highly desirable to have tools for quantifying the control plane resilience in such scenarios.

$$P_d = 1 - e^{-\lambda \Delta t (1+P_L)} \sum_{k=0}^{C} \binom{C}{k} \left[\left(e^{\mu \Delta t} - 1 \right) (1 - P_L) \right]^k \qquad (1)$$

Equation (1) reproduces the analytical Pd expression obtained in [Per07], where Poisson traffic arrivals to the network were assumed. In this expression, $C \approx \lceil \lambda/\mu \rceil$ identifies the number of active connections in the network at the failure time. Note that the mathematical analysis behind P_d is valid to any network scenario, as it basically depends on the traffic characteristics. The parameter that captures the network topology under study (and the traffic distribution over it) is P_L, which was particularized for symmetrical ring topologies in [Per07]. This section targets at a general P_L expression to allow P_d computation in asymmetrical meshed control planes.

Let G_{DP} (N_{DP}, E_{DP}) and G_{CP} (N_{CP}, E_{CP}) identify the data and control plane graphs of a GMPLS-enabled transport network, respectively. For the ongoing model we assume that G_{DP} is bi connected and planar. In fact, G_{DP} topology can be seen as a set of interconnected sub-rings, that for highly meshed networks can be as small as triangles. We also assume G_{CP} bi-connected, providing survivability to the control plane. Particularly, we restrict the control plane topology to be a subset (or the complete set) of the data plane one. Thus, G_{DP} and G_{CP} can be related as:

$$N_{CP} \equiv N_{DP} \equiv N \qquad (2)$$

$$E_{CP} \subseteq E_{DP} \qquad (3)$$

In this scenario, we define a minimal bi-connected covering topology over G_{DP} (e.g. a Hamiltonian cycle or a minimum n-tree), so that $E^{it}{}_{DP}$ identifies the link subset in this minimal topology and $E^{ot}{}_{DP}$ the subset containing the rest of the data plane links. Hence, $E_{DP} = E^{it}{}_{DP} + E^{ot}{}_{DP}$. In what follows, this additional relation between G_{DP} and G_{CP} is imposed:

$$E_{CP} \supseteq E^{it}{}_{DP} \qquad (4)$$

A *minimal* control plane topology ($E_{CP} \equiv E^{it}{}_{DP}$) is defined. On this basis, any intermediate topology (hereafter, partially meshed) is created by adding links to the *minimal* topology, finally getting the *symmetrical* topology ($E_{CP} \equiv E_{DP}$). From the assumptions above, GCP consists at least on one ring. Every link in $E^{ot}{}_{DP}$ added to E_{CP} creates a new sub-ring, either by sub-ring partitioning (splitting an existing sub-ring in two) or sub-tree closing (adding a new sub-ring external to the minimal topology). In any case, two data plane adjacent nodes will belong to the same sub-ring at the control plane.

Let us define H_{DP} as the average hop length of the data paths. In a similar way, H_{CP} defines the average hop length of the control paths. As the RSVP-TE messages forwarded on the control plane should visit (i.e., configure) the same node sequence comprised in the computed data plane route, H_{DP} becomes a function of G_{DP} and G_{CP}.

At this point, we can define $P_L = D_L/D_T$, that is, the ratio between the amount of demands supported in the failed link L (D_L) with respect to the total number of demands (D_T). This finally leads to:

$$P_L = \frac{D_L}{D_T} = \frac{C \cdot H_{CP} / |E_{CP}|}{C} = \frac{H_{CP}}{|E_{CP}|} \qquad (5)$$

As shown, P_L directly depends on the average hop length of control plane paths. As mentioned above, end-to-end RSVPTE messages are processed hop-by-hop at every node in the route of the Label Switched Path (LSP) being signalled/torn-down. As a consequence of equation (3), adjacent nodes in the data plane may be not adjacent in the control plane. Thus, H_{CP} is proportional to H_{DP}, and can be expressed as:

$$H_{CP} = \tau \cdot H_{DP} \qquad (6)$$

where the parameter τ adjusts the distance (the number of hops) in the control plane between two adjacent nodes in the data plane. Without loss of generality, we consider that every demand is routed through the shortest path. Besides, as in [Per07], we assume the traffic uniformly distributed in the network. Then, the average length of the shortest paths in a mesh network can be approximated by [Kor04]:

$$H_{DP} \approx \sqrt{\frac{|N|-2}{\delta_{DP}-1}} \qquad (7)$$

where δ_{DP} is the average node degree in the data plane. To calculate τ we compute the distance at the control plane of all adjacent node pairs at the data plane. Being also adjacent at the control plane their distance equals to 1. Otherwise, their distance in the control plane ($h_j^{G_{CP}}$) is computed. Finally, τ can be expressed as:

$$\tau = \left(\sum_{\forall i \in E_{CP}} 1 + \sum_{\forall j \in E_{DP} \setminus E_{CP}} h_j^{G_{CP}} \right) \frac{1}{|E_{DP}|} = \frac{|E_{CP}|}{|E_{DP}|} + \frac{|E_{DP}| - |E_{CP}|}{|E_{DP}|} \cdot \frac{\sum_{\forall j \in E_{DP} \setminus E_{CP}} h_j^{G_{CP}}}{|E_{DP}| - |E_{CP}|} = \alpha + (1-\alpha) \cdot \kappa \qquad (8)$$

where α is the proportion of links at the control plane to those at the data plane, and k represents the average distance of nonadjacent nodes at the control plane. We have focused on a minimal G_{CP} topology consisting on a Hamiltonian cycle, where the average lengths of E^{ot}_{DP} and E^{it}_{DP} links are similar. There, we have concluded (after several tests) that k can be accurately estimated as $(|N|)^{1/2}$. In a more general case, every sub-ring in the control plane acts as a cycle covering a subset of nodes of G_{DP}. Based on the previous results, we approximate $k \approx (V_{CP})^{1/2}$, where V_{CP} is the mean number of nodes in a sub-ring.

As mentioned before, every pair of adjacent nodes at the data plane belongs to the same sub ring at the control plane. Let R_{CP} denote the number of sub-rings at the control plane, and T_{CP} the sum of nodes in every individual sub-ring.

Thus, V_{CP} satisfies:

$$V_{CP} = \left\lceil \frac{T_{CP}}{R_{CP}} \right\rceil \qquad (9)$$

$$T_{CP} \approx \left|E_{DP}^{it}\right| + 2 \cdot \left|E_{DP}^{ot}\right| = 2 \cdot \left|E_{CP}\right| - \left|E_{DP}^{it}\right| \quad (10)$$

$$R_{CP} = \left|E_{CP}\right| - \left|N\right| + 1 \quad (11)$$

Note that equation (10) gives an exact TCP value when all sub-rings have been created by sub ring partitioning. In any other case, however, it still represents a valid approximation, since sub-ring partitioning is much more frequent than sub-tree closing. Finally, combining equations (6), (7), (8) and (9), P_L can be stated as:

$$P_L \cong \left[\alpha + (1-\alpha) \cdot \sqrt{V_{CP}}\right] \cdot \sqrt{\frac{|N|-2}{\delta_{DP}-1}} \cdot \frac{1}{|E_{CP}|} \quad (12)$$

The obtained P_d model has been validated over different networks with different average node degrees. To this end, we consider a quite sparse 28-Node NSFNET topology, a moderately meshed 14-Node Deutsche Telekom (DT) network, and a highly meshed 28-Node European Optical Network (EON). Besides, for each topology, we define four different control plane alternatives: a *symmetrical* topology, a *minimal* topology, and two partially meshed topologies in between. Table 3 reviews the most relevant parameters of each topology under evaluation. The column on the right presents $|\mathrm{ECP}|$ in the *symmetrical*, partially meshed 1, partially meshed 2 and *minimal* topologies, respectively. The performance of the model has been validated by simulation results. For them, enough wavelengths per link to guarantee that all requests are routed through the shortest path (accomplishing the wavelength continuity constraint) are assumed. In such scenarios, uniformly distributed connection requests arrive at each node following a Poisson process, and connection holding times are exponentially distributed.

	$\|N\|$	$\|E_{DP}\|$	δ_{DP}	$\|E_{CP}\|$
NSFNET	28	37	2.64	37 - 34 - 31 - 29
DT	14	23	3.28	23 - 20 - 17 - 14
EON	28	61	4.36	61 - 41 - 34 - 28

Table 3: Network topology parameters

The model and the simulation results for P_d as a function of Δt are plotted in Figure 40. Each simulation is conducted in order to reach steady state results within a 95% confidence interval. As seen, the P_d model and the simulation results are really close in every experimented topology.

Aiming to measure the discrepancy between the obtained P_d values and the expected ones, we have computed the Chi-square goodness of fit test in each scenario. To this goal, we compare the number of affected connections obtained by simulation with respect to the expected value of this variable (i.e., multiplying the P_d analytical value by the number of total simulated connections). In all cases, the null hypothesis can be clearly accepted (the difference between simulation and analytical results is zero), which highlights the accuracy of the model.

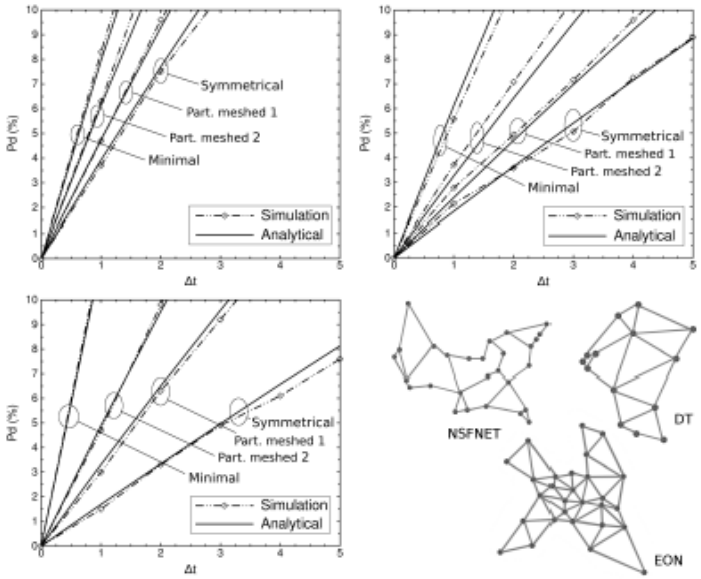


Figure 40: Model vs. simulation results: NSFNET (top left); DT (top right)

Motivated by the necessity of quality of resilience parameters, P_d could be proposed to quantify the maximum recovery time to meet certain control plane resilience requirements (i.e., a certain P_d value). In particular, the minimal topology requires very restrictive Δt values (Fig. 1). Since multiple demands are supported on each control link, the performance degradation caused by control link failures is very high. For instance, aiming at a P_d = 5% objective in the 28-Node EON, Δt < 500 ms must be assured. However, by increasing the connectivity at the control plane, P_d steadily decreases. In the symmetrical topology, as only one demand is supported on each control link, Δt ≈ 3 s already fits P_d = 5%. Between both extremes we have the partially meshed topologies, which target at a trade-off between resilience and required resources. Network operators could benefit from the proposed model to quantify the number of control plane links needed to fit certain P_d requirements, given a Δt achievable by their control plane recovery mechanisms (e.g., IP layer re-routing, dedicated link protection). This value could be afterwards used as input data for an optimal control plane topology design.

2.7.2 Implementation and evaluation of GMPLS-like Control Plane (AGH)

The main aim of this activity was to study and validate suitable failure detection and recovery mechanisms for the GMPLS-enabled control plane, achieving the resilience level required for next-generation optical transport networks performance.

Besides analysis of necessary GMPLS enhancements there were also efforts to practically implement the open-source DRAGON software, enabling cooperation between Control Plane and Transport Plane must be analysed. DRAGON (Dynamic Resource Allocation via GMPLS Optical Networks) is an US-based project with the main aim to set-up and manages dynamic, end-to-end network transport services for high-end e-Science applications. The main motivation to use DRAGON solution is to test rather low-cost, yet credible solution mapping the functionalities of GMPLS messages (i.e. transferred through Control Plane) onto Transport Plane, e.g. via SNMP (Simple Network Management Protocol).

We have adapted the DRAGON software to support Catalyst 3560 series switches and created own GMPLS network named KT-GMPLS. During first phase, reported in previous WP22 deliverable D22.2, two scenarios with different topologies and amount of elements in CP and TP, supporting L2SC type of switching were tested successfully and then reported. First scenario was single-domain, no-NARB network, the second was single-domain, NARB-enabled network.

The main building blocks for DRAGON are: VLSR (Virtual Label Switching Router) which is responsible for participation in GMPLS protocols' exchanges and provision for supervised switch according to protocol events (PATH setup, PATH tear down, etc.). The NARB&RCE server (Network Aware Resource Broker with Resource Computation Element) is equivalent to PCE (Path Computation Element). NARB is responsible for path computation, inter-domain routing and intra-domain listening. CSA (Client System Agent) is the end system or client software responsible for signalling towards network (UNI or peer mode) and participate in path computation.

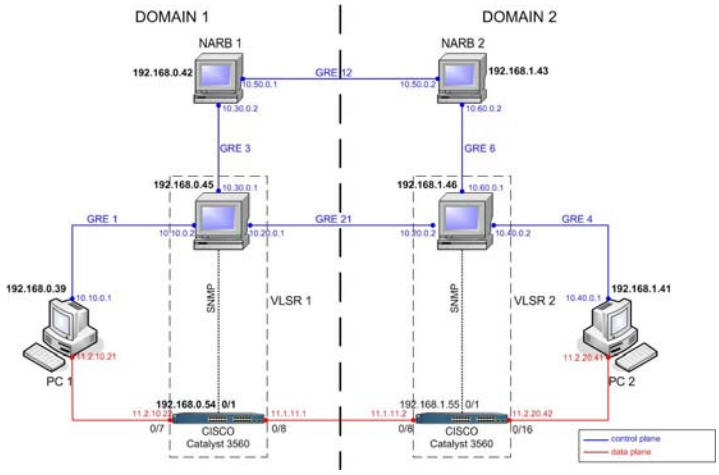


Figure 41: Multidomain scenario tested in KT-GMPLS network

The network shown in Figure 41 consists of two domains, each of them contains:

- End System Client – PC1 and PC2 (zebra, ospfd, RSVP, dragon demons running);
- VLSR router – PCs configured as virtual routers which manage (using SNMP) the CISCO Catalyst 3560 device (zebra, ospfd, RSVP, dragon demons running);
- NARB – PC with relevant software (zebra, ospfd-interdomain, ospfd-intradomain, rce, narb demons running);

We can see one inter-domain TE link addressed (Figure 41): 11.2.30.30/32 where the LSPs between PC1 which is also called CSA1 (Client System Agent) and PC2 acting as CSA2 were created. The intra- and inter-domain OSPF LSA (Link State Advertisement) are flooded through GRE tunnels.

With inter-domain OSPF configured, NARB will check the status of the OSPF adjacency for the originating interface. NARB will originate a topology only if the OSPF adjacency on this originate-interface is in Full state. Otherwise, NARB will enter a loop waiting for the OSPF adjacency.

For the LSP set-up process, the CSA1 acts as a NARB API client and sends an LSP query request to NARB1, which co-operating with RCE and NARB2 computes a full explicit route

for this request. Then RSVPD can use the obtained ERO (Explicit Route Object) to signal up an end-to-end LSP based on Ethernet tagged VLAN from PC1 to PC2 crossing the two domains.

2.7.3 List of publications

- M. Ruiz, J. Perelló, L. Velasco, S. Spadaro, J. Comellas, "*An Analytical Model for GMPLS Control Plane Resilience Quantification*", IEEE Communications Letters, Vol. 13, nº 12, December 2009.

2.8 Multi-domain provisioning/recovery in GMPLS all-optical networks

This Joint Activity is focused on the experimental validation and evaluation of different GMPLS restoration strategies operating at several granularities such as link, node and wavelength channel within a multi-domain Wavelength Switched Optical Network (WSON). In a WSON, a link or node failure causes a huge amount of data loss. Therefore, fast and efficient recovery schemes are needed to minimize these effects and recover the disrupted services. In particular, this JA concentrates on restoration schemes where the backup path is computed and set up after the failure is detected, localized and notified to the responsible to recover the optical connection (i.e., lightpath). This allows and efficient utilization of the network capacity at the expense, however, of compromising both the restoration success and time. WSON restoration has been extensively studied in the last years. In general, these works focused on restoration within a single domain where the routing schemes for recovery are aware of the complete topology and network resource status, allowing high restoration efficiency. However, in this JA, restoration is addressed in multi-domain network scenarios. In such a network, only abstracted topology and reachability information is shared among the network domains. While each domain is, in principle, responsible for the routing of the path segment traversing its respective network, the restoration of the end-to-end lightpaths may be far from the optimal yielding to an inefficient use of the overall resources. We consider that the lightpaths are transparently (i.e., within the optical layer) set up regardless of the traversed domains. That is, no optical-electronic conversions are used. In this regard, the working and backup paths must fulfill the wavelength continuity constraint (WCC) since no all-optical wavelength converters (WCs) are placed.

In the GMPLS restoration, once a failure occurs, the selected nodes to repair the failed lightpaths/s are notified with different information levels regarding the failed resources: no information, wavelength, link and node basis. This granularity allows devising different restoration strategies which operate at each of these information levels. The goal of this JA is to validate the feasibility of the GMPLS protocols to restore lightpaths within multi-domain transparent WSON when using one of these restoration strategies. Specifically, when using information at either link or wavelength channel granularity. The validation is experimentally carried put within a multi-domain network connecting at the control plane level both CTTC ADRENALINE testbed® and UPC CARISMA testbed.

2.8.1 Multi-domain restoration strategies: problem statement

This JA considers a multi-domain network formed by several OSPF-TE areas connected through the backbone area 0. The nodes connecting two or more areas are the *Area Border Routers* (ABRs). The ABRs allow summarizing the flooded topology information exchanged among domains. In consequence, a domain has a limited visibility of the topology and

resource status in the other domains. The routing of lightpaths (*Label Switched Paths*, LSPs in GMPLS) is thus attained in a *per-domain* basis. That is, for end-to-end LSPs spanning multiple domains, each ABR along the route computes, using its own intra-domain TE information, the segment of the LSP to the next egress ABR until the destination is reached. Therefore, a route expansion is required at each traversed ABR.

In the RSVP-TE signalling protocol, a set of Notify_Request objects may be added to the Path and/or Resv messages indicating the node IP addresses to be notified when a LSP failure occurs. To this end, GMPLS uses the so-called Notify message. The node receiving such a message is termed as *Point of Repair* (PoR), and is one of the responsible to restore the failed LSP. In the JA, it is assumed that the source and the ABRs traversed by a LSP act as PoRs.

In the Notify message, the inclusion of failure information (e.g., node, link, or wavelength) allows the PoR to avoid the failed resources when computing an alternate path. Typically, such information is conveyed during the backup LSP setup as the *eXclude Route Object* (XRO) within the Path message, so nodes doing a route expansion (e.g. ABRs) may avoid these failed resources. However, in multi-domain networks, it may be that a PoR within a given domain does not have sufficient topology information to compute a strict *Explicit Route* excluding the failed resources conveyed in the XRO object. This has an impact on the efficiency of the restoration and may require more advanced topology aggregation mechanisms using a hierarchical routing.

Let's use Figure 42 to analyze the impact of reporting the failure information to the PoRs when restoring a LSP. Two GMPLS domains (CTTC and UPC) are connected through the area 0. A working LSP is established along the path formed by nodes 4, 2, 8, 9 and 10. A failure occurs between nodes 9 and 10. Three different failure types may occur: a wavelength channel (e.g., problem on a receiver of any of the OXC ports), link (e.g., fiber cut or optical amplifier problem) and node (i.e., optical switch). A Notify message is sent upstream by the node 9 reporting the failure to the PoR (i.e., node 8). This PoR is required to compute an alternate route, within the UPC domain, to detour the failure, maintaining the same wavelength. In the example, we assume that node 8 cannot restore the LSP due to either the lack of resources or the WCC failure. Thus, the Notify message is sent to the upstream PoR of the route (i.e., node 2). Let's consider that the node 8 removes/filters from the Notify message the failure information to avoid sharing such information among domains. In consequence, node 2 is only aware that node 8 needs to be avoided as long as the same wavelength must be kept (due to the WCC), since it cannot restore. It is worth noting that if wavelength converters were available in node 2, node 8 may be traversed with another wavelength. Thereby, node 2 is also unable to route the LSP within the area 0. Note that the other outgoing links from node 2 (i.e., with nodes 4 and 1) do not belong to the area 0. The Notify message is finally sent to the upstream PoR (i.e., node 4). In this case, node 4 (the source) can apply two routing policies: *pessimistic* and *optimistic*. In the former, the route detours the node 2 (i.e., path along the nodes 4, 3 and 1) and considers any available wavelength including the failed one. In the latter, the route does not detour node 2 but exclude the failed wavelength channel. Focusing on the pessimistic policy, a route expansion is needed at node 1. Since no XRO information is carried in the Path message, node 1 has two routing options: through node 8 or node 7. In the first, the restoration LSP would definitely fail. In the second option, the route expansion done at node 7 could route through either node 11 or node 9. Note that through node 9 the LSP would also fail. Hence, we observe that filtering information between domains may lead to block the restoration of LSPs even if a feasible backup LSP exists.

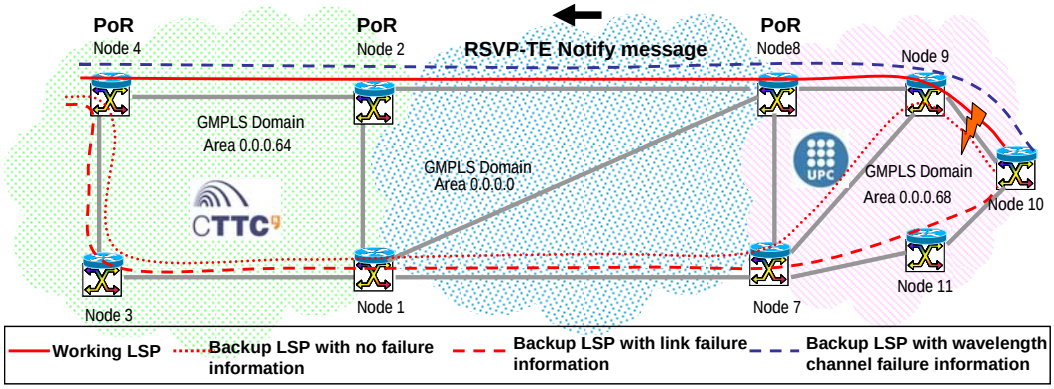


Figure 42: Multi-domain LSP restoration with either no failure information and/or with link/wavelength-based information shared among domains

In light of the above, to increase the LSP restorability, failure information needs to be exchanged beyond the domain boundaries. Accordingly, node 8 sends the Notify message to the neighbouring domain without filtering the failure information. In case of a link failure, the Notify message specifies that failed link (i.e., 9-10). Node 2 receives the Notify message and, again, node 2 cannot route the LSP due to the WCC. Consequently, the Notify message is sent to node 4. Assuming the pessimistic policy, the backup path is computed along the nodes 4, 3 and 1 within the CTTC domain. The Path message contains the XRO to inform to the subsequent route expansions (nodes 1 and 7) about the nodes (i.e., 2 and 8) and the links (i.e., 9-10) to be avoided. This allows computing a feasible backup LSP along the nodes 4, 3, 1, 7, 11 and 10.

The above solution is adequate when the failures are either link or node. However, if the working wavelength channel fails in the link between nodes 9 and 10, the complete link should not be discarded for the backup LSP. Indeed, the backup LSP could be set up using a different wavelength channel on that link. The only restriction is that the WCC for the backup LSP must be satisfied. Thus, the source node may consider the optimistic approach (i.e., not detouring node 2 but without using the failed wavelength), based on a policy decision. In our example, the node 8 receives the Notify message including the link and the failed wavelength channel. Due to the WCC, node 8 sends the Notify message to the next PoR (node 2). Since node 2 is also unable to restore the LSP, a Notify message is finally sent to the node 4. As stated, node 4 may or may not reroute via 2. Let's assume not detouring is performed (optimistic policy). In this case, the Path message to set up the backup LSP includes the XRO which indicates the link and the failed wavelength channel. This information is taken into account within the UPC domain when performing the route expansion at node 8. In other words, until reaching such a node, the carried XRO information does not affect the needed path computations. However, at node 8, the wavelength channel failure information is used to constraint the path computation and the wavelength assignment algorithm. In the example, the computed segment path is formed by nodes 8, 9 and 10, in which the wavelength assignment algorithm excludes the failed wavelength channel from the *Label Set* object. By doing this, usable wavelength channels on the link between nodes 9 and 10 are not discarded, and the LSP restoration may attain a better use of all the resources.

In short, in a WSON with WCC, coarse granularity regarding link and node failures prevents the source node from deciding whether to reuse part of the failed path with an alternate wavelength or to exclude it completely. This choice is enabled by disseminating finer granularity failure information (i.e. at the wavelength level). Additionally, more efficient

restoration mechanism can be deployed even in multi-domain scenarios knowing whether it was a failed wavelength/link/node without knowing exactly which one.

2.8.2 Experimental setup and validation

Figure 43 shows the network used to validate the feasibility of the GMPLS protocol for dynamically restoring the LSPs in a multi-domain WSON. Two domain networks at the control plane level, namely, the CTTC ADRENALINE and the UPC CARISMA testbeds are considered.

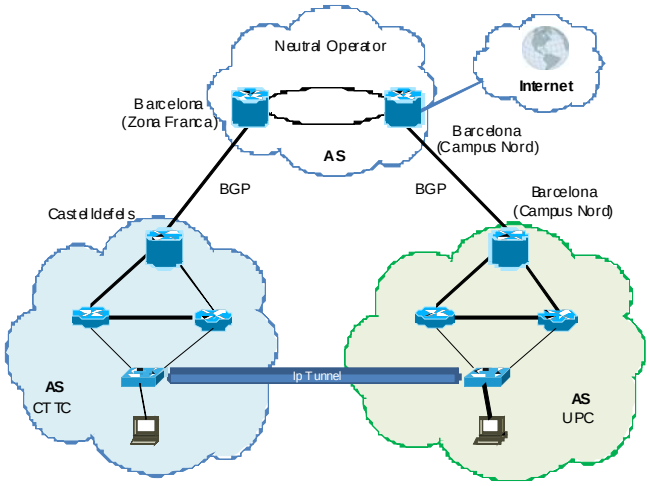


Figure 43: Simplified scheme of the IP network interconnection

The considered transport topology is depicted in Figure 44. It is formed by 9 optical switches (i.e., OXCs) with 13 TE links which are divided into three domains: CTTC, UPC and Area 0.0.0.0. Each link supports 8 wavelength channels operating at 10 Gb/s. In order to compute the path for a new incoming connection request, the used routing strategy is based on a modified Dijkstra algorithm. This algorithm computes the shortest path in terms of hops satisfying the WCC. Recall that for LSPs encompassing more than one domain, this routing algorithm operates in a per-domain basis, and thus is executed at each node expanding the route. Furthermore, for the restoration purposes, as long as the XRO is present, such information is considered constraining the path computation.

Besides verifying the intra- and inter-domain routing information, the conducted tests validate the exchange of the RSVP-TE messages for notifying the same failure but a different information granularity: link and wavelength channel. Accordingly two restoration strategies are used. The working LSP is set up through the nodes 4, 2, 8, 9 and 10 and occupies the wavelength label id. 637534212. The failure is generated at the working wavelength channel over the link between nodes 9 and 10.

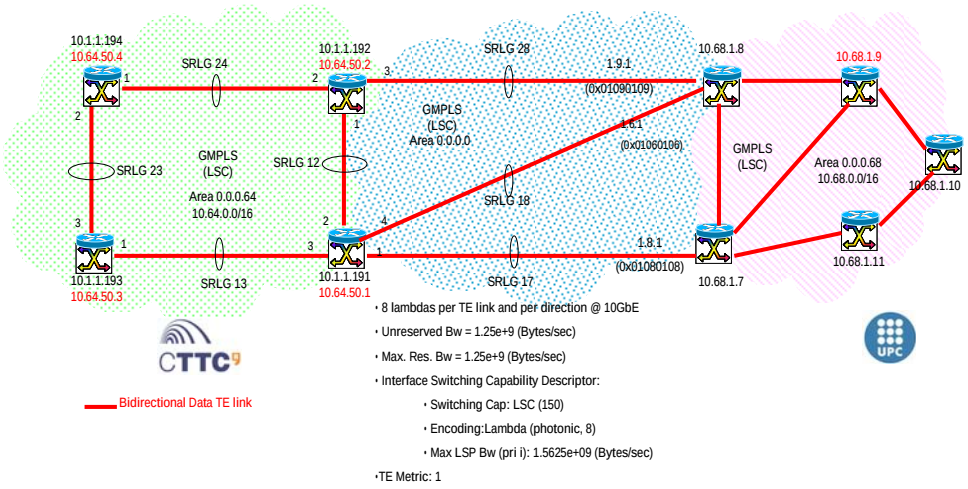


Figure 44: Emulated transport network topology between CTTC ADRENALINE and UPC CARISMA testbeds

Figure 45 depicts the RSVP-TE messages captured at node 8 when the failure information is notified at the wavelength channel basis. We observe that after the Notify message, the working LSP is torn down (see message frames 9-12). Next the backup LSP is set up along the same route as the working path (see message frame 13). In that Path message, the XRO indicates the resource at the wavelength channel (i.e., link and label id) to be avoided when performing the route expansion.

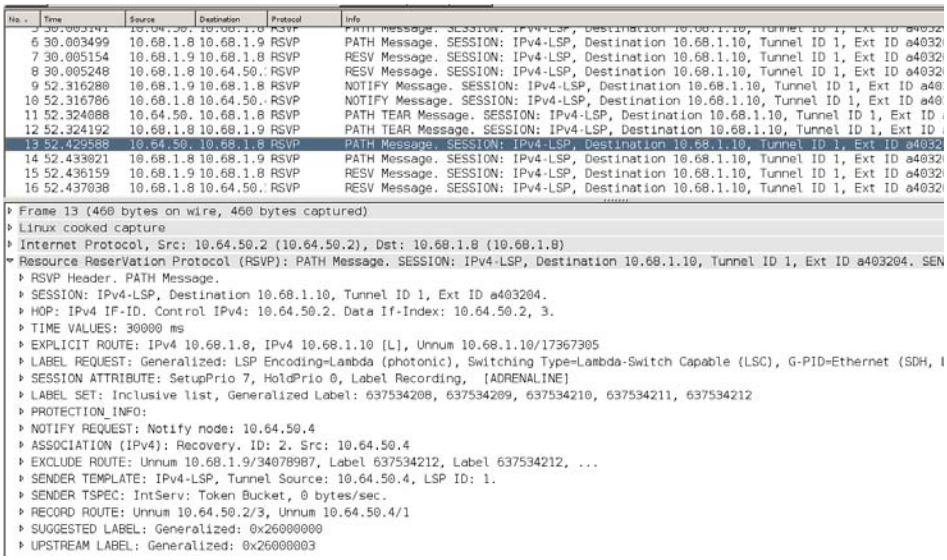


Figure 45: RSVP-TE messages at node 8 for the LSP restoration at wavelength channel

Table 4 gathers the results for both validate restoration strategies in terms of the computed backup LSPs along with their respective restoration time. As said, using failure information at wavelength channel granularity does not discard usable wavelength channels on specific links. Furthermore, it allows increasing the LSP restorability ratio as well as computing shorter (in terms of traversed hops and links) backup LSPs. The latter leads to both lower the restoration time and increase the likelihood of satisfying the WCC. However, these benefits

are achieved at the expense of increasing the signalling control overhead and compromising the confidentiality among domains.

LSP Restoration using	Working LSP	Backup LSP	Restoration time (msec)
Failure information at wavelength channel basis	Nodes 4, 2, 8, 9 and 10; label id: 637534212	Nodes 4, 2, 8, 9 and 10: label id: 637534215	Around 135 msec
Failure information at link basis	Nodes 4, 2, 8, 9 and 10; label id: 637534212	Nodes 4, 3, 1, 7 11 and 10: label id: 637534215	Around 185 msec

Table 4: Numerical results

2.8.3 Conclusions

In this JA, we have addressed the GMPLS restoration within multi-domain transparent WSON. The strengths of such a work are twofold: first, the restoration in multi-domain scenarios is considered, at the time being, as a hot research topic. Indeed, an interesting aspect is to focus on the trade-off between scalability/confidentiality and restoration performance. Second, the validation and evaluation of the GMPLS-enabled restoration strategies were experimentally conducted through interconnecting two testbeds (i.e., UPC CARISMA and CTTC ADRENALINE).

We believe that the Joint Activity has met the original goals that were set when it was defined. It has successfully combined development, experimentation and model validation on a key subject and hot topic.

2.8.4 List of publications

- R. Martínez, R. Casellas, L. Velasco, F. Agraz, R. Muñoz, S. Spadaro, "*Experimental validation of end-to-end GMPLS-enabled restoration in multi-domain transparent WSON*", accepted for publication to the OFC/NFOEC 2010, 21-25 March 2010 – San Diego, CA, USA.

2.9 GMPLS-based control plane for optical packet-based technologies

Optical Burst Switching (OBS) and Optical Packet Switching (OPS) networks need to be capable to be rapidly reconfigured with the aim of achieving an efficient use of bandwidth, low latency and high degree of transparency. However, the bufferless architecture and the one way (on the fly) reservation scheme intrinsic of the OBS/OPS networks bring several challenges to its development. Several mechanisms have been proposed to improve the OBS/OPS performance: there are generally based on including more intelligence in the switching layer. Nonetheless, these complex control processes together with the highly dynamics of OBS/OPS networks make impossible the objective of achieving fast per-burst decisions.

In our proposal, the idea is to move the intelligence to the control plane keeping the switching layer only responsible of local decisions with limited choices. Our challenge is therefore the definition of a Control Plane, which must be able to respond to the just mentioned highly dynamic and complex control requirements. In line with this, although its features fit with

wavelength switched networks, GMPLS could be considered as a reference to design such OBS/OPS-capable control plane. The adaptation/interoperation of GMPLS and OBS/OPS is catching the research community attention. Several recent research relevant papers [Lon06], [Man07] and [Guo07] deal with the design of a multi-layer network architecture for interworking GMPLS and OBS networks.

This fact, together with our convincement that the deployment of OBS/OPS will necessarily take place from the migration of OCS networks, and for this reason they will need to coexist in the transition, this activity is addressing the problem of designing a GMPLS-controlled OBS/OPS network.

In this second year, we continue the work on the GMPLS-based OBS architecture described in D22.2 [D222]. In particular, we design two different schemes to set up and maintain the TE-tunnels with QoS guarantees: 1) a scheme able to create multi-bus channels, 2) a scheme able to create multi-tree channels. In the following section we remind the details of the GMPLS-based OBS architecture followed by the brief description of the two schemes.

2.9.1 GMPLS-based OBS architecture

The GMPLS control layer works as an overlay control network (uses out-of-band, either out-of-fibre or in-fibre signalling), which is in charge of configuring a virtual topology for the OBS network. Its purpose is setting up and tearing down Traffic Engineering tunnels (TE-tunnels) –in our context, a TE-tunnel is a group of wavelengths, representing one or multiple parallel LSPs established in a single signalling session, and the whole set of established TE-tunnels can be seen as a virtual overlay network where to route the data bursts–. The layer for the OBS specific control functions interact directly with the data plane; meaning that if a TE-tunnel has W wavelengths available, one wavelength is reserved for transmitting the Burst Control Packets (BCPs), while the rest (W-1 wavelengths) can be allocated (by the BCPs) for transmitting the data bursts.

Consequently, under such an architecture, the OBS network approximates the connection oriented behaviour, i.e., the source-destination path is determined by the OBS source node among the available preset TE-tunnels reaching the desired destination; but the wavelength to send the bursts are chosen by the BCPs at each transit node along the selected path (i.e., the TE-tunnel), meaning that a burst can be switched from one wavelength to another (always within the same TE-tunnel) according to contention avoidance policies or occupancy ratio. In this way, we achieve the idea of keeping the switching layer as fast as possible since only simple, local and limited decisions are required (select a wavelength among a set of pre-selected ones).

2.9.2 Addressed schemes

For both alternatives, we address the problem of optimizing the wavelength allocation in an OBS network subject to given (absolute) QoS constraint. More specifically, we are looking for such network routing that for given set of traffic demands and end-to-end requirements on the burst loss rate, minimizes the usage of wavelength in the network. The result of the optimization problem is a set of TE-tunnel connecting the nodes properly. The key aspect is to determine the number of wavelengths required in each link of the TE-tunnels.

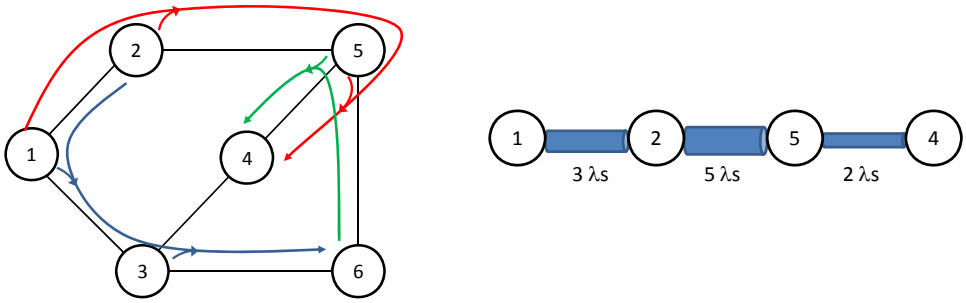


Figure 46: left) Example of three TE-tunnels established with the multi-bus scheme; right) Example of the wavelength allocation per link in a bus connection four nodes

For the case of multi-bus scheme, a TE-tunnel is a unidirectional bus. Figure 46 shows an example. Three TE-tunnels have been established according to the traffic matrix: one TE-tunnel connects node 1, 2, 5 and 4; the second TE-tunnel connects node 6, 5 and 4; the third TE-tunnel connects node 2, 1, 3, and 6. These TE-tunnels allow the transmission of the bursts in downstream direction, i.e., in the first bus, node 1 can send bursts to all other nodes of the bus, node 2 only to its downstream node (5 and 4), and so on. It is worth to mention that the nodes apply a MAC protocol to avoid burst contention. In such a way there are no losses in the optical domain since the intermediate nodes check the resource availability in the bus before the transmission of the burst. Besides the TE-tunnels, the scheme also determines the minimum number of wavelengths required between nodes to allow the transmission of the bursts with zero losses.

For the case of multi-tree scheme, a TE-tunnel is a unidirectional tree. Figure 47 shows an example. Three TE-tunnels have been established according to the traffic matrix: one TE-tunnel connects node 1, 3, 6 and 4; the second TE-tunnel connects node 6, 5, 4, 2, and 1; the third TE-tunnel connects node 2, 4, 5, and 6. As in the previous scheme, the TE-tunnels allow the transmission of the bursts in downstream direction. The difference in this case is that two (or more) leafs of the tree may merge in an intermediate node which can experience burst contention and, consequently, burst losses. The key point of this scheme is that these burst losses are bounded to an (absolute) level by the optimization problem which determines the minimum number of wavelengths required between nodes to allow the transmission of the bursts with such level.

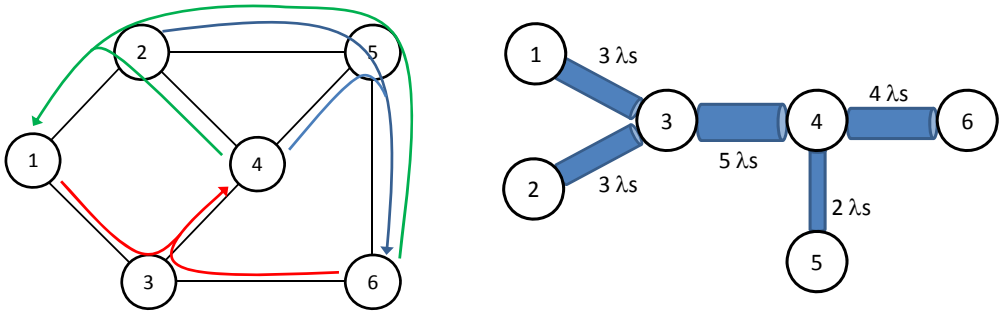


Figure 47: left) Example of three TE-tunnels established with the multi-tree scheme; right) Example of the wavelength allocation per link in a tree connection six nodes

2.9.3 Conclusions

This JA is almost finalised. Two final steps need to be completed yet: the harmonisation of the set of results achieved with the above described schemes and the preparation of two papers.

2.9.4 List of publications

- P. Pedrosa, D. Careglio, R. Casellas, M. Klinkowski, J. Solé-Pareta, "*An interoperable GMPLS/OBS Control Plane: RSVP and OSPF extensions proposal*", in Proc. of 1st Colloquium on Photonic Communication Systems and Networks (PCSN2008), Graz, Austria, July 23-25, 2008.

2.10 Bidirectional service signalling in GMPLS networks

During the second year of this Joint Activity, the study focused on two main topics. In the first task, we investigated the benefits of prioritizing bidirectional connection requests aiming at decreasing the use of WCs. In the second task, we proposed and evaluated a PCE-based architecture for bidirectional lightpath set up, improving the performance of the distributed architecture with both standard and enhanced signalling schemes.

2.10.1 Prioritization of Bidirectional Connection Requests in GMPLS Optical Networks

The Wavelength Division Multiplexing (WDM) technology has shown to be the primary solution for fulfilling the ever increasing demand for capacity in optical transport networks. To setup a connection in such a network a route and a wavelength must be identified for each connection request. This process is referred to as the routing and wavelength assignment (RWA) problem. If possible, connections are allocated on continuous wavelength paths (wavelength continuity constraint). This is due to the fact that even though emerging technologies allow for the conversion between wavelengths [Ram98], these devices are still very expensive and therefore only a limited number of wavelength converters (WCs) is likely to be introduced in the networks.

Previous studies have shown that economic use of these WCs significantly decreases the blocking probability during the connection provisioning phase [And06, Kos08, Rue08]. If all connection demands are known beforehand, the RWA problem can be solved off-line, allowing for the globally most resource efficient assignment of routes and wavelengths. However, in real-life networks traffic demands most often arrive dynamically, which obstructs the aforementioned global optimization, since decisions on connection allocation must be made on the fly without knowledge of other ongoing and future connection demands. There are however two situations in particular, where such knowledge is available so that the off-line and the dynamic RWA can be favourably combined: first, if a set of connections have to be admitted to the network all at once as a bulk; and second, if reconfiguration of the network is carried out [Ros08]. In this sub-activity, we therefore show how dynamic and off-line RWA can be combined to save critical WC resources.

In optical core networks, connections are generally bidirectional and for ease of control, it is desirable that both directions are routed along the same links and use the same wavelength in both directions [Ber09].

The two following subsections detail the proposed off-line prioritization and dynamic bidirectional connection provisioning approach.

Off-line prioritization of requests

The off-line optimization tool proposed in [Kos08] returns an ordering of the connection requests, so that instead of the requests being treated based on a memory-less distribution, the same set of requests is reordered so that all requests using the same wavelength in the off-line solution are grouped. Demands that need conversion along the path are prioritized according to the wavelength used. Alternatively, we prioritize the demands according to the hop count measure. Due to this prioritization, the sequential dynamic setup of the connections is influenced to assign the routes and wavelengths of the off-line solution.

Dynamic bidirectional connection setup

Dynamic connection setup is carried out by using the RSVP-TE protocol. A flag within the Path message indicates that the connection request is bidirectional [Far09]. In order to minimize the usage of expensive resources, e.g., WCs, the Label Set has been enhanced with the Suggested Vector [And06], which is an additional object passed during connection setup that can be used to minimize the use of WCs. The concept is illustrated in Figure 48. At each intermediate node, the Label Set is updated to reflect which wavelengths are available in both directions on a given link, while the Suggested Vector reveals the number of necessary wavelength conversions the choice of a given wavelength entails. When the connection request arrives at the destination node, it selects the wavelength requiring fewest WCs within the received Label Set (i.e., λ_2 in the example) and initiates the backward reservation of both connection directions on the chosen wavelength by passing the Resv message.

Simulation study

OPNET Modeler, a commercial discrete event simulation tool, was used in combination with CPLEX-based off-line optimization to simulate connection allocation in a GMPLS-controlled network.

As a test instance, we are using the Pan-European topology, consisting of 28 nodes and 61 links, each equipped with one fibre of 10 wavelengths per direction.

In the dynamic scenario we only use Suggested Vector based dynamic wavelength assignment, i.e., no prioritization scheme is implemented. In the wavelength prioritized scenario, the optimization tool prioritizes requests that require WCs according to the highest wavelength used. In the hop count prioritized scenario, requests are prioritized according to descending hop counts.

The results illustrated in Figure 49 show that the two proposed prioritization schemes decrease the WC usage, and hence avoid WC-bottleneck situations causing connection blocking. At low network loads, prioritizing the requests according to their assigned wavelength returned by the off-line optimization gives the best performance, while prioritizing requests in relation to decreasing hop count is more advantageous for high network loads.

This is likely due to the increased number of conversions needed in the off-line optimized sequence as well.

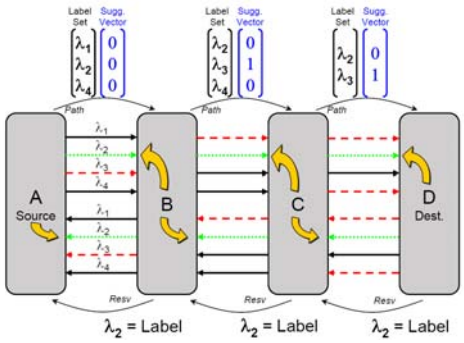


Figure 48: Bidirectional Connection Setup enhanced with Suggested Vector for WC minimization

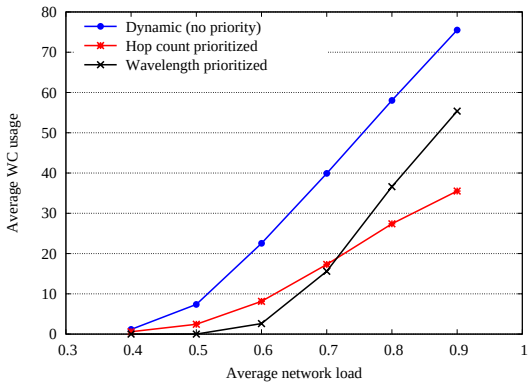


Figure 49: Average wavelength converter usage for dynamic wavelength assignment compared to when requests are prioritized according to wavelength usage or hop count

2.10.2 PCE-based vs. Distributed Set Up of Bidirectional Lightpaths in GMPLS Optical Networks

This task continues the work performed in this JA in the previous year, by comparing two bidirectional lightpath set up schemes in two different network architectures, as detailed in the following.

Network Architectures and Path Computation

Two network architectures have been evaluated, the first being completely distributed, and the second one where the PCE is in charge of path computation. In both architectures, the path for an LSP request between the node pair (*s*, *d*) is selected within a set of candidate paths called $P_{s,d}$. If more than one path in $P_{s,d}$ satisfies the condition detailed below, one of them is randomly selected.

- **Distributed architecture** exploits OSPF-TE to flood aggregated wavelength availability information (i.e., the number of wavelength channels available on each link). In particular, OSPF-TE floods the information by means of link-state advertisements (LSAs). To limit LSA generation, an LSA update timeout is used. Once an LSA has been generated for a given link, all link-state changes detected on the link before the timeout are not immediately advertised, but delayed after the timeout expiration. Path computation is performed at the source node by choosing the

path in $P_{s,d}$ with the largest number of available wavelength channels on its most congested link.

- PCE-based architecture exploits PCE Communication Protocol (PCEP [Vas09]) to provide the PCE with detailed and updated wavelength availability information (i.e., the status of each wavelength channel). Every time an LSP is established (released) the source node communicates to the PCE the reserved (freed) wavelength channels along the used path using a notification message (i.e., PCNtf). By exploiting such a detailed information the PCE selects the path in $P_{s,d}$ that can accommodate the largest number of wavelength-continuous lightpaths.

Signaling schemes

After the path selection, signalling session is triggered. In case of errors further set up attempts are triggered considering the updated wavelength availability information received by the source node in the Acceptable Label Set object included in the PathErr message [Ber03]. The two considered signalling schemes are detailed in the following.

- **Upstream Label (UL)** scheme adheres to [Ber03]. A Path message is sent from source to destination. It contains the Upstream Label object, which reserves a wavelength in the reverse direction, and the Label Set object, which in case of blocking is used to fill the Acceptable Label Set object carried in the PathErr message. If no error occurs, the reverse path is completely set up when the Path message reaches the destination. A Resv message is then sent in the upstream direction along the forward path to reserve the same wavelength used on the reverse path. When the Resv message reaches the source node, the LSP is established.
- **Label Set (LS) scheme** has been proposed in [Ber09] to avoid the drawbacks of the forward reservation performed with the Upstream Label. The request for a bidirectional LSP is indicated with a flag carried by the Path message. The Label Set is updated at each intermediate node by jointly checking the wavelength availability on both directions. When the LSP set up request arrives at the destination, it selects an available wavelength within the received Label Set and starts the backward reservation of both LSP directions on the chosen wavelength.

Simulation studies

Simulations are performed with a custom-built C++ event-driven simulator. A Pan-European network topology is considered with 27 nodes, 55 bidirectional WDM links and 32 wavelengths per link along each direction. Link lengths are equal to the geographical distances between the two end-nodes. Estimated lightpath set up times consider: path computation time (0.5 ms), packet queuing delay and processing time, packet propagation and transmission delays, and OXC switching time (10 ms). Lightpath requests are generated according to a Poisson process and uniformly distributed among all node pairs. Both, inter-arrival and holding times are exponentially distributed. The average of the inter-arrival time is fixed to 1 second. The set $P_{s,d}$ includes all the paths whose hop length is within one hop from the shortest path. The considered LSA update timeout is 30 seconds. Up to five set up attempts are performed before refusing a lightpath request. Random wavelength assignment is used with both signalling schemes. In the PCE-based architecture, the PCE provides the source node with the indication of a wavelength which is included in the Suggested Label object.

Figure 50 shows the blocking probability as a function of load after several set up attempts with distributed and PCE-based architecture. We notice a significant performance

improvement brought by the PCE. In particular, the blocking after one set up attempt exploiting the standard UL scheme with the PCE is lower than the one after the fifth attempt exploiting the enhanced LS scheme in the distributed scenario without the PCE. Nevertheless, Figure 50 (a) shows that the UL scheme in the distributed scenario performs very poorly: at a load of 200 Erlang the blocking after the fifth set up attempt is still greater than 10^{-2}.

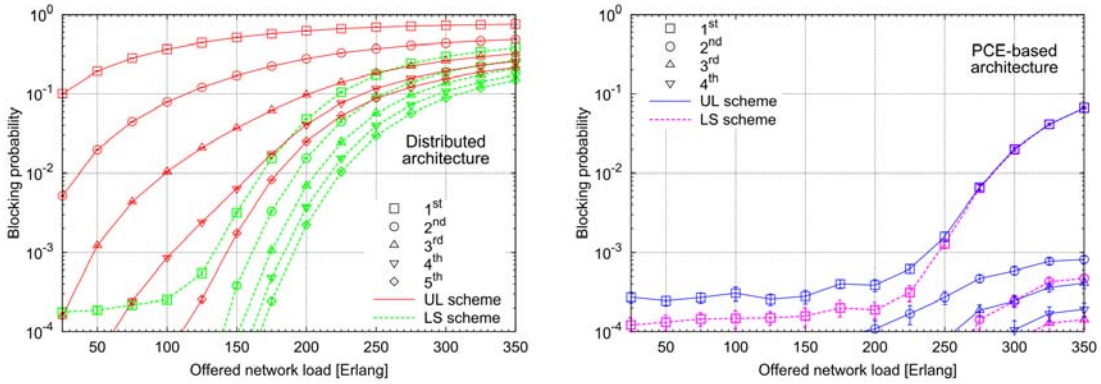


Figure 50: Blocking probability with the distributed architecture (a) and with the PCE-based architecture (b)

Figure 51 (a) details the average lightpath set up time as a function of load, while Figure 51 (b) details the distribution of lightpath set up attempts at a network load of 200 Erlang. Both figures consider the UL and the LS schemes in the two network architectures. In the distributed architecture, both signalling schemes experienced a set up time increasing with the load, i.e., more lightpath requests need further set up attempts to be successfully established. However, LS experiences a significantly lower blocking (see Figure 50 (a)) and thus, since only few lightpath requests exploit more than one set up attempt (see Figure 51 (b)), a lower average set up time is guaranteed with respect to UL at all network loads. On the contrary, with the PCE-based architecture, the two signalling schemes experience the same set up time, almost independent of the load. In this case the set up time is increased because of the PCEP message exchange between the source node and the PCE. However, at high loads the set up time ensured by the PCE-based architecture is the shortest because the majority of connections are established at the first set up attempt (see Figure 51 (b)).

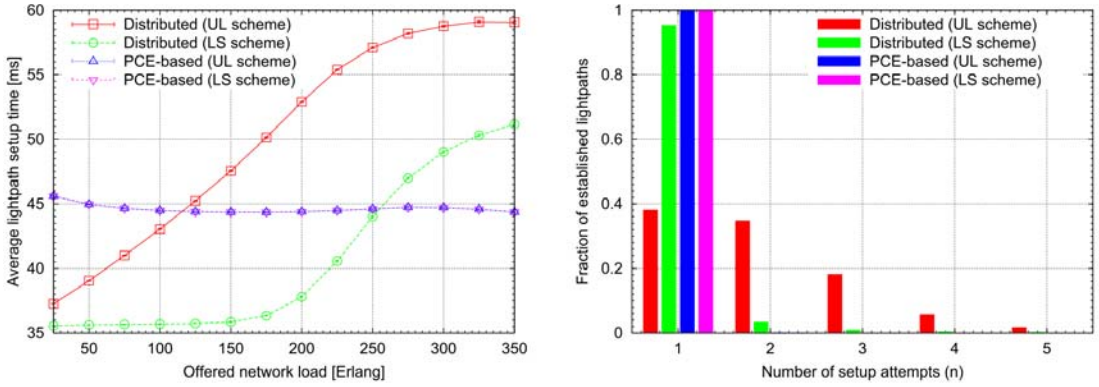


Figure 51: Average signalling time in ms (a), fraction of lightpaths established after n set up attempts (b)

2.10.3 List of publications

- S. Ruepp, A. Koster, N. Andriolli, and A.V. Manolova, "*Prioritizing Connection Requests in GMPLS-Controlled Optical Networks*" in Proc. of Photonics in Switching 2009, Pisa, Sep. 15-19, 2009, paper WeII3-6.

2.11 Monitoring for GMPLS Control Plane in Optical Networks

This JA proposes how to utilize monitoring information in order to provision lightpaths in GMPLS-controlled optical networks within a limited number of set up attempts. The proposed scheme permits a fast lightpath set up procedure.

Transparent dynamic optical networks are affected by physical impairments, which can potentially bring lightpaths' signals below an unacceptable threshold [Tom08]. During last years, QoT admission control mechanisms have been proposed to establish lightpaths while guaranteeing the required QoT. If the evaluated QoT (through QoT estimation or measurements) is acceptable the lightpath is established. Otherwise, alternative paths are exploited, delaying lightpath establishment.

This JA investigates a monitoring-based approach. It is possible to measure the physical characteristics of the links at installation time, but this method is not applicable to dynamic optical networks where the physical layer parameters vary with time. End-to-end monitoring may be used to verify the actual feasibility of a "candidate" lightpath by injecting probe traffic on the lightpath that has been set up before transmitting real data [Sam08]. Then, if the measured QoT meets the transmission requirements, the lightpath is activated. Otherwise, the lightpath is rejected and another setup trial is required, delaying data transmission and wasting temporarily resources that would otherwise have been available for data transmission.

This could be avoided if the knowledge of the lightpath QoT was known a priori: we use an end-to-end estimation framework called "network kriging" [Chu06] to perform QoT estimation by exploiting the knowledge of the network physical layer gained through past probing. We propose an innovative distributed lightpath establishment mechanism that includes network kriging, which reduces the number of successive attempts to successfully establish lightpaths. Simulations show that, for a sample transparent optical network and for medium load values, utilization of network kriging decreases the number of required number of establishment attempts from 3 to 2 to achieve a blocking rate of 10^{-3}.

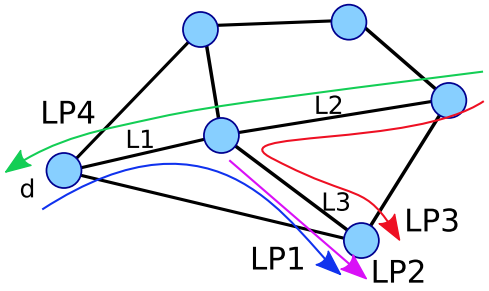


Figure 52: Principle of Network Kriging

2.11.1 Lightpath Provisioning with Network Kriging

We consider the following framework: a transparent optical network is equipped with a distributed, impairment-aware GMPLS control plane that can disseminate QoT parameters to network nodes (i.e., through signalling protocol extensions [Cug08]). At some point in the

operation of the network, some lightpath QoT parameters have been monitored, and hence, QoT information for a number of lightpaths (still established, or already torn down) is available at the network nodes. Assume that a new lightpath request arrives, for which no QoT parameter information exists yet. A QoT parameter for a lightpath is link-additive if there is a linear relationship between the end-to-end parameter and the parameters for each link traversed by the lightpath. Network kriging can estimate link-additive parameters by combining the following information: a) QoT parameters from some other lightpaths already known from past probing; and b) the network topology. Kriging exploits the correlation in terms of physical layer impairments (and hence QoT) between the lightpaths that share the same links; indeed lightpaths that use the same link(s) sustain similar physical degradations. For instance, Figure 52, node d has probing information for lightpaths LP_1, LP_2, and LP_3 then d can determine the contribution in terms of physical impairments of links L_1, L_2, and L_3 and hence compute the physical impairments sustained by LP_4 and hence its QoT. We propose the lightpath provisioning with network kriging scheme (NKS), where we estimate in turn the four following link-additive metrics: OSNR (additive through its inverse, 1/OSNR), PMD (additive through PMD^2 in ps^2), CD (in ps/nm), and SPM (additive through the nonlinear phase shift Φ_{NL}); each of these can be measured at a node using appropriate monitors; in particular, Φ_{NL} can be estimated with power monitors [Ant02]. BER can be computed from these parameters [Cug08]. In the considered network, each node maintains a measurement database (MD) containing the performed end-to-end measurements of OSNR, PMD, CD and Φ_{NL} QoT parameters. The MD may be filled through signalling protocol extensions, hence the MD is distributed and each node has its own view of the network's physical layer parameters.

Upon lightpath request from source *s* to destination *d*, s computes a path *p*. If the MD holds the QoT parameters for *p* from previous probing, then BER is derived. Otherwise, by applying network kriging to the parameters (1/OSNR; PMD^2; CD and Φ_{NL}) contained in the MD at *s*, the parameters related to *p* are estimated and the BER is derived. In both cases, if the derived BER is acceptable, s starts signalling the lightpath request along p, otherwise another path is computed and tentatively established. During the signalling session, link resources (i.e., a common wavelength along *p*) are reserved and optical cross-connects configured. To verify the lightpath QoT is acceptable, probing is performed and QoT measurements are gathered at d, which sends the measured values back towards s. Each node along p fills its proper MD entry with the updated end-to-end measurements. Note that the QoT measurements are not flooded in the network, to keep the scheme scalable. If the measured parameters indicate an unacceptable QoT for *p*, *s* frees resources along *p* and performs another setup attempt. Otherwise, the lightpath is activated and data transmission begins.

2.11.2 Simulation Results

The performance of the proposed scheme is evaluated for a Pan-European topology with 17 nodes, 33 bidirectional links, and 40 wavelengths per direction [Cug08]. To benchmark NKS, we disable the network kriging estimation step and call measurement database based scheme (MDS) this new establishment technique, which uses the information contained in MD without further processing. Lightpath requests are uniformly distributed among all node pairs, following a Poisson process with mean inter-arrival time $1/\lambda$, and holding times are exponentially distributed with a mean $1/\mu$. The offered network load in Erlang is λ/μ. Upon lightpath request, *s* randomly selects a path *p* within a set $P_{s;d}$ of pre-computed paths. $P_{s;d}$ is the set of all paths connecting *s* and *d* that are within one hop from the shortest path and wavelength assignment is first fit (without loss of generality). NKS and MDS are compared in terms of blocking rate after a variable number of setup attempts *n*: blocking occurs if no

wavelength can be found on any path of $P_{s;d}$ or if the monitored QoT parameters (using probing, after establishment) indicate unacceptable lightpath QoT.

Figure 53 shows the blocking rate of NKS and MDS for a fixed load (200 Erlang, low enough such that blocking is due to QoT only) after $n \in \{1,2,3\}$ setup attempts along alternate routes as a function of time, measured after a varying number of lightpath requests. The plotted results are obtained by averaging 100 randomly generated sequences of lightpath requests and blocking rate is computed for a sliding window containing the last 100 requests. In both schemes, as the MD is populated, more information is gathered and the blocking rate decreases with lightpaths demands. But convergence is faster for NKS, which is able to better exploit the information contained in the MD than MDS. For instance, if a single set up attempt is permitted, MDS achieves a 1%-blocking rate after the arrival of 1200 lightpaths, as opposed to only 600 arrivals for NKS.

In Figure 53, we show the lightpath blocking rate for a varying traffic load after $n \in \{1,2,3\}$ setup attempts. Each point is obtained by averaging 100 independent trials of 1500 lightpath requests each. In the range [200,300] Erlang, blocking probability is constant for $n \in \{1,2\}$ because within 1500 requests, the MD is not completely filled in case of MDS, or, in the case of NKS, network kriging does not have enough information to provide confident estimations for every (s; d) pairs. For low and medium loads, if a target blocking rate of 10^{-3} is set, 3 setup attempts for each lightpath arrival are required by MDS while NKS needs only 2.

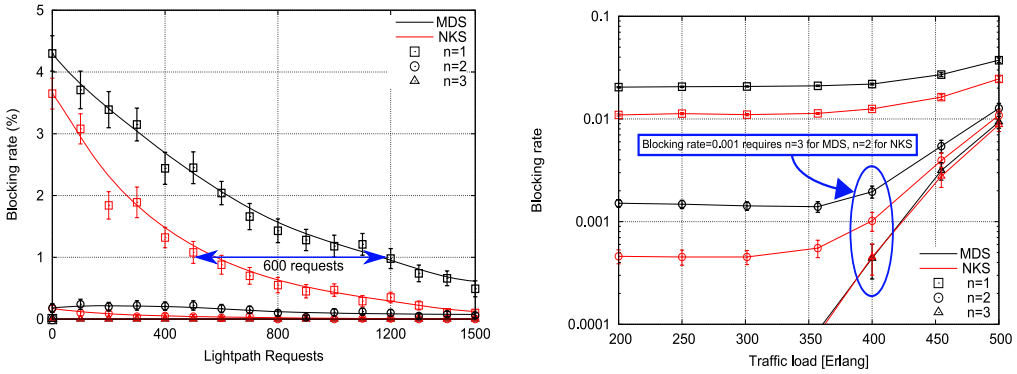


Figure 53: Temporal evolution of the blocking rate after $n \in \{1,2,3\}$ and Blocking rate after $n \in \{1,2,3\}$

2.11.3 Conclusions

In conclusion, we harnessed the "network kriging" estimation framework to estimate lightpaths' QoT before establishment based on prior measurements and using the correlation between lightpaths' QoT induced by the network topology. Simulation results show that with network kriging, fewer attempts are required to successfully establish lightpaths (e.g., 2 set up attempts instead of 3 to obtain 10^{-3} blocking).

2.11.4 List of publications

- N. Sambo, Y. Pointurier, F. Cugini, L. Valcarenghi, P. Castoldi, I. Tomkos, "*Lightpath establishment in distributed transparent dynamic optical networks using network kriging*", in Proc. of ECOC, Vol. Mo, No. 1.5.3, Vienna, Austria, September 2009.

- N. Sambo, Y. Pointurier, F. Cugini, P. Castoldi, I. Tomkos, "*Lightpath establishment in PCE-based dynamic transparent optical networks assisted by end-to-end Quality of Transmission estimation*", in Proc. of 11th International Conference on Transparent Optical Networks, ICTON 2009, Island of São Miguel, Azores, Portugal, June 2009.
- A. Giorgetti, N. Sambo, I. Cerutti, N. Andriolli, P. Castoldi, "*Label Preference Schemes for Lightpath Provisioning and Restoration in Distributed GMPLS Networks*", Journal of lightwave Technology, Vol. 27, No. 6, pp. 688 - 697, DOI 10.1109/JLT.2008.917380, March 2009.
- A. Giorgetti, N. Sambo, I. Cerutti, N. Andriolli, P. Castoldi, "*Suggested Vector Scheme with Crankback Mechanism in GMPLS-controlled Optical Networks*", in Proc. of GTTI meeting, Parma, June 2009.

3. Conclusions

The research activities done within the WP22 Topical Project on “MPLS, GMPLS and routing” represented significant contributions on issues still open related to the evolution of IP-MPLS multi-service networks to all-optical networks. Such contributions have led to several publications in prestigious international journals, magazine and conferences. Specifically, an overall number of 39 papers have been published being 15 of them joint papers, which witnesses the joint research work done in most of the Joint Activities.

In addition, 3 mobility actions were performed contributing in such a way to the mobility of the European researchers and the work on this WP gave the possibility to also have joint experimental activities through the interconnection of the UPC-CTTC testbeds, the UC3M-UPC testbeds and the CTTC-UST-IKR testbeds.

Finally, WP22 significantly helped to perform joint research activities.

4. References

[Agg08] R. Aggarwal, Y. Kamite, and L. Fang, "*Multicast in VPLS*", IETF, Internet draft draft-ietf-l2vpn-vpls-mcast-04.txt, work in progress, Jun. 2008

[And06] N. Andriolli, J. Buron, S. Ruepp, F. Cugini, L. Valcarenghi, and P. Castoldi, "*Label preference schemes in GMPLS controlled networks*", IEEE Communications Letters, vol. 10, no. 12, December 2006.

[Ant02] J. C. Antona, S. Bigo, and J.P. Faure, "Nonlinear cumulated phase as a criterion to assess performance of terrestrial WDM systems," in Proc. OFC, 2002.

[Bat00] P. Batchelor et al. "*Study on the implementation of optical transparent transport networks in the European environment-results of the research project COST 239*", Springer Photonic Network Communications, 2000, vol. 2, no. 1, 15-32.

[Ber03] L. Berger, "*Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description*", IETF RFC 3471, Jan. 2003.

[Ber09] G. Bernstein et al., "*Signaling Extensions for Wavelength Switched Optical Networks*", Jul. 2009, Internet Draft, exp. Jan. 2010.

[Cas07] P. Castoldi, F. Cugini, L. Valcarenghi, N. Sambo, E. Le Rouzic, J. Poirrier, N. Andriolli, F. Paolucci, A. Giorgetti, "*Centralized vs. Distributed Approaches for Encompassing Physical Impairments in Transparent Optical Networks*", ONDM 2007.

[Cas08a] R. Casellas, R. Martínez, R. Muñoz, "*Design, implementation and validation within ADRENALINE® testbed of a Path Computation Element for Wavelength Switched Optical Networks*", in Proc. 4th International Conference on IP over Optical (iPOP2008), Tokyo (Japan), June 2008.

[Cas08b] R. Casellas, R. Muñoz, R. Martínez, "*A Path Computation Element for Shared Path Protection in GMPLS-enabled Wavelength Switched Optical Networks*", in Proc. 34th European Conference on Optical Communications, ECOC2008, Brussels, (Belgium) September 20-25 2008.

[Cas09a] R. Casellas, R. Martínez, R. Muñoz, S. Gunreben, "*Enhanced BRPC for multi-domain PCE-based path computation in Wavelength Switched Optical Networks under Wavelength Continuity Constraint*", Journal of Optical Communications and Networking (JOCN) Vol. 1, No. 2 pp. A180-A193, July 2009.

[Cas09b] R. Casellas, R. Muñoz, R. Martínez, "*Experimental Field-Trial of Multi-domain PCE-based Path Computation for OSNR-aware GMPLS enabled translucent WSON*", in Proc. of 35th European Conference on Optical Communications, ECOC2009, September 2009.

[Chu06] D. B. Chua, E. D. Kolaczyk, and M. Crovella, "*Network kriging*," IEEE Journal on Selected Areas of Communications, vol. 24, no. 12, pp. 2263-2272, Dec. 2006.

[Cug08] F. Cugini, N. Sambo, N. Andriolli, A. Giorgetti, L. Valcarenghi, P. Castoldi, E. Le Rouzic, and J. Poirrier, "*Enhancing GMPLS signaling protocol for encompassing quality of transmission (QoT) in all-optical networks*" J. Lightwave Technol., vol. 26, no. 19, pp. 3318-3328, Oct. 2008.

[D222] Deliverable D22.2, "*Report on Y1 and updated plan for activities*", November 2008.

[Dos99] B. T. Doshi., "*Optical Network Design and Restoration*", Bell Labs Tech. J., 58-84,(1999).

[Dyna] Dynamipswebpage.http://www.ipflow.utc.fr/index.php/Cisco_7200_Simulator

[Dynab] Dynagenwebpage.http://dynagen.org/

[Far09] A. Farrel et al. "*Encoding of Attributes for MPLS LSP Establishment Using Resource Reservation Protocol Traffic Engineering (RSVP-TE)*", RFC 5420, Feb. 2009.

[Fei01] A. Fei, J.H. Cui, M. Gerla, M. Faloutsos. "*Aggregated Multicast: an Approach to Reduce Multicast State*". In Proc. of Sixth Global Internet Symposium (GI2001). San Antonio, Texas, USA, November 25-29, 2001.

[Gro03] W. Grover, "*Mesh-based Survivable Transport Networks: Options and Strategies for Optical, MPLS, SONET and ATM Networking*", Prentice Hall, 2003.

[Gun09] S. Gunreben, R. Casellas, R. Martínez, R. Muñoz, J. Scharf, "*Experimental Validation and Assessment of Multi-domain and Multi-layer Path Computation*", submitted for publication to the 6th conference on Testbeds and Research Infrastructures for the Development of Networks and Communities (Tridentcom 2010) 18-20 May 2010 - Berlin, Germany.

[Guo07] H. Guo, T. Tsuritani, Y. Yin, T. Otani, J. Wu, "*Proposal of a multi-layer network architecture for OBS/GMPLS network interworking*", in Proceedings of SPIE, vol. 6784, 67842I, 2007.

[Han97] S. Han and K. G. Shin, "*Efficient spare resource allocation for fast restoration of real-time channels from network component failures*", Proceeding of IEEE Symposium on Real-Time Systems, IEEE 1997, pp. 99-108.

[ITU8080] ITU-T G.8080, "Architecture for the automatically switched optical network (ASON)," (2001).

[Jaj06] A. Jajszczyk, P. Rozyki, "*Recovery of the control plane after failures in ASON/GMPLS networks*", IEEE Network, vol. 20, no. 1, Jan. 2006.

[Ji08] Y. Ji, H. Wang, and L. Guo, "*MSWS method to support shared-mesh restoration for wavelength switched optical networks*," Internet Draft, July 2008, IETF web site, Intended Status: Informational.

[Kom08] O. Komolafe, J. Sventek, "*Impact of GMPLS Control Message Loss*", IEEE/OSA Journal of Lightwave Technology, vol. 26, no. 14, Jul. 2008.

[Kor04] S.K. Korotky, "*Network global expectation model: a statistical formalism for quickly quantifying network needs and costs*", IEEE/OSA Journal of Lightwave Technology., vol. 22, no. 3, March 2004.

[Kos08] A. Koster and S. Ruepp, "*Benchmarking RWA Strategies for Dynamically Controlled Optical Networks*", in Proc. of Networks 2008.

[Li02] G. Li, J. Yates, D. Wang, C. Kalmanek, "*Control plane design for reliable optical networks*", IEEE Comm. Mag., vol. 40, no. 2, Feb. 2002.

[Li05] J. Li, K. L. Yeung , "*A Novel Two-Step Approach to Restorable Dynamic QoS Routing*", Journal of Lightwave Technology, vol. 23, no. 11, 3663-3670 (2005).

[Lon06] K. Long, X. Yang, S. Huang, Y. Kuang, "*A GMPLS-based OBS Architecture for IP-over-WDM Networks*", Network Architectures, Management, and Applications IV, Proceedings of SPIE, vol. 6354, 63540H, 2006.

[Man04] E. Mannie, "*Generalized Multi-Protocol Label Switching Architecture*", IETF RFC 3945, Oct. 2004.

[Man07] A. Manolova et al., "*Advantages and Challenges of the GMPLS/OBS Integration*", in Proc. of VI Workshop GMPLS, Girona, Spain, April 2007.

[Mar06] R. Martinez, R. Martinez, C. Pinart, F. Cugini, N. Andriolli, L. Valcarenghi, P. Castoldi, L. Wosinska, J. Comellas, and G. Junyent, "*Challenges and requirements for introducing impairment-awareness into the management and control planes of ASON/GMPLS WDM networks*" IEEE Commun. Mag., vol. 44, no. 12, pp. 76–85, Dec. 2006.

[Mar07] G. Markidis, S. Sygletos, A.Tzanakaki and I.Tomkos, "*Impairment aware based routing and wavelength assignment in transparent long haul optical networks*", LNCS Optical Network Design and Modeling, Springer 2007, Vol. 4534, pp. 48-57.

[Mar07b] I. Martínez-Yelmo, D. Larrabeiti, and I. Soto, "*Multicast Traffic Aggregation in MPLS-Based VPN Networks*", IEEE Communications Magazine, vol. 45, pp. 78-85, 2007.

[Mar08] G. Markidis and A. Tzanakaki, "*Routing and wavelength assignment algorithms in survivable WDM networks under physical layer constraints*", Proceedings of the 3rd International GOPS Workshop, Broadnets (IEEE 2008).

[Mun08] R. Muñoz, R. Casellas, R. Martínez, "*An Experimental Signalling Enhancement to Efficiently Encompass WCC and Backup Sharing in GMPLS-enabled Wavelength-Routed Networks*", in Proc. of IEEE International Conference on Communications (ICC 2008), 19-23 may 2008, Beijing (China).

[Per07] J. Perelló, S. Spadaro, J. Comellas, and G. Junyent, "*An Analytical Study of Control Plane Failures Impact on GMPLS Ring Optical Networks*", IEEE Commun. Lett., vol. 11, no. 8, Aug. 2007.

[Pin07] C. Pinart, Le Rouzic, I. Martinez, "*Physical-layer considerations for the realistic deployment of impairment-aware connection provisioning*", 9th International Conference on Transparent Optical Networks (ICTON). Rome (Italy), 2007.

[Pin08] C. Pinart, "*Relation of OSNR and PLR for nonintrusive fault detection in all-optical IP-Ethernet-WDM networks*", OSA Journal of Optical Networking, Vol. 7, Issue 4, pp. 256-265, 2008.

[Ram98] B. Ramamurthy and B. Mukherjee, "*Wavelength conversion in WDM networking*", IEEE J. Sel. Areas Commun., vol. 16, no. 7, pp. 1061–1073, 1998.

[Ram99] B. Ramamurthy et al., "*Impact of transmission impairments on the teletraffic performance of wavelength-routed optical networks*", IEEE/OSA Journal of Lightwave Technology, vol. 17, no. 10, pp. 1713–1723, 1999.

[Ram99b] S. Ramamurthy and B. Mukherjee, "*Survivable WDM Mesh Networks, Part I - Protection*", in Proceeding of IEEE Conference on Computer and Communication Societies (IEEE, 1999), pp. 744-751.

[Rek06] Y. Rekhter, T. Li, S. Hares, "*A Border Gateway Protocol 4 (BGP-4)*", RFC 4271, January, 2006.

[Ros06] E. Rosen and Y. Rekhter, "*BGP/MPLS IP Virtual Private Networks (VPNs)*", IETF, RFC http://www.ierf.org/rfc/rfc4363.txt, Feb. 2006.

[Ros07] E. Rosen and R. Aggarwal, "*Multicast in MPLS/BGP IP VPNs*", IETF, Internet draft draft-ietf-l3vpn-2547bis-mcast-08.txt, work in progress, Apr. 2007.

[Ros08] R. Roshani et al. "*Reconfigurable Optical Networks: Is it Worth?*" OTuA2, in Proc. OFC 2008.

[Rue08] S. Ruepp et al. "*Restoration in All-Optical GMPLS Networks with Limited Wavelength Conversion*", Computer Networks, vol. 52, no. 10, pp. 1951–1964, July 2008.

[Sal07] E. Salvadori, Y. Ye, A. Zanardi, H. Woesner, M. Carcagni, G. Galimberti, G. Martinelli, A. Tanzi, D. La Fauci, "*Signaling-based architectures for impairment-aware lightpath set-up in GMPLS networks*" in Proc. IEEE Global Telecommunications Conference, GLOBECOM 2007, Dec. 2007.

[Sam08] N. Sambo, F. Cugini, I. Cerutti, L. Valcarenghi, P. Castoldi, J. Poirrier, E. Le Rouzic, C. Pinart, "*Probe-based schemes to guarantee lightpath Quality of Transmission (QoT) in transparent optical networks*", in Proc. 34th European Conference and Exhibition on Optical Communication (ECOC 2008). Brussels (Belgium), September 21-25, 2008.

[Sam08b] N. Sambo et al., "*Impact of routing and wavelength selection strategies on GMPLS-controlled distributed restoration*", JON, Vol. 7, No. 5, 2008 .

[San06] J. Sanchez, P. Manzanares, and J. Malgosa, "*Spanish Telco Strategies Facing New Integrated Digital Transmission Advances*", Global Communications Newsletter. IEEE Communications Magazine, vol. 44, 2006.

[Sta08] D. Staessens, D. Colle, U. Lievens, M. Pickavet, P. Demeester, W. Colitti, A. Nowe, K. Steenhaut, and R. Romeral, "*Enabling High Availability over Multiple Optical Networks*", IEEE Comm. Mag. Vol. 46, Issue 6, pp. 120-126, June 2008 .

[Tom08] I. Tomkos, S. Azodolmolky, M. Angelou, D. Klonidis,Y. Ye, C. V. Saradhi, E. Salvadori, A. Zanardi, R. Piesiewicz, "*Impairment Aware Networking and Relevant Resiliency Issues in All-Optical Networks*", in Proc. 34th European Conference and Exhibition on Optical Communication (ECOC 2008). Brussels (Belgium), September 21-25, 2008.

[Vas09] J. P. Vasseur and J. L. Le Roix, "*Path computation element (PCE) communication protocol (PCEP)*", IETF RFC 5440, Mar. 2009.

[Vel08] L. Velasco et al., "*On the design of MPLS-ASON/GMPLS Interconnection Mechanisms*", VII Workshop in G/MPLS Networks. Vilanova i la Geltrú, Cataluña, Spain, 11-12 Marzo 2008.