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/SSSUP/R/PU/D12.2
Contractual Date of Deliverable:	30/11/09
Actual Date of Delivery:	30/11/09
Title of Deliverable:	D12.2: Report on inventory of available expertise, planned strategy towards integration, and activated joint activities
Workpackage contributing to the Deliverable:	WP12: Virtual Center of Excellence on Services and Applications (VCE S&A)
Nature of the Deliverable	Make your choice from: R (Report)
Dissemination level of Deliverable	Make your choice from: PU (Public)

Editors: Partner / Names
SSSUP / Piero Castoldi, Luca Valcarenghi, Isabella Cerutti

Abstract:

This document is the report of the second year of activities in the VCE-WP12 Virtual Center of Excellence on Services and Applications (VCE S&A). It includes the results of the restructuring finalized during the year that have included a few joint activities from former WP21 Topical Project on Service Oriented Optical Networks. This reports the status of all joint activities currently operating within VCE-WP12.
There are twelve partners involved in this workpackage and seven joint activities are currently running. During the second year, the collaboration between partners increased and helped to set-up joint testbeds for experimental evaluation of service-oriented platforms for optical networks. Joint research work indicated that the problem of improving and guaranteeing service availability is of paramount importance and strategies for addressing such problem were proposed. Dissemination of the joint activities in international conferences and journals and in the Open Grid Forum standardization body has been actively pursued.

Keyword list:

Application-network integration, resource virtualization, service engineering.

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 deliverable of the work package "Virtual Center of Excellence on Services and Applications (VCE S&A)", reporting the activity carried out from 1/1/2009 to 30/11/2009.

During Y2, the ongoing activity in WP21 Topical Project on "Service Oriented Optical Networks" were moved to WP12, as they address similar aspects of application and services in optical networks: (i) Service definition, architectures and implementation, (ii) Application definition, architectures, requirements, (iii) Services and application in an integrated view.

Based on common research interests and complementary expertise of the partners, in Y1, four (4) joint activities (JA) were proposed in WP12 and three (3) JAs were proposed in WP21. Therefore, this document is reporting the work, the progress and the planned actions of the seven (7) JAs carried out in Y2.

The joint collaborative work carried out in Y2 focused on the following areas:

- *Definition of a standard Network Service Interface* (NSI). A draft of recommendations for a NSI to support advanced network connectivity requests from heterogeneous network users and providers has been proposed to the Open Grid Forum (OGF). The novelty of the proposed NSI is the possibility to support both the Grid approach which treats networks as resources and the network approach which treats networks as collaborating entities that provide a service to requestors. These approaches have been unified in the NSI recommendation, allowing network resource infrastructure to be deployed and configured in a variety of ways.
- *Ensuring availability of services*. The main objective is to ensure the required level of availability to services with stringent requirements. In particular, two types of services are considered:
    - grid services, namely network-wide distributed storage and computing services. Grid services avoid the need of huge private supercomputers and data storage centers, thanks to the virtualization and sharing of the resources, i.e., network resources and computing resources. For improving the service availability in terms of reliability, novel approaches based on multicost optimization of routing and relocation of grid services are proposed and studied. Different resilience strategies for OBS networks are compared. Also, an approach that ensure the service availability in terms of bit error ratio (BER) is proposed.
    - service interconnections in integrated wired and wireless networks. To ensure a seamless connectivity to mobile stations of wireless networks, availability and reliability should be ensured taking into account the availability of both the wireless and the wired networks.
- *Strategies for service abstraction and resource virtualization.* Service-oriented optical networks allows the provisioning of services (such as grid services or cloud computing services) without disclosing the technology-related details of the network resources to applications. To offer such features, the network needs to be enabled with service

abstraction and resource virtualization capabilities. Algorithms and strategies for achieving service abstraction and resource provisioning are investigated and their impact on the quality of service requested by the applications and on network performance has been evaluated.

- *Planes for service-oriented optical networks*. To support the service abstraction and the resource virtualization required by grid services and cloud computing, the design and implementations of the service plane, along with the configuration of the service-oriented functionalities have been carried out in optical networks based on
    - (generalized) multi-protocol label switching (G)MPLS control plane. Services can be provisioned by a module implemented in the Path-Computation Element (PCE) or by a service plane on top of the control plane. The PCE module for resource and admission control is implemented and tested for grid interconnection services. Also, an implementation of service plane based on session initiated protocol (SIP) signaling is realized and shown to provision the requested services.
    - optical burst switching (OBS) networks. The functionalities and the signaling for a service-oriented OBS network are defined and implemented in a OBS network with multi-granular OXC.
- *Experimental evaluation of service-oriented networks*. Three testbeds have been realized. A large collaborative effort was necessary to organize and configure the testbeds. However, once set-up the testbeds could be operated remotely, thus avoiding the need of personnel visits and mobility actions. Testbed consisted of networks with either (G)MPLS commercial equipment or innovative OBS technology and configured to support services and applications. The testbeds were set up to experimentally validate the concepts and evaluate the performance of real scenarios.

Finally, the report documents the dissemination of the results and achievements in international conferences and journal publications and the mobility actions that took place.

1.1 Meetings

The VCE organized a plenary meeting, jointly with WP12 WP21 WP22 WP24 WP26. The meeting was held in Bologna in June 2009, shifted of about 6 months with respect to the project plenary meeting. The meeting lasted 2 days (June 8 and 9). The meeting was organized as a mini workshop, with about 60 participants and mostly devoted to technical presentations of results achieved by the JAs.

Within the various JAs, during Y2 several meetings or conference calls were set-up to discuss the workplan and the advance of results.

1.2 Summary of Joint Activities

The current Joint Activities (JAs) defined within the WP12 are the indicated in the table below. Please note that JA5, JA6, and JA7 were inherited from WP21 on Service Oriented Optical Networks.

No	Joint Activity Title	Responsible person	Participants	Mobility Action in 2009	Time frame
1	Service Interconnection Fault-Tolerance	Luca Valcarenghi (SSSUP)	SSSUP, AGH, IBBT, KTH	YES	Extended For 3rd year
2	Cloud Computing	Dimitra Simeonidou (UESSEX)	UESSEX, RACTI, SSSUP	NO	Extended For 3rd year
3	Service Plane functionalities and demonstration	Fabio Baroncelli (SSSUP), Barbara Martini (SSSUP)	SSSUP, UESSEX, UNIBO, FUB	YES	Extended For 3rd year
4	Joint Optimisation of Grid and Network Resources	Tibor Cinkler (BME)	BME, SSSUP, PoliTO	NO	April 2008-Dec. 2010
5	Programmable Service Composition Algorithms for Service Oriented Optical Networks	Chinwe Abosi (UEssex)	UESSEX, UoP, RACTI	YES	Extended For 3rd year
6	UNI extensions for Service Oriented Optical Networks	Eduard Escalona (UEssex)	UESSEX, RACTI, AIT	NO	Extended For 3rd year
7	Photonic Grid Dimensioning & Resilience	Chris Develder (IBBT)	IBBT, AIT, RACTI	NO	Extended For 3rd year

1.3 Mobility Actions

The following mobility actions took place during the second year of project within WP12:

1. **Service interconnection availability in Mobile WiMAX**

Luca Valcarenghi, Assistant Professor at SSSUP,

hosted by KTH from 07/02/2009 to 15/02/2009

2. **SIP signalling for user-oriented optical networks**

Walter Cerroni, Researcher at UNIBO,

hosted by SSSUP from 14/05/2009 to 14/05/2009

3. **Economic viability of resource infrastructure abstractions in Grid networks** (2 mobility actions)

Chinwe Abosi, Research Student at UEssex,

hosted by GET from 01/06/2009 to 25/06/2009

Miss Rosy AOUN, PhD student at GET

Hosted by UEssex from 4/4/2009 to 5/5/2009

4. **SIP-based Service Platform for Transport Networks** (2 mobility actions)

Molka Gharbaoui, PhD student at SSSUP,

Barbara Martini, Research Scientist at SSSUP

hosted by UNIBO from 25/11/2009 to 25/11/2009

1.4 Dissemination

During the second year, the JAs of WP12 generated the publications listed in the table below.

JA	Authors' Affiliation	Title	Journal/ Conf.	Published in
JA1	AGH, SSSUP	Experimental Evaluation of PCE-Based Batch Provisioning of Grid Service Interconnections	Conf.	Proc. Globecom '09
JA1	SSSUP, KTH	Issues and Solutions in Mobile WiMAX and Wired Backhaul Network Integration	Conf.	Proc. ICTON '09
JA2	RACTI	A Service-Transparent and Self-Organized Optical Network Architecture	Conf.	Proc. ICC '09
JA3	SSSUP, UNIBO	SIP-based Service Platform for On-demand Optical Network Services	Conf.	Proc. OFC '09
JA3	SSSUP, UEssex	Service-Oriented Multi-Granular Optical Network Testbed	Conf.	Proc. OFC '09
JA3	SSSUP, UEssex, UNIBO	Data-Plane Architectures for Multi-Granular OBS Network	Conf.	Proc. OFC '09
JA3	SSSUP, UNIBO	SIP-based Service Architecture for Application-aware Optical Network	Conf.	Proc. ICTA '09
JA5	GET	Service differentiation based on flexible time	Conf.	Proc.

		constraints in market-oriented grids		Globecom '09
JA5	GET	An exact optimization tool for market-oriented grid middleware	Conf.	Proc. CQR workshop '09
JA5	GET	Towards a fairer benefit distribution in grid environments	Conf.	Proc. GRIDCOM '09
JA5	SSSUP	A cooperative approach for the automatic configuration of MPLS-based VPNs	Journal	Int. J. of Grid Comp. & Multi Agent Systems
JA7	AIT, KTH	Impairment Constraint Based Routing (ICBR) with Service Differentiation in Survivable WDM Networks	Conf.	Proc. ECOC '09
JA7	IBBT	On dimensioning optical grids and the impact of scheduling	Journal	Photonic Netw. Commun. (17), '09
JA7	IBBT, RACTI	Multi-cost job routing and scheduling in grid networks	Journal	Future Gen. Computer System. (25):8, '09

1.5 Standardization

A standardization effort has been carried out by JA6 partners within the in the Open Grid Forum (OGF). A unofficial draft version of Network Service Interface (NSI) user-case deliverable (work in progress) has been submitted.

1.6 Joint Testbeds

Within the various JAs, the joint testbeds indicated in the table below were set-up.

JA	Involved Partners	Testbed Name	Testbed Type	Joint Collaboration
JA1	AGH, SSSUP	PCE-based batch provisioning of grid service interconnections	Network testbed with commercial MPLS routers	Testbed set-up at SSSUP, remotely operated from AGH
JA3	SSSUP, UEssex	Service-oriented multi-granular optical network	OBS network (at UEssex) combined with service-plane software (implemented by SSSUP)	Testbed set-up and operated at UEssex during a mobility action
JA3	SSSUP, UniBO	SIP-based service platform for GMPLS-enabled transport network	MPLS network infrastructure enhanced with service plane (at SSSUP) integrated with enhanced SIP-based signaling (developed by UniBO)	Network set-up at SSSUP and SIP software located at UniBO → tested remotely run from UniBO

2. JAs Description

WP12 merges the activities on services and applications with the activities that started in WP21 in 2008 on service-oriented optical networks. Thus, the main focus of WP12 is on application and services for optical networks.

This section describes the technical progress achieved by the joint effort of the partners in Y2, within each JAs. Planned activities and publications for 2010 are indicated for each JA.

2.1 JA1 - Service Interconnection Fault-Tolerance

Participants: SSSUP, AGH, IBBT, KTH

Responsible person: Luca Valcarenghi (SSSUP)

2.1.1 Objectives for Y2

The activities carried out during Y2 focused on two main objectives that originated two sub-activities: the experimental evaluation of path computation element (PCE)-based batch provisioning of grid service interconnections and the evaluation of the reliability of service interconnections in integrated wired and wireless networks.

2.1.1.1 PCE-based Batch Provisioning of Grid Service Interconnections

The first sub-activity is the follow up of the activity started during Y1 and conducted in collaboration between SSSUP and AGH within JA1. During Y2, the activity focused on the experimental evaluation of the performance of the proposed PCE-based system for service interconnection provisioning in wired networks.

The proposed system is referred to as Resource and Admission Control Function for Grid Services or (GRACF) [1] and its interactions with the elements involved in service interconnection provisioning (e.g., Label Switched Paths --- LSPs) are depicted in Figure 1. The *Client* represents a service running on a workstation connected to the grid data and control plane network. The service utilizes a *GRACF-client interface*, implemented in Java, to submit connection (e.g., LSP) set-up and tear down requests to the GRACF. Communication with the GRACF is carried out using an HTTP-based protocol (i.e., Hessian) [2]. The *Transport Network* consists of *Network Elements (NEs)* (e.g., Label Switched Routers, LSRs) utilized to build the grid data and control plane network and configured by the GRACF. The GRACF queues the requests from the clients and communicates with the PCE for connection route computation. Based on a implemented policy, it accumulates a certain amount of requests and then communicates with NEs for connection activation. A cached network topology is maintained along with active and pending LSP data. The GRACF is a J2EE application based on Spring Framework [3] and runs within a dedicated container — in this case Apache Tomcat 6.0. Within the GRACF, the following functional elements can be distinguished. The *Request Broker* receives the LSP (de)activation requests from the client and notifies the clients about the successful/unsuccessful set up/tear down of a connection. It also updates a list of pending LSPs in the Topology Cache (i.e., the LSPs for which requests have been received but that have not been established yet), according to the received requests. Moreover, it queues the requests in the Request Queue after positive admission control. The

Topology Cache contains a cached view of the network topology with full information on available network resources and active and pending LSPs. The initial topology information is obtained from the NEs by requesting a snapshot of current Traffic Engineering Database (TED) through the Network Controller. The topology cache is maintained by either adding or subtracting the bandwidth utilized by LSPs from the link available bandwidth stored in the Topology Cache, once LSPs are either removed or established. The update is triggered by the Network Controller once it is notified that an LSP has been either set up or torn down. The update of the active and pending LSP lists is synchronized with the Topology Cache update. The Topology Cache may also include policy parameters utilized by the Request Broker for admission control (e.g., additional link information unavailable in the TED). The *Request Queue* is used to store admitted LSP set up/tear down requests. The requests are represented by means of Java objects. The *Path Computation Scheduler (PCS)* is responsible for serving the queued requests based on the implemented service policy and the interaction with the PCE.

The policy proposed in this study, for simplicity called *batch provisioning policy*, is a bulk-service policy that serves connection requests in batches of fixed size *b*. Once the number of requests in the Request Queue reaches the size *b*, the PCS, by acting as a Path Computation Client (PCC), sends a Path Computation Request (PCReq) message to the PCE. The PCReq message contains all the *b* requests in the batch. After having computed the routes for the *b* requests, by solving an Integer Linear Programming (ILP) formulation of the LSP routing problem, the PCE returns them to the PCS through PCRep messages. The PCS groups the LSPs by the ingress LSR (NE in which the LSPs must be activated and configuration changes must be applied). Then, it forwards the grouped requests to the Network Controller. Optionally, the PCS might skip the PCE-based routing and delegate route computation to the NEs, instead of using PCE. In this case, the LSRs utilize Constraint Shortest Path First (CSPF) algorithm for distributed route computation. The *Network Controller* interacts with the NEs through the management network to both forward LSP (de)activation commands and obtain network topology and LSP status information. As stated previously, the Network Controller is provided by the PCS with separate groups of LSP set-up requests. Each group contains LSPs that originate from the same physical NE. Note that this may in fact include multiple logical NEs in case router virtualization features are used within the network. At this point, the LSPs are converted into necessary NE configuration changes, expressed in terms of XML-based commands. All commands that originate from a single group of requests are sent to the physical NEs in a single session and, once accepted, they are committed together. Each physical NE is configured separately which allows to perform this process in parallel for all physical NEs. Upon successful LSP set up the Network Controller moves the set up LSPs from the *Pending LSP List* to the *Active LSP List* and it updates the Topology Cache. The Network Controller interacts with the NE by means of a proprietary XML-based communication protocol.

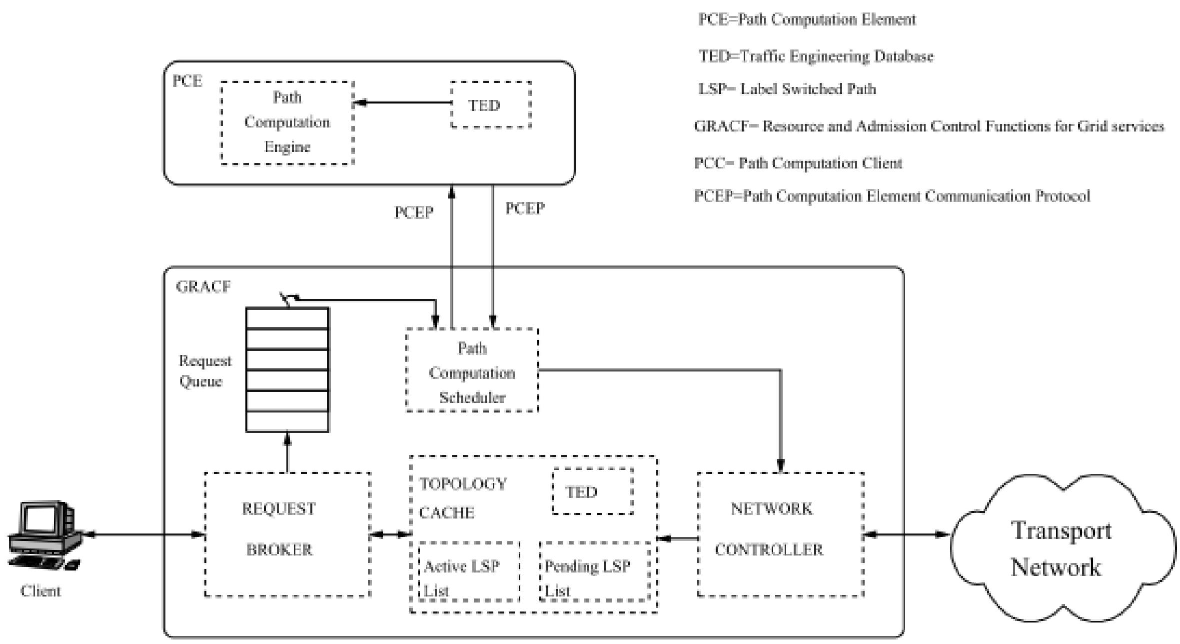


Figure 1: System architecture

2.1.1.2 *Reliability of Service Interconnections in Integrated Wired and Wireless networks*

The second sub-activity focused on the issue of providing reliable service interconnections in integrated wired and wireless networks. In particular the integration of a wireless WiMAX (802.16) network and a fixed network utilizing Mobile IPv6 (MIPv6) to support user mobility is considered. The functional architecture of MIPv6 to support MS mobility is displayed in Figure 2.

To ensure IP-connectivity, the Mobile Station (MS) is assigned a Care of Address (CoA) by the Access Service Network (ASN) Gateway (ASN-GW) responsible of the Base Station (BS) to which the MS is currently connected. The CoA is registered in the Home Agent’s (HA) Binding Cache (BC) that maps the MS Home Address (HoA) (i.e., the MS address in its Home-CSN) to the current CoA. In this way, a Correspondent Node (CN) located in another network can communicate with the MS by sending IP packets addressed to the MS HoA. Packets are received by the HA that, in turn, forwards them to the MS CoA, through IP tunnelling. Similarly, the MS can communicate with the CN by tunnelling packets through the HA that then forwards them to the CN (i.e., reverse tunnelling). Other optimized route communications can be implemented where the MS communicates directly with the CN and the CN communicates directly with the MS (i.e., Route Optimization).

The MS may roam between different BSs but also between different ASNs. When the MS roams between two BSs that are connected to the same ASN Gateway (ASN-GW) (i.e., when they belong to the same sub-network to which the ASN-GW belongs), MIPv6 leverage the Layer-2 mobility functions provided by the ASN to support user mobility. In this case the MS does not need to change its IP Care of Address (CoA) and, thus, Layer 3 operations are not utilized. On the other hand, when the MS roams between BSs belonging to different subnetworks (i.e., connected to a different ASN-GW), a new CoA is assigned and a binding cache update is triggered. The binding cache update allows to change the entry CoA-HoA in the HA binding cache [5]. In the case of a fast moving MS, the binding update procedure is triggered frequently. To reduce the high signaling overhead, a solution, called Hierarchical Mobile IPv6 (HMIPv6) [6], has been proposed by IETF. With the HMIPv6 protocol, a mobility anchor point (MAP) is introduced to act as a proxy of the HA. The MAP is used to handle, in a localized manner, binding update (BU) exchange procedures when the handoffs are within the same MAP domain. According to the HMIPv6 protocol, the MS configures two CoA: a regional Care of Address (RCoA), i.e., the MS address in the MAP subnet, and a Link Care of Address (LCoA) i.e., a local CoA for the MS interface. The MAP acts as a local HA. It receives all packets on behalf of the MS and it re-tunnels them to the MS current address. If the MS changes its current address within a local MAP domain, it only needs to register the new LCoA with the MAP. The RCoA does not change as long as the MS moves within the same MAP domain. This makes the MS mobility transparent to the CNs and the HA and it reduces both latency and signaling overhead. The aforementioned procedures are typical of MIPv6 and HMIPv6, but can be mapped to the WiMAX ASN-anchored mobility and CSN-anchored mobility.

When looking at the end-to-end all-IP WiMAX network, an important aspect to consider is the availability. The choice of one MAP over another may impact differently end-to-end connections. One MAP selection approach, that achieves an efficient reduction of the signaling overhead, may not be the best choice in terms of availability (e.g., the farther the MAP from the MS, the worse the availability of the connection). In addition, the communication between the CN and the MS may require some degree of diversity (e.g., node, SRLG), thus requiring the MS to be connected to more that one MAP, one for each disjoint end-to-end path. In turn, this impacts the way MAP should be placed in the optical backhaul network. However the MAP selection and the MAP placement schemes proposed so far do not consider network availability as part their objectives. Current MAP selection schemes can be efficient in minimizing handover latency but lack to optimize the availability.

In this activity, the impact of current MAP selection schemes on the end-to-end connection availability between mobile client (MS) and fixed server is evaluated. The considered MAP selection schemes are three distance-based selection schemes. In the first scheme, the *farthest MAP selection scheme*, the MS selects the farthest MAP in terms of hops among the ones whose advertisements were received by the MS. The *Farthest MAP selection (FMS)* scheme has been already proposed in the literature [7] and it is utilized by fast moving MS to minimize the number of handovers. In the second scheme, the *Closest MAP selection (CMS)*, the advertised MAP closest to the MS is chosen. CMS can potentially improve the availability of the connection between MS and MAP but it implies a higher number of handovers. In the third scheme, the *Half-Way MAP selection (HMS)* scheme, the advertised MAP at a distance closest to the average of all the distances advertised by the MAP to the MS is chosen by the MS. The HMS scheme represents a compromise between availability and number of handovers.

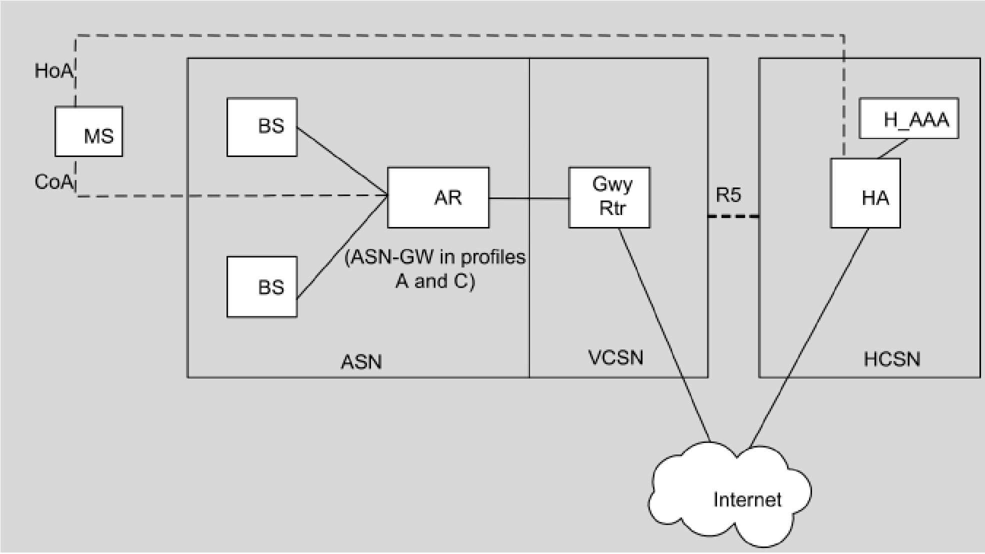


Figure 2: Functional architecture of MIPv6 [4]

2.1.2 Achieved Results in Y2

2.1.2.1 PCE-based Batch Provisioning of Grid Service Interconnections

Experimental evaluation of the PCE-based batch provisioning of grid service interconnections has been performed by means of a remote collaboration by SSSUP personnel and AGH personnel. SSSUP personnel contributed to set up the testbed and implement the ILP-based PCE LSP computation. AGH personnel contributed by implementing the GRACF and integrating it with the PCE code provided by SSSUP. In addition the AGH personnel run the experiments remotely on the testbed and collected the results.

The testbed consists of N = 9 NEs and L = 12 bidirectional links. Commercial LSRs equipped with GMPLS control plane are used as NEs. The network topology is obtained by utilizing the logical router feature of the commercial LSRs. This feature allows to divide physical routers into several logical ones that emulate independent routers. The interfaces of a physical router are assigned to specific logical ones. However, the management interface still belongs to the physical router and it is the one that receives all (i.e., related to both physical and logical router) configuration requests. Thus, the configuration changes are sent to the same physical interface even if they are referring to different logical routers. In the considered network, 9 logical routers are implemented in 3 physical routers. The boundaries of physical routers are depicted by dashed lines. The LSRs are connected through 100 Mb/s links with the exception of 3 links: (r4, r7), (r5, r8) and (r6, r9) which are 1 Gb/s links in both directions.

Each experiment consists of generating a set of B LSP set-up requests between service access routers (ingress and egress LSRs) following a Poisson process with an average arrival rate of λ = 25 requests per second. For each request, ingress and egress LSRs are selected randomly based on a uniform distribution from the set of all routers {r1, r2, r3, r4, r5, r6, r7, r8, r9}. In a single experiment, all of the generated requests are identical in terms of demanded QoS parameters and require a bandwidth of W Mb/s. Set up LSPs are never torn down, thus traffic is incremental.

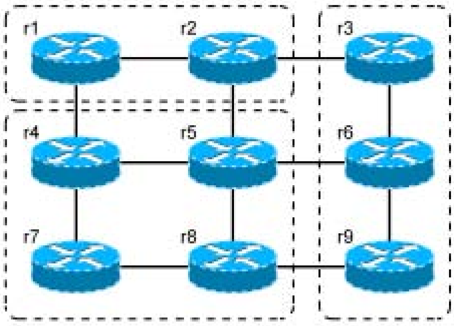


Figure 3: Experimental network testbed

The proposed batch provisioning is compared against CSPF-based provisioning. In CSPF-based provisioning the PCE is not utilized for LSP route computation. Connection requests are still temporarily stored in the request queue. Then, they are forwarded to the ingress routers that compute and set up the requested LSPs in a distributed manner.

Figure 4 summarizes the experimental results when $B = 150$ LSP requests are generated. The bandwidth requested by each LSP is $W = 10$ Mb/s. The batch size b varies from 25 requests per batch to 150 requests per batch. The performance are evaluated in terms of *average number of accepted LSPs (B′)*, *average service time* (τ), and *average number of fully congested links* L' for PCE-based batch provisioning and CSPF-based provisioning. B' is defined as the number of set up LSPs averaged over all the performed experiments for a fixed value of batch size b. τ is defined as the time elapsing between the LSP set up request reception by the GRACF and the client notification of the successful LSP set up, averaged over all the accepted requests and all the experiments for a fixed value of batch size b. L' is defined as the number of unidirectional links whose capacity is exhausted in each experiment averaged over all the experiments for a fixed value of b.

From Figure 4, it can be observed that PCE-based batch provisioning achieves a better performance in terms of B' than CSPF-based provisioning. In particular, the higher b the larger is the improvement in terms of B' that is achieved by PCE-based batch provisioning with respect to CSPF-based provisioning. At the limit, for $b = B$, the improvement is of about 6%. In addition, if PCE-based batch provisioning is utilized, L' is lower than if CSPF-based provisioning is utilized. Figure 4 shows that the utilization of the batch queue is also slightly beneficial if CSPF-based provisioning is utilized. This is due to the fact that, in the considered testbed, many equal cost shortest paths are present between LSR pairs. Thus, LSRs are capable of better balancing their routing by computing the routes of many LSPs between the same *(s,d)* pair at once. Finally, the PCE-based batch provisioning and the CSPF-based provisioning achieve similar results in terms of τ for any considered value of b. In particular, unexpectedly, the larger is b the smaller is τ. This is because the utilized commercial LSRs are irresponsive to any external command while configuration changes are activated. Therefore, if queued LSP requests are served before the set up of LSPs belonging to a previous LSP request batch terminates, they must wait to be set up. Thus, their τ increases. On the other hand, if the batch size b is such that LSP requests are served after the termination of the set up procedure of LSPs belonging to a previous LSP request batch, they can be immediately set up. Thus, they experience a lower average service time.

b	B'	L'	τ [s]
150	127.83	12.5	41.98
75	119.5	13.16	49.22
50	118.83	13.5	58.8
25	118.33	14.83	154.33

b	B'	L'	τ [s]
150	120.33	13.83	43.88
75	118.33	14.83	48.37
50	117	13.66	60.64
25	115.66	14.33	152.53

Figure 4: PCE-based batch provisioning with B=150 (left hand side) and CSPF-based provisioning with B=150 (right hand side)

2.1.2.2 Reliability of Service Interconnections in Integrated Wired and Wireless networks

In the second sub-activity related to the issue of providing reliable service interconnections in integrated wired and wireless networks, the three MAP selection schemes have been evaluated analytically in two sample fixed networks: ring topology and tree topology. The results have been obtained during the mobility action SSSUP → KTH. The two topologies are depicted in Figure 5 and Figure 6.

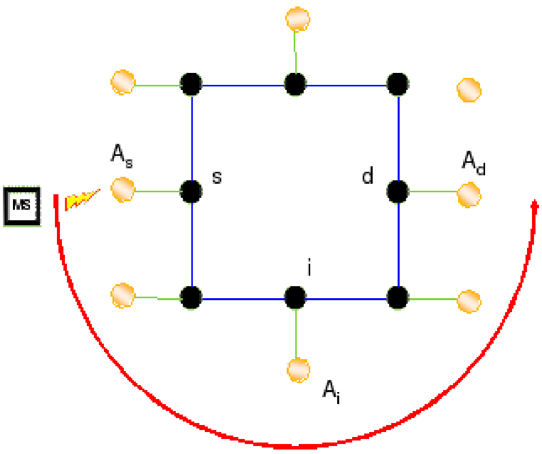


Figure 5: Ring topology

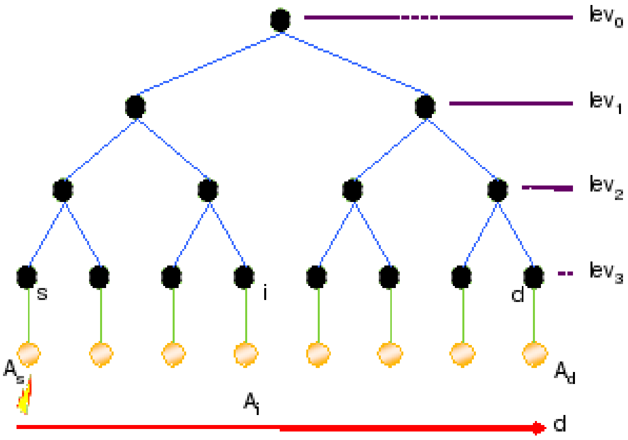


Figure 6: Tree topology

The expected probability that a packet transmitted from node *s* (MS) to node *d* (BS) belonging to the *i-th* access network A_i is correctly received p_a. The expected probability that any link of the backhaul wired network is working (i.e., link reliability) is p_{ml}. The MS moves between access networks following the patterns depicted in Figure 5 and Figure 6.

In the ring network, the average availability of the connection between the MS and the MAP for the FMS, HMS, and CMS can be computed based on the following equations:

$$\overline{A}_v^{FMS} = \frac{\sum_{j=0}^{h_{sd}} p_a \left[1-\left(1-p_{ml}^{(h_{sd}-j)}\right)\left(1-p_{ml}^{(h_{\max}-h_{sd}-j)}\right)\right]}{h_{sd}+1} \quad (1)$$

$$\overline{A}_v^{HMS} = \frac{\sum_{j=0}^{h_{sd}} p_a \left[1 - \left(1 - p_{ml}^{\left(\frac{h_{sd}}{2} - j\right)}\right)\left(1 - p_{ml}^{\left(h_{\max} - \frac{h_{sd}}{2} - j\right)}\right)\right]}{h_{sd} + 1} \quad (2)$$

$$\overline{A}_v^{CMS} = p_a \, , \quad (3)$$

where h_{sd} is the number of hops between s (i.e., the MS) and d (MAP) and h_{max} is the maximum number of links in the ring. The average number of handovers is 1 for the FMS, 2 for the HMS, and 20 for the CMS.

In the tree network the average availability of the connection between a node s (i.e., the MS) and the node d (i.e., MAP) is

$$\overline{A}_v^{lev_k} = p_a \, p_{ml}^{(lev_{\max} - lev_k)} \, , \quad (4)$$

where lev_k is the tree level at which the MAP is placed and lev_{max} is the number of levels in the tree.

The average number of handovers is

$$tot_{handover} = 2^{lev_k} \, . \quad (5)$$

Figure 7 shows the average availability of the ring network normalized to the average availability to the one achieved by the FMS,t as a function of the probability that a link of the backhaul wired network is working, p_{ml}. The results show that HMS availability is 4% better than the one obtained by FMS while requiring the double of handovers. The average availability achieved by CMS is 12% better than the one achieved by FMS but the number of handovers required is 20 times larger. Therefore, HMS is a good compromise between the obtained availability and the number of required handovers.

The results obtained for the tree network are depicted in Figure 8. They confirm that placing the MAP at a higher level of tree improves the linear availability but it needs an exponentially increasing number of handovers.

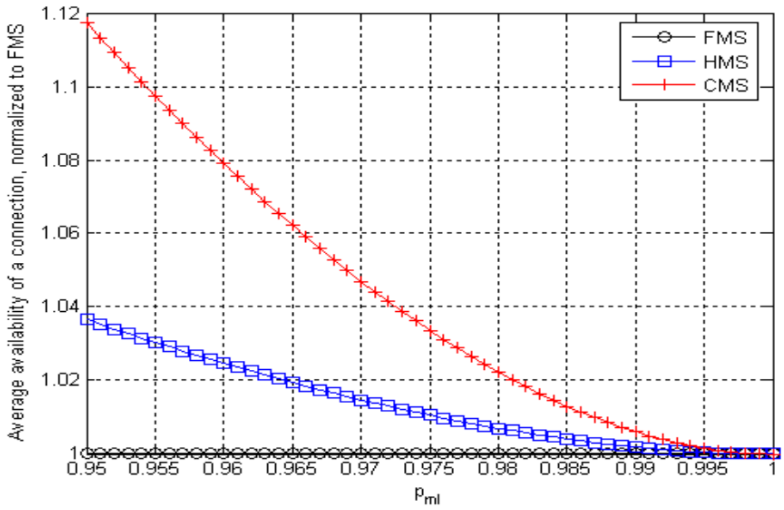


Figure 7: Average availability of a connection between MS and MAP normalized to the one achieved by CMS

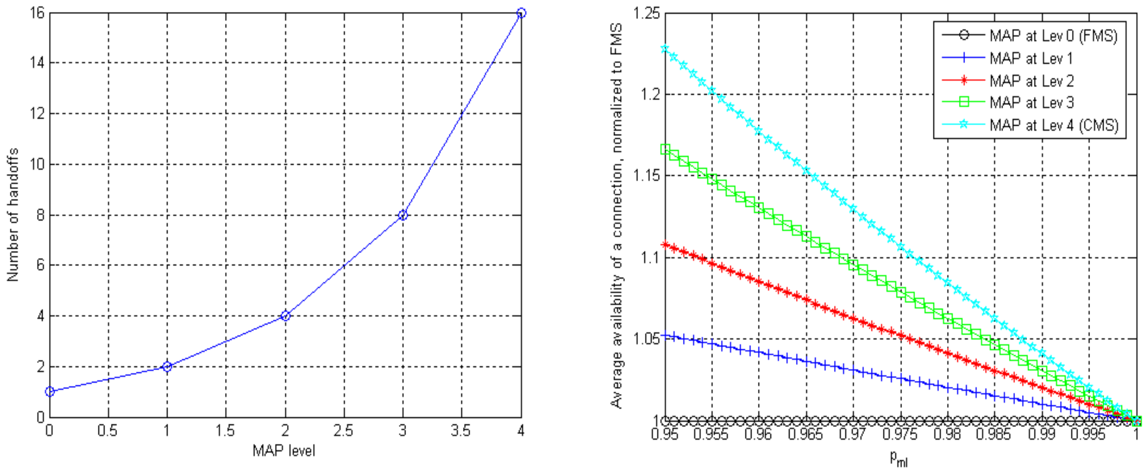


Figure 8: Tree network results

2.1.2.3 References

[1] K. Knightson, N. Morita, and T. Towle, "NGN Architecture: Generic Principles, Functional Architecture, and Implementation," IEEE Communications Magazine, 2005.

[2] "Homepage of hessian," http://hessian.caucho.com

[3] "Spring framework," http://www.springsource.com

[4] WiMAX Forum End-to-End Network Systems Architecture Stage 2–3, Release 1.0.0, March 2007 (http://www.wimaxforum.org/resources/documents/technical/release1) [accessed on May 1, 2009]

[5] R. J. Koodli and C. E. Perkins, "Mobile Inter-Networking with IPv6: Concepts, Principles, and Practice", John Wiley & Sons, 2007

[6] H. Soliman, C. Castelluccia, K.E. Malki, L. Bellier, "Hierarchical Mobile IPv6 Mobility Management (HMIPv6)", IETF RFC 4140, August 2005

[7] S. Pack, M. Nam, T. Kwon, Y. Choi, "An Adaptive Mobility Anchor Point Selection Scheme in Hierarchical Mobile IPv6 Networks", Computer Communications, vol. 29, n. 16, 2006, pp 3066 – 3078.

2.1.3 Published Joint Papers (joint and single partner)

L. Valcarenghi, P. Korus, F. Paolucci, F. Cugini, M. Kantor, K. Wajda, P. Castoldi, "Experimental Evaluation of PCE-Based Batch Provisioning of Grid Service Interconnections", to appear in Proceedings of Globecom 2009, Nov. 30-Dec. 4, 2009, Honolulu, HI, USA

L. Valcarenghi, P. Monti, I. Cerutti, P. Castoldi, L. Wosinska, "Issues and Solutions in Mobile WiMAX and Wired Backhaul Network Integration", ICTON 2009, Jun. 28-Jul. 2, 2009, Island of S˜ao Miguel, Azores, Portugal [INVITED PAPER]

2.1.4 Planned Activities for Y3

Within the activity related to the batch provisioning of grid service interconnections, an analytical model and a simulative performance evaluation are planned.

Within the activities related to the reliable service interconnections in integrated wired and wireless networks, further investigation that takes also into account also the availability of the connection between MAP and corresponding node, is planned.

2.1.4.1 Targeted publications

Computer Communications, Globecom 2010, ICC 2010.

2.2 JA 2 - Cloud Computing

Participants: UESSEX, RACTI, SSSUP

Responsible person: Dimitra Simeonodou (UEssex)

2.2.1 Objectives for Y2

Cloud computing enables a transparent access to Information Technology (IT) services such that, the users do not need to know the location and characteristics of the relevant resources [1]. Data transport within the cloud and its efficient control have not received much attention in the technical literature. In fact, connectivity is, itself, a service that contributes to the overall performance of the cloud.

- Clouds provide utility computing as a commercial service, that can offered by commercial infrastructure/service providers (Google, Amazon, eBay, Microsoft, IBM. Cloud computing providers require high-bandwidth optical networks in order to co-locate computing resources in remote locations (i.e. secure sites, renewable energy sites)
- Define optimised end-to-end service offering scenarios (routing algorithms, NML, virtualisation)

Connection-oriented transport technologies, such as the (Generalized) Multi Protocol Label Switching ((G)MPLS), allow the dynamical setup of broadband connections thanks to the control plane functionality [2].

To facilitate the interaction between the control plane and the service request, two approaches were proposed in Y1:

- A *service layer* has been introduced for each service offered in the network. The provision of each service to end-user means that there exist service specific resources in the network that can be computing elements, storage, VoD servers etc. Each resource can be replicated more than one time in the network either for survivability issues or for serving more or different customers.
- a *Network Virtualization Platform* (NVP) has been proposed to support virtualized connectivity to cloud users at different levels of reliability, traffic Quality of Service (QoS) and transparency. Following cloud terminology, this class of services is named as *Network as a Service* (NaaS).

In Y2, the proposed layer and platform were further defined and finalized in the following sub-activities:

- *design and integration of service virtual plane into GMPLS*. This sub-activity aims at proposing how to combine the different metrics (network and non-network resources) and how to implement the proposed scheme using GMPLS extensions. With respect the question on how to combine metrics of different nature for path computation purposes, there exists numerous research approaches. However, there has not been yet proposed a method either to convert the multi-cost to single cost routing problem as in physical layer impairment aware routing algorithms using analytically calculated Q-factors or truly handle the problem as a multi-cost routing problem.

- *design and validation of the Network Virtualization Platform*. This sub-activity aims at providing design requirements and directives to integrate NaaS into Cloud Computing infrastructure such that it can be orchestrated with other Cloud Services, i.e., Software as a Service (SaaS), Infrastructure as a Service (IaaS).

2.2.2 Achieved Results in Y2

The Y2 progress of the two sub-activities are as follows.

2.2.2.1 Design and Integration of Service Virtual Plane into GMPLS

The architecture for cloud computing proposed in Y1 is illustrated in Figure 9.

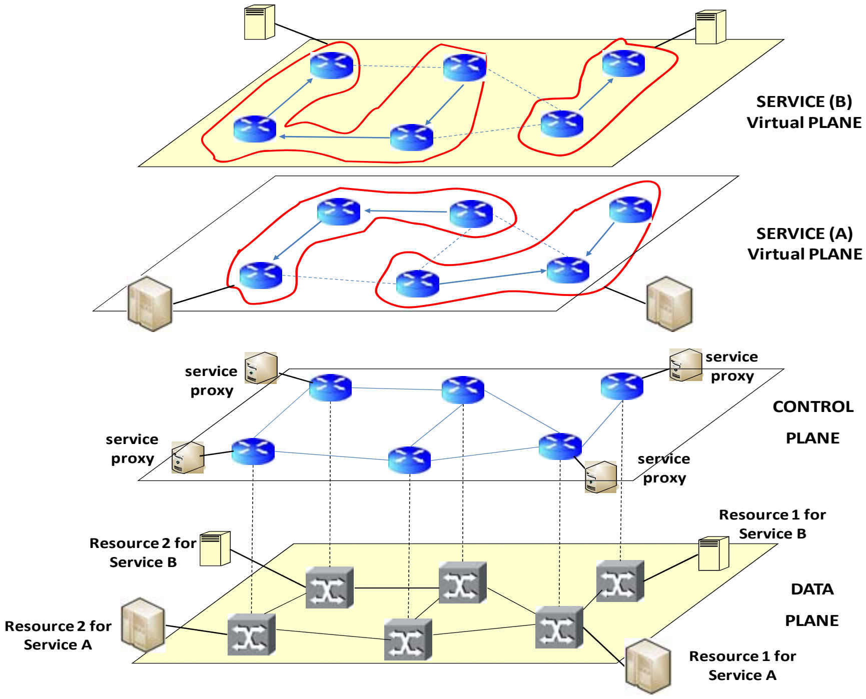


Figure 9: Cloud computing paradigm employing service transparent optical islands

Each service plane is fragmented into self-organized and self-managed entities hereinafter called *service transparent islands*. For each service, there exists a virtual service plane (replica of the control plane), where nodes are self-organized per resource in such a way that each service request within that network domain is transparently routed to that resource for execution. A *service proxy* is responsible for *service addressing* and is placed at the entry points of the network. Self –organization of nodes is carried out using Local Network State Information that is being exchange between nodes in an upstream fashion starting from resources towards end-users.

We propose the use of classical optimization algorithms to address the problem of combining together the different metrics (network and non-network resources). Among them, meta-heuristics (evolutionary algorithms, ant colony, etc) are the only one that can truly handle a multi-objective optimization problem but however are too complex for path selection especially for a small set of candidate paths. This is because each core node usually has a small number of neighbouring nodes. On the other hand, traditional methods of converting the multi-objective to single-objective optimization path selection are more suitable. Such methods include the *weighted sum* and the *ε-Constraint* approaches. In the first case, a single-objective model is created by weighting *n* objective functions. For example, a candidate score function from node R_i to S_r^k resource could be:

$F_{s_r^k} = a_1 D_{R_i \to s_r^k} + a_2 h^{-1}_{R_i \to s_r^k} + a_3 D^{-1}_{R_i \to s_r^k}, \ldots, a_k CPU_{s_r^k}$ subject to $\sum_{t=1}^{k} a_t = 1$ and $a_t \in [0,1], t = \{1 \ldots k\}$

Optimization function F_s can be modified to represent a usable absolute number by using for example the % deviation over the minimum-maximum range value as follows:

$$F_{s_r^k} = a_1 \frac{B_{R_i} - \mathrm{B_{min}}}{\mathrm{B_{max}} - \mathrm{B_{min}}} + a_2 \left(\frac{h_{R_i \to s_r^k} - h_{\mathrm{min}}}{h_{\mathrm{max}} - h_{\mathrm{min}}} \right)^{-1} + a_3 \left(\frac{D_{R_i \to s_r^k} - D_{\mathrm{min}}}{D_{\mathrm{max}} - D_{\mathrm{min}}} \right)^{-1}, \ldots, + a_k \frac{CPU_{s_r^k} - \mathrm{CPU_{min}}}{CPU_{\mathrm{max}} - CPU_{\mathrm{min}}}$$

where the min, max refer to the corresponding min, max values among the set of all candidate islands in the vicinity of R_i.

The proposed design can be implemented with currently proposed GMPLS extensions. We will make use of the concept of *forwarding adjacency* proposed by K. Kompella et al.[3]. The concept of forwarding adjacency was proposed to aggregate label switched paths (LSPs) by creating a hierarchy of such LSPs. Routing protocols like open-shortest-path-first (OSPF) floods information for forward adjacent (FA) links as for normal TE links and thus each Label Switching Router (LSR) may keep FA LSPs in its link state database along with its TE information. A *forward adjacent* LSP is treated as normal TE links by LSRs. To make use of FA notion in the proposed Cloud computing architecture, FA-LSPs must be setup dynamically between more than two GMPLS nodes. In that case, each FA-LSP is initiated from the first node (egress router), adjacent to a resource. The newborn FA-LSPs (single link) can be created statically whereas we may assume that there exists an external mechanism (software) to pass performance metrics (i.e CPU availability etc) of that resource. These non-network metrics, together with networking ones must be communicated upstream to extend the service specific FA-LSPs. This is done by network layer protocols like OSPF-TE that keep and flood link state information. To this end and according to [3], the attributes of a FA-LSP are inherited from the LSP that induced its creation. Thus, TE information may propagate with the extension of the FA-LSP. During that process, LSR treats FA-LSPs similar to other TE links and computes which forwarding adjacency to join in a way similar to TE links for path computation. To this end, each LSR node may autonomously decide to join (or not) a FA-LSPs, thus extending a "*service island*". It must be noted here that in our case, an LSR has only FA-LSPs in its database on a per-service basis and this set of FA-LSPs corresponds to the number of resources of that specific service in the network.

2.2.2.2 *Design and Validation of the Network Virtualization Platform (NVP)*

A NVP is able to translate a NaaS request consisting in an abstract connection request in terms of end-point IP addresses and high-level QoS parameters (i.e., desired bandwidth and maximum tolerated delay and jitter) into a set of CP directives to the network. The NVP implements NaaS by composing and coordinating the IP-based or sub-IP-based connectivity services provided by the actual network resources. NaaS may be invoked on demand or may be combined with IaaS, PaaS, and SaaS to provide richer or advanced cloud services, i.e., coupled with QoS-enabled connectivity services.

From an architectural point of view, the NVP is logically placed alongside (i.e. at the same level of abstraction of) the Cloud Virtualization Platform, and it accepts on demand NaaS requests from the user via a dedicated interface as shown in Figure 10. The NaaS is a novel class of services for Cloud Computing that provides virtualized connectivity to end users at different levels of reliability, traffic Quality of Service (QoS) and transparency in a flexible and scalable way. Flexibility represents the possibility for the end user to dynamically change QoS parameters, like the bandwidth or the delay time, and to monitor the bill, the number of access and their durations, according to its Service Level Agreement (SLA). Scalability refers to the opportunity for the user to add or remove network virtual nodes easily without affecting its traffic.

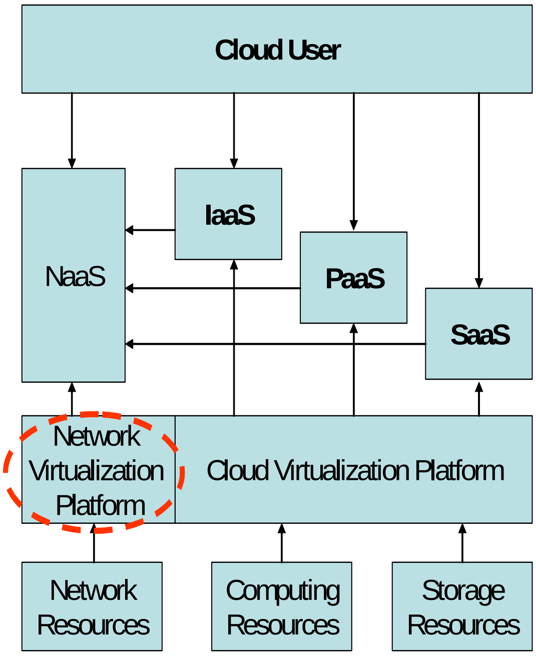


Figure 10: Cloud architecture extension

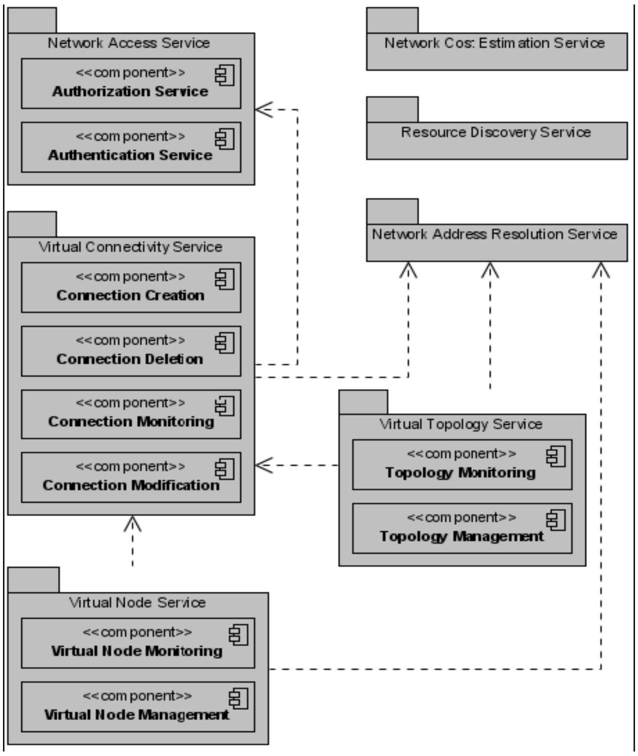


Figure 11: Network as a Service classification

The parameters that characterize a NaaS can be classified in two categories: the Perceived QoS Parameters and the User Address Parameters. The Perceived QoS Parameters refers to the connectivity attributes that an end-user can perceive and monitor, for example, using software or hardware probes (e.g., the bandwidth, the packet delay, the packet jitter). The User Address Parameters are relevant to the address space domain of the end-user (e.g., the IP addresses of the user's hosts). NaaS can be classified in the following service sets: Network

Access Service, Virtual Connectivity Service, Virtual Topology Service, Virtual Node Service, and Network Cost Estimation Service as sketched in Figure 11.

The Network Access Service consists of the Authentication Service and the Authorization Service. The former concerns the user identification following a service request in order to grant the access to the network service. The latter is related to service acknowledgement and provisioning according to user service level agreement.

The Virtual Connectivity Service offers the monitoring and management of the virtual connections established among users. It consists of the Connection Creation service that allows the creation of connections between a pair of user addresses with specific attributes, the Connection Deletion service that allows the deletion of an existing connection, the Connection Monitoring service that monitors the status of certain connection parameters, and the Connection Modification service that allows the modification of the parameters of an already established connection.

The Virtual Topology Service consists of the Topology Monitoring and the Topology Management for monitoring and managing virtual connectivity information, respectively (such as the available bandwidth, the packets delay and jitter, and the restoration time). The Virtual Node Service consists of the Virtual Node Monitoring and the Virtual Node Management for monitoring and managing information about virtual nodes, respectively. The information provided is strictly related to the type of virtual connectivity established (e.g., VPN, VPLS). The Network Cost Estimation Service allows end users to enhance the performance of their applications through the use of information on the status and the transmission behaviour of the controlled virtual links.

The previously presented NaaS needs network-specific information to perform the mapping between a NaaS request and the set of CP-based directives needed to configure the network for the provisioning of that service. In particular, the parameters that characterize the request of a connectivity service provided by a NVP can be classified in three categories: the Connectivity QoS Parameters, the Transport Network Address Parameters, the Network Technology Parameters.

The Connectivity QoS Parameters are QoS parameters that characterize the connectivity established in the transport network such as the bandwidth, and the class of service used by a specific network technology (e.g., DiffServ). The Transport Network Address Parameters concern the edge nodes of the transport network involved in the provisioning of a service, e.g. the Network Service Address Point (NSAP) of a transport network optical device. The Network Technology Parameters are the technology-specific parameters needed by the NVP for setting up the network devices.

To perform the mapping from these parameters and the Perceived QoS and User Address parameters, two basic services are conceived: the Resource Discovery Service and the Network Address Resolution Service. Since these services deal with private network information, their use is only conceived to compose other NaaSs and Cloud Services in general.

The Resource Discovery Service searches for resources that match the requirements of the user application. The resources may be both network-specific or IT-specific. The former are relevant to the technology-specific information about the network resource controlled by the NVP (e.g., the available bandwidth, the available ports and channels) that should be utilized by the Virtual Connectivity Service before the establishment of a new virtual connection. The latter are relevant to the information about the connectivity, reachability, and availability of the IT resources controlled by the Cloud such as the list of Storage servers connected to a specific sub-network or the IP addresses associated to a given cluster. The Network Address Resolution Service bounds a user address to the relevant address of the network access node.

The NVP for cloud computing has been deployed into the Service Oriented Network Architecture (SOON) model whose testbed has been widely validated experimentally [4]-[6]. In [4], it has been demonstrated that the taken approach is scalable since the network size does not significant degrade the performance of the system, especially the service provisioning latency.

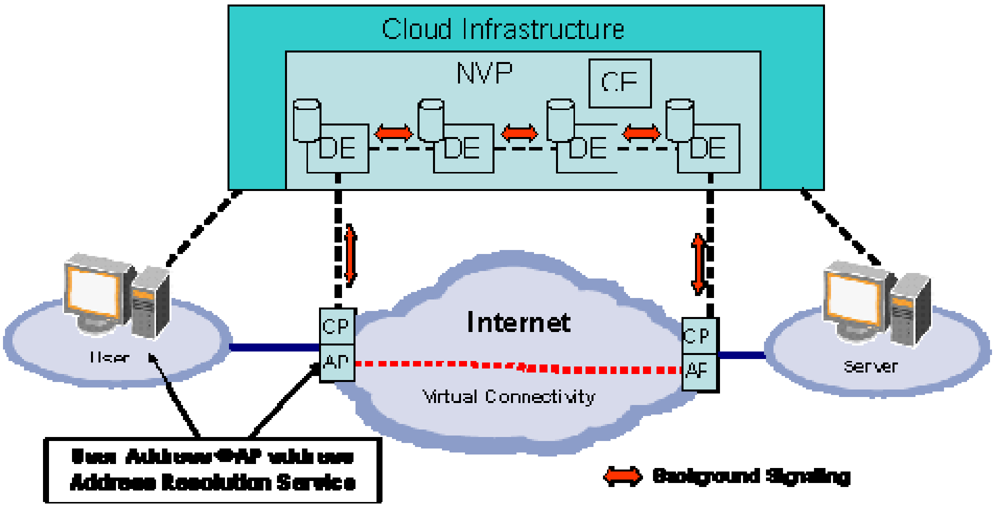


Figure 12: NVP architecture

As shown in Figure 12, the NVP consists of Distributed Entities (DEs) and a Centralized Entity (CE). In particular, the CE implements a database with the customer profiles and relevant Service Level Agreement for the end user authentication and relevant network service authorization. Since network

Following the architecture proposed in [4], NVP is based on a decentralised resource discovery mechanism, named *Background Signalling*, for the automatic collection of the network topology information. In particular, the *Background Signalling* consists in a message exchange at regular intervals between the DEs that periodically collect from the APs the network information on resource status and boundary topology . Each DE floods to other DEs and locally stores the received information in a local Network Resource Database. Once summarized, such information support the network resource virtualization logic and thus the mapping between NaaS requests and relevant set of CP directives.

In addition two additional signalling, independent of the CP signalling of the controlled network, named *Service Provisioning Signalling*, and *Inter-Domain Signalling* are conceived among the NVP entities [6,7].

The *Service Provisioning Signalling* is responsible for the actual NaaS provisioning by performing the consistent settings on the network nodes. In fact, when the NVP entity receives a NaaS request, it computes the path among the network sub-domains. Then, it decomposes the requested connectivity (e.g., the creation of a VPN) into a set of basic connectivity requests (e.g., the creation of Label Switched Paths) to be issued to specific network edges through the relevant NVP DEs. The addresses of the edge nodes and the relevant NVP DEs are discovered using the network boundary information previously collected through the *Background Signalling*. In turn, each NVP DE translates the QoS technology-independent network parameters of the connectivity request (e.g., the bandwidth) into a set of CP-specific parameters (e.g., the number of virtual circuits in the case of SDH networks) that, in turn, are used to issue a set of directives to the controlled CP element at the boundary of the network for the actual resource reservation devoted to the user-generated data stream. These directives may include buffer queuing and scheduling, packet filtering, marking and policing, traffic shaping, firewall. Upon receiving connectivity directives, each CP element triggers a signalling session within the network CP for establishing the part of service under its competence.

The *Inter-Domain Signalling* occurs among NVP DEs deployed in different domains in order to provide edge-to edge connectivity across a multi-domain network. In particular, this signalling is able to discover the service capabilities in each domain.

2.2.2.3 References

[1] Foster, I.; Yong Zhao; Raicu, I.; Lu, S., "Cloud Computing and Grid Computing 360-Degree Compared," Grid Computing Environments Workshop, proc. GCE conf., Nov. 2008

[2] E. Mannie Ed. "Generalized Multi-Protocol Label Switching (GMPLS) Architecture", RFC 3945, October 2004, http://www.ietf.org/rfc/rfc3945.txt?number=3945

[3] K. Kompella, Y. Rekhter, “Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE)” RFC 4206, Oct. 2005

[4] B. Martini, V. Martini, F. Baroncelli, K. Torkman, P. Castoldi, “Application-driven Control of Resources in MultiService Optical Networks”, IEEE/OSA Journal of Optical Communications and Networking, vol. 1, n. 2, July 2009, pag. A270-A283

[5] F. Baroncelli, B. Martini, V. Martini, P. Castoldi, “A cooperative approach for the automatic configuration of MPLS-based VPNs”, International Journal of Grid Computing and Multi Agent Systems (GCMAS), to appear

[6] F. Baroncelli, B. Martini, V. Martini, and P. Castoldi, “A distributed signalling for the provisioning of on-demand VPN services in transport networks,” Integrated Network Management (IM) 2007, Germany, Munich , May 2007

[7] Damandeep Kaur, Jyotsna Sengupta “Resource Discovery in Web-services based Grids” World Academy of Science, Engineering and Technology, vol. 31, July 2007, issn: 2070-3724

2.2.3 Published Joint Papers (joint and single partner)

Kyriakos Vlachos and Apostolos Siokis, “A Service-Transparent and Self-Organized Optical Network Architecture”, in proceed. of ICC 2009, Page(s):1 – 6, 14-18 June 2009.

2.2.4 Planned Activities for Y3

During Y3, RACTI plans to experimentally implement the proposed architecture with service virtual layers using a set of linux boxes and OXCs. The testbed will consist of optical cross connects (OXC), controllable by dedicated, Spartan II FPGAs. Linux PCs will be employed to serve as control plane nodes and which will communicate via a dedicated spi interface with the FPGAs for parsing commands. At the network edges, service proxies (linux PCs as well) will parse user service requests. In addition, GMPLS extension will be also proposed.

Concerning the sub-activity on NVP, future works will investigate concrete solutions for integrating NaaS within cloud services using commercial tools. As starting point, definition of use cases will provide a valid approach to effectively study possible migration and integration strategies, either in case of NaaS directly invoked by users or in case of a NaaS used to enhance already existing Cloud services

2.2.4.1 Targeted publications

Terena Conference 2010, journal publication into the Annals of Telecommunications published by Springer

2.3 JA 3: Service Plane Functionalities and Demonstration

Participants: SSSUP, UEssex, UNIBO

Responsible persons: B. Martini (SSSUP) and F. Baroncelli (SSSUP)

2.3.1 Objectives for Y2

The activities carried out during Y2 have the main objective to experimentally evaluate the performance of service platforms for optical networks that were designed during Y1. Such platforms are required to expose abstract levels of information regarding network resources to IT applications, e.g., Cloud computing or Multimedia Service Platform, in order to provide enhanced and advanced IT services, i.e., coupled with QoS-enabled connectivity services.

Consider an heterogeneous network based on optical burst switching (OBS) and (G)MPLS as core network technologies. In such scenario, we employed the Service Oriented Optical Network (SOON) architecture. SOON implements service provisioning functionalities for optical transport networks, i.e., fulfills connectivity set-up requests, issued by applications in terms of end-host addresses and perceived quality of service, while performing self-consistent and network-wide device settings based on a distributed signaling among “service nodes” called Distributed Service Elements (DSEs).

Two different architectures were investigated in Y2: a Service-Oriented Multi-Granular Optical Network (SO-MGON) architecture and a SIP-based SOON architecture. The sub-activities concerning the two architectures are detailed next.

2.3.1.1 Service-Oriented Multi-Granular Optical Network (SO-MGON)

The first activity is the follow up of the activity started during Y1 and aims at further investigating the performance of the proposed Service-Oriented Multi-Granular Optical Network (SO-MGON) architecture shown in Figure 13. SO-MGON consists of a SOON platform enabled for a multi-granular optical network to support dynamic wavelength and sub-wavelength granularities with different transport formats (i.e., OBT, OBS), reservation protocols (one-way, two-way) and different QoS levels per service type.

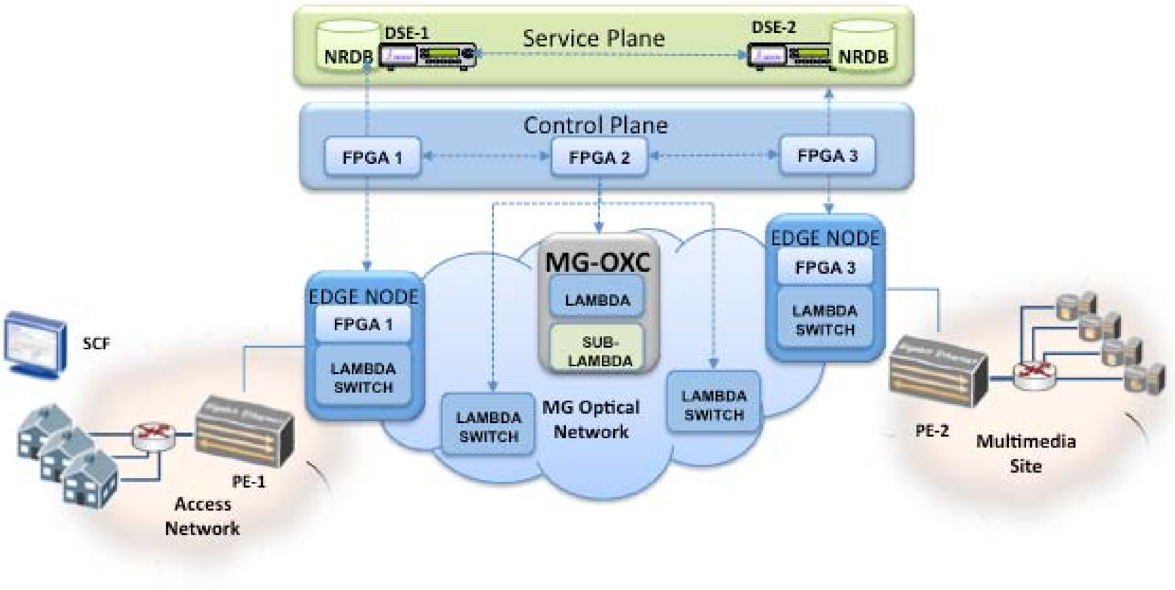


Figure 13: Architectural block diagram of the SO-MGON testbed

The SOON supports Service Abstraction and Resource Virtualization capabilities. Thanks to such capabilities, applications can issue service request without specifying technology-specific parameters. In fact, SOON can map a set of application-specific parameters into a set of parameters used by the network for the actual configuration of a service, while avoiding the disclosure of the technology-related details of network resources (such as burst-dimension, assembly time, network transport service (i.e., OBS, OBT), end-to-end lightpath) to applications. In this context, applications are represented by software modules implementing Service Control Functions (SCFs) that are responsible for session-based signaling, e.g., SIP-based call or multimedia transfer, aiming at the control of application resources involved in the content transmission, such as codec, delay bound, max bandwidth.

Specifically, the SOON platform has the ability to coordinate the different OBS edge devices to create unidirectional and bidirectional end-to-end wavelength (OBT) and sub-wavelength (OBS) paths. This is obtained through an ad-hoc signalling protocol among DSEs for correlating the configuration parameters and thus for carrying out a consistent settings of network nodes. DSEs also periodically collect and correlate the boundary topology information (e.g., the address of the client networks) from the controlled network devices, and then store the overall boundary topology information in the Network Topology and Resource Database (NTRD) to be used of the resolution of edge node address from the end-host address. To enable the interaction with the OBS network nodes, a specific technology-dependent module installed in each DSE has been conceived, which translates the information received from the DSE into a set of specific directives comprehensible by the OBS edge devices.

The service-aware edge OBS router, which utilizes a network processor and FPGA devices, is able to differentiate between SOON signalling messages (i.e., among DSEs), network requests, and data packets. In case of SOON signalling messages, the edge router forwards them to the Control Plane. In case of incoming SOON network requests, the burst aggregation scheduler is triggered to reflect service requirements into buffer size and the time thresholds.

The SOON signaling message is also used to decide the network provisioning system (i.e., either OBS or OBT) and the technology-specific parameters to configure on network devices. The OBS provisioning system is carried out by aggregating bursts and in turn generating and transmitting a Burst Control Header (BCH) ahead of time (5μs) in order to configure the acousto-optic switch on per burst basis. The OBT provisioning system reserves an end-to-end lightpath for the duration of the service by generating a BCH upon reception of a SOON message. The combination of wavelength selection, carried out by controlling a tuneable laser connected to a MEMS switch (for connectivity with all core nodes), together with the BCH processed at the core FPGA provides the end-to-end dynamic lightpath. Finally the data packets are buffered and then transmitted over different wavelength or sub-wavelength lightpaths.

The provisioning system and the technology-specific parameters are decided based on service that is requested by the SCF application. Four different services are considered on this experiment: 1) QuadHD-VoD, 2) VoIP, 3) Video Streaming, and 4) raw data transfer. Each of these has been mapped to different functions provided by the network according to the application's QoS requirements, as shown in Table 1.

Service (SOON message)	Technology specific parameters (OBS functionalities)			
Application	Burst Size (bytes)	Assembly time(μs)	Wavelength/ Path	Transport Service
1) QuadHD –VoD	5120	5120	Lambda I/ Path 1	OBS
II) VoIP	4096	4096	Lambda II / Path II	OBT
III) Video Streaming	3072	3072	Lambda III / Path III	OBT
IV) Raw data transfer	2048	2048	Lambda IV/ Path IV	OBT

Table 1: SOON services mapped on technology-specific parameters

The service-aware core OBS router comprises three nodes, all controlled by a centralized FPGA-based control plane module. Two nodes are equipped with MEMS switches (10 ms) while the third one consists of both MEMS switch and acousto-optic switch (4 μs) to form a multi-granular optical cross-connect (MG-OXC). The Control Plane module utilizes a network processor and FPGAs and it can process and forward service layer information on the fly as well as allocate resources in the MEMS switch for either OBS or OBT provisioning.

SOON concept is the result of the decoupling of network technologies from services. In this way, the Control Plane is unburdened from the task of performing service-oriented functionalities and it can focus on the provisioning of connectivity services. Unlike the existing service oriented architectures, which use the legacy IP network for carrying SOON signals and messages, in the proposed architecture SOON messages are carried in the optical domain via JIT signaling.

2.3.1.2 SIP-based Platform for SOON

The second sub-activity of JA3 is the follow up of the activity started during Y1 and carried out in collaboration between SSSUP and UniBO. During Y2, the performance of a SIP-based platform for service oriented optical transport networks operating on top of a GMPLS-enabled infrastructure according to SOON principles was further investigated. In particular, the system performance has been evaluated in terms of capacity required to satisfy multiple requests of a single media flow channel, issued at a different rate.

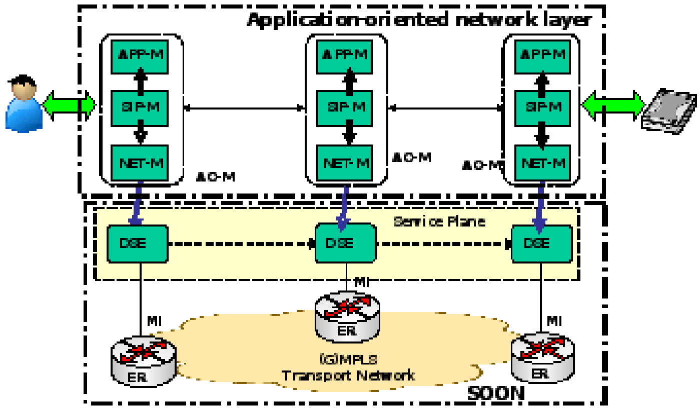


Figure 14: SIP-based service platform: building blocks and functional layers

The functional architecture of the proposed service platform has been extensively discussed in D12.1. Here it is briefly described and depicted in Figure 14. From an architectural point of view, the service platform can be decomposed into three functional layers or planes:

- *Application-Oriented Network Layer*, that manages session-based signaling among applications enhanced with resource management capabilities, regarding both IT resources (i.e., remote storage space, multimedia content) and network resources (i.e., QoS-enabled media flows). It resides in the *Application-Oriented Module* (AO-M) where different sub-modules are in charge of processing the incoming resource requests (APP-M); managing the service sessions by means of the SIP protocol (SIP-M); interacting with the underlying Service Plane for issuing connectivity service requests to activate the QoS-enabled media flow transfer (NET-M). The latter requests trigger the resource provisioning mechanism across the network and are specified in terms of end-host addresses and perceived quality of service.
- *Service Plane* (SP), that elaborates the connectivity service requests issued by the AO-M. Specifically, the SP performs the admission control and, in case of resource availability, it fulfills each request by configuring the network-specific traffic policies on the relevant edge routers in a consistent way, in order to enable a data flow with the required QoS assurances between end-hosts. This operation is based on a distributed signaling among “service nodes” called *Distributed Service Elements* (DSE).
- *GMPLS-enabled Network Layer*, which is responsible for data forwarding.

The coupling of the GMPLS transport network with the SP represents a Service Oriented Optical Network (SOON), i.e., a service-enabled network infrastructure.

2.3.2 Achieved Results in Y2

The results achieved during Y2 in the two sub-activities are the followings.

2.3.2.1 Service-Oriented Multi-Granular Optical Network (SO-MGON)

The experimental evaluation of the SO-MGON has been performed on a OBS network infrastructure located at the Essex University facility, using the SOON prototype implemented by SSSUP personnel and adapted for OBS network nodes. The testbed, shown in Figure 15, consists of the SOON platform composed of DSEs operating of top of OBS control and data plane according to the architecture shown in Figure 13.

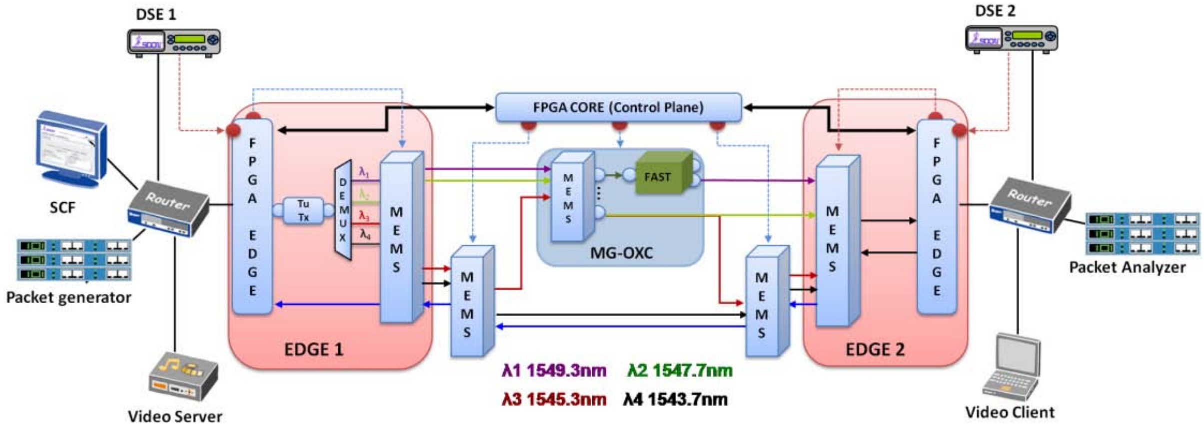


Figure 15: SOON-enabled OBST testbed setup

Measurements of SOON provisioning time i.e., overall time elapsed during the SOON network service provisioning process, were taken during the experiments. In addition, the Frame Error Ratios (FER) experienced along three different paths have been evaluated using different wavelengths and different numbers and types of optical switches (i.e., MEMS and acousto-optic).

Figure 16 shows the sequence of messages generated by service provisioning request and characterizes the latency of the different operations performed by SOON and the overall SOON provisioning time. On the left of the figure, it is possible to see that the SOON processing time is of about 30 ms. Once the request is processed, the signalling (blue peak) is triggered. Then, additional latency is required for generating specific directives to configure the edge node (pink peak). Once the edge node is configured, the SOON receives the ACK message from the DSE module and sends an ACK to SCF application (red line).

The SOON signaling protocol performance has been evaluated by measuring the end-to-end service time, which includes the edge and core node parsing and forwarding time. This value (about 60 ms) is mostly dependent on the end-host performance used for SOON elements and not on the network device performance.

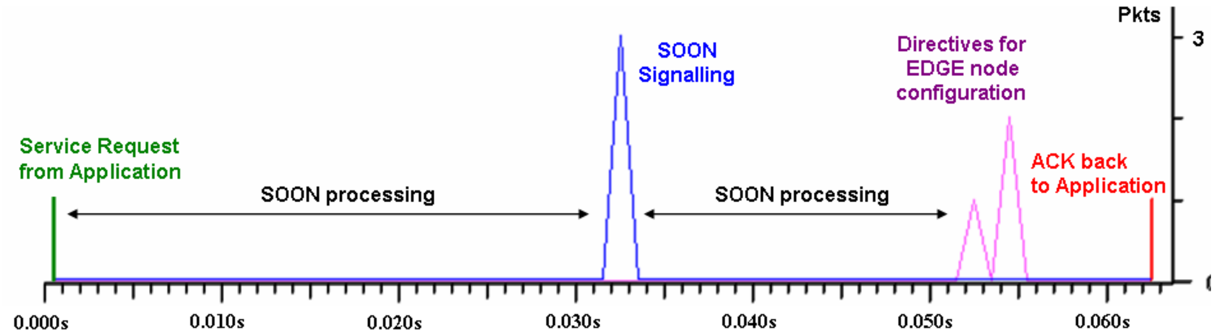


Figure 16: SOON-enabled OBST Testbed Setup

Finally, Figure 17 illustrates the physical layer performance (FER) of three different paths using three different wavelengths and different numbers and types of optical switches. For comparison purposes, the back-to-back (B2B) performance of the end-to-end testbed are included. The paths (e.g., Lambda I/Path I, Lambda II/Path II, Lambda III/Path III, all shown on both Figure 16 and Figure 17) experience a power penalty of 1.5 – 2.5 dB, which is well within acceptable levels considering the fact that the whole testbed is optimized once for all possible paths and the fact that each path is using different number of MEMS switches (two up to five) or even a combination of MEMS switches and acousto-optic switch.

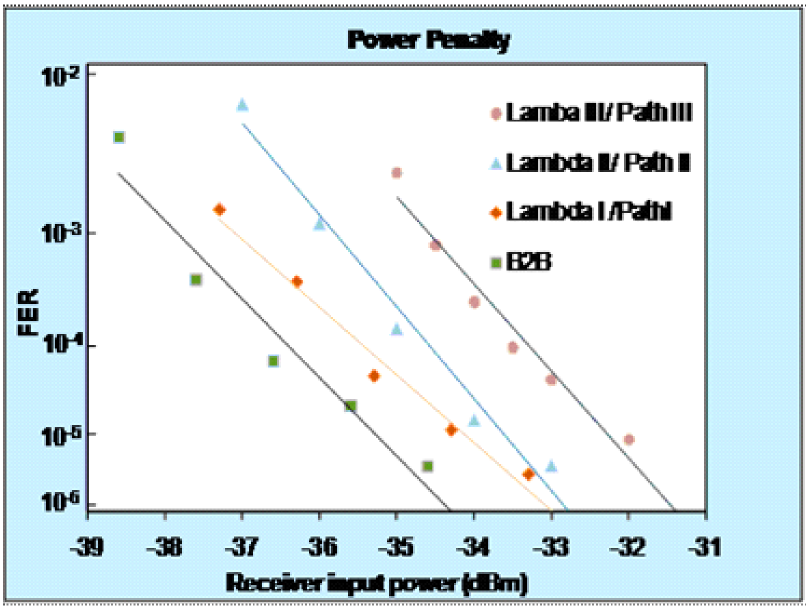


Figure 17: Frame Error Ratio (FER) of three different network paths

2.3.2.2 SIP-based Platform for SOON

The experimental evaluation of the SIP-based service platform for GMPLS-enabled transport network has been carried out by means of a collaboration by SSSUP personnel and UniBO personnel for the purpose of setting up a testbed reproducing the three functional layers depicted in Figure 14. Specifically an MPLS network infrastructure enhanced with SP functionality, i.e., a SOON infrastructure, was set-up by SSSUP personnel. An enhanced SIP-based signaling among AO-Ms triggered by media flows set-up request including network resource requirements was set-up by UniBO personnel. The two building blocks have been integrated. In this way, a request for network resource provisioning issued by the enhanced SIP signaling can be elaborated by the SOON. SOON translates it into a set of consistent settings for involved network nodes.

The integrated test-bed set-up is shown in Figure 18. The AO-M has been deployed on a Linux Box by extending open-source software implementing a SIP proxy server and by building the APP-M and NET-M functionalities. In principle, an AO-M operates for each DSE. In this work, without lack of generality, a single AO-M was included and coupled with DSE3. The SP is made of 3 Linux Boxes running an instance of the DSE module and connected to both the public Internet and to the Edge Router (ER). Each DSE controls the corresponding ER by issuing NETCONF directives [1] while the inter-DSEs communication is based on XML messages. The metro-core network comprises six IP/MPLS routers, with ER1, ER2 and ER3 configured as provider ERs and the CRs configured as core routers. All routers support MPLS, OSPF and RSVP with extension for DiffServ-aware Traffic Engineering (DS-TE) [2] for enabling the reservation of network resources on a per-traffic class basis. The SIP User Agent Client (UAC) and Server (UAS) are connected to the transport network via a router configured as Customer Edge (CE) node, representing the gateways out of the customer premises. These are connected to the transport network through VLANs that isolate the traffic flows of different application categories, e.g., video, voice, data. Without lack of generality, we consider the case of video traffic. Therefore an LSP is established between ER1 and ER2 (LSP-video) with reserved bandwidth of 80Mbit/s and with queue schedulers configured accordingly in both core and edge routers to guarantee rate assurance to video packet delivery. Upon service request, a portion of such bandwidth is guaranteed for the media flow between the UAC and UAS through traffic policy settings.

The testbed demonstrates and evaluates the performances of on-demand establishment of QoS-guaranteed media flow. Performance are evaluated in terms of service set-up latency due to the signaling scheme using different request arrival rate and different delay for network node configuration. We considered a request demanding a single resource (e.g., a user request for a single media flow transfer across the network, for instance for a video application).

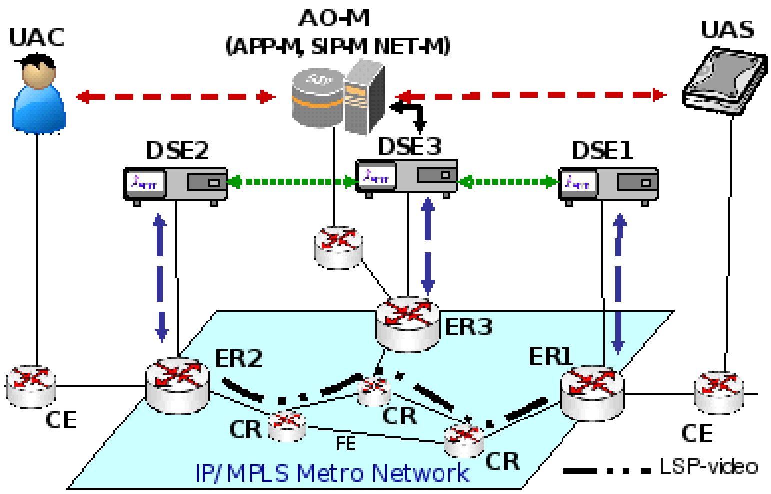


Figure 18: SIP-based service platform: experimental set-up

The signaling workflow adopted by the proposed service platform is shown in Figure 19. When the UAC sends, at time t1, a requests for a given service (e.g., HD video) to be provided by the UAS, an INVITE message is forwarded to the AO-M including the description of the required network resources (e.g., media flow with HD video quality). The SIP-M sub-module, acting as an enhanced SIP Proxy Server (PS), processes the INVITE message, extracts its content and passes the network resource specifications (NRDL [3]) to the NET-M, thus triggering NET-M to issue a connectivity service request (SRQ) to the DSE3 for network resource reservation. The service request consists of an XML file including the UAC and UAS IP addresses and the service requirements: service type (i.e., video flow) and quality level (i.e., HD). The DSE3 determines the edge routers to be involved as well as the traffic category and amount of bandwidth (e.g., 20 Mbit/s for HD video) to be reserved within LSP-video. As shown in Figure 19, with the support of DSE1, the DSE3 verifies (RARQ message) whether such bandwidth is available across the network (BARQ message) and, in case of positive response, it triggers the bandwidth reservation by asking the DSE1 and the DSE2 (CSRQ message) to enforce traffic policies on both ER1 and ER2 (NSRQ message) that ensure the proper forwarding of video packets across LSP-video. When the reservation is completed, the DSE3 informs NET-M that the requested resources are available (SRP message) and the NET-M asks SIP-M to complete the session set-up (NRDL message). Thus, SIP-M sends an INVITE message. The INVITE message is then received at time *t2* by the UAS that replies with an OK message at *t3*, as required by the SIP three-way handshake. The ACK sent by UAC at *t5* concludes the session setup and data can now flow on the LSP-video with the required QoS.

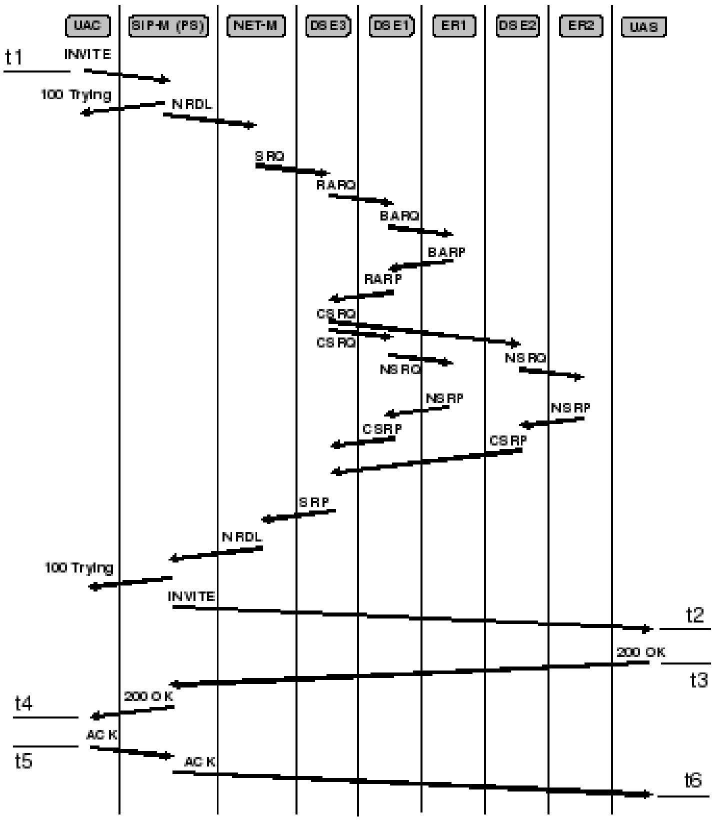


Figure 19: Message flow for application-driven service configuration and CSD components: CSD = (t6-t1) – (t3-t2) – (t5-t4)

NRDL	Network Resource Description Language
SRQ	Service Request
RARQ	Resource Availability Request
BARQ	Bandwidth Availability Request
BARP	Bandwidth Availability Reply
RARP	Resource Availability Reply
CSRQ	Connectivity Set-up Request
NSRQ	Network Setting Request
NSRP	Network Setting Reply
CSRP	Connectivity Set-up Reply
SRP	Service Reply

Table 2: List of acronyms used in Figure 7

The service set-up latency has been evaluated in terms of *Call Set-up Delay* (CSD) defined as in ITU-T recommendation Y.1531 [4]. Y.1531 reports the specifications for the performance parameters defined for call processing, where by call is intended a generic media session between end-users established using SIP or H.323. The CSD is defined as the overall time needed to establish the call between the end-hosts (i.e., SIP terminals), i.e., the time elapsed between the caller's issuance of a SIP INVITE message at the caller UNI (*t1*) and the callee's receipt of the corresponding SIP ACK message at the callee UNI (*t6*), excluding:

- the callee delay, between the callee's receipt of the INVITE message (*t2*) and issue of the corresponding 200 OK message (*t3*);
- the caller delay, between the caller's receipt of the 200 OK message (*t4*) and issuance of the corresponding ACK (*t5*).

In practice, the CSD takes into account the transfers and delays associated with the internal network functions that occur between end-hosts, which implement admission control and resource reservation across the transit networks, i.e., the IP/MPLS network.

The performance has been evaluated by averaging the CSD on N request samples, generated according to two different traffic profiles: in one case, N = 30 requests are periodically generated from the UAC interleaved by a fixed inter-arrival time (IAT) value; in the other case, N = 100 requests arrive according to a Poisson process with a given average IAT. All requests are related to the same LSP-video resource as previously described. In order to evaluate the latency due to the signaling scheme, the first set of experiments has been executed assuming a negligible delay due to bandwidth availability check and router configuration operations. The measured CSD is shown in Figure 20 as a function of the request IAT for both the deterministic and Poisson profiles, with 95% confidence interval. The CSD is much lower than the 7.5s threshold recommended by ITU-T in Y.1530 [5], even in case of high request rates (e.g., two per second) asking for the same network resource.

A more realistic measure of the set-up delay experienced by the UAC when requesting a service can be obtained by observing the impact on the CSD of the latency due to network equipment configuration, as shown in Figure 21 for four different values of the average IAT with Poisson arrivals and 95% confidence interval. Using the plot, it is possible to dimension the rate of requests asking for a given network resource that can be satisfied while maintaining the CSD under the recommended 7.5s threshold. The upper limit of 3.5s refers to the router configuration delay attainable with the current technology, as resulted from measurements performed on our test-bed.

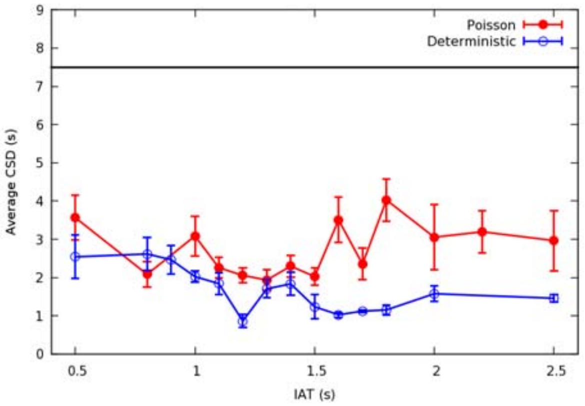


Figure 20: Average CSD vs. request IAT (signalling only)

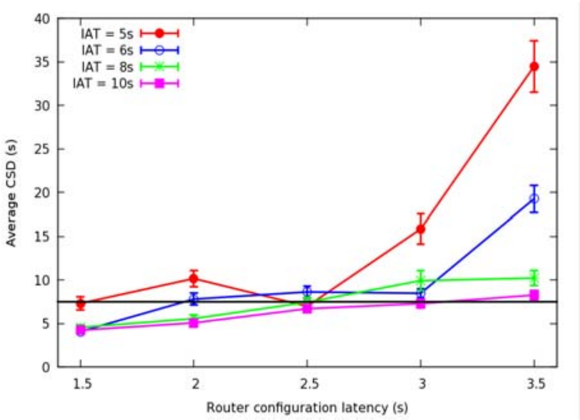


Figure 21: Impact of the router configuration latency on CSD

These experiments show that the proposed scheme is actually feasible and provides scalable performance in line with ITU-T recommendations. The impact of the router configuration latency on the service set-up delay has also been evaluated showing how it limits the highest supported request rate while meeting CSD requirements.

2.3.2.3 References

[1] R. Enns, “NETCONF Configuration Protocol,” RFC 4741, Dec. 2006.

[2] F. Le Faucheur, W. Lai, “Requirements for Support of Differentiated Service Aware MPLS Traffic Engineering,” RFC 3564, Jul. 2003.

[3] F. Callegati, A. Campi, “Network Resource Description Language”, Proc. 3rd IEEE Workshop on Enabling the Future Service-Oriented Internet, in conjunction with IEEE Globecom 2009, Honolulu, HI, Dec. 2009.

[4] ITU-T Rec. Y.1531 “SIP-based call processing performance”, Nov. 2007.

[5] ITU-T Rec. Y.1530 “Call processing performance for voice service in hybrid IP networks”, Nov. 2007.

2.3.3 Published Joint Papers (joint and single partner)

“SIP-based Service Platform for On-demand Optical Network Services”, B. Martini, F. Baroncelli, V. Martini, K. Torkmen, P. Castoldi, A. Campi, F. Zangheri, W. Cerroni, F. Callegati, Proc. of OFC 2009, March 24-26 2009, San Diego, California USA.

“Service-Oriented Multi-Granular Optical Network Testbed” , Y. Qin, G. Zervas, V. Martini, M. Ghandour, M. Savi, F.Baroncelli, B.Martini, P.Castoldi, C. Raffaelli, M. Reed, D. Hunter, R. Nejabati, D. Simeonidou, Proc. of OFC 2009, March 24-26 2009, San Diego, California USA.

“Data-Plane Architectures for Multi-Granular OBS Network” , M. Savi, G. Zervas, Y. Qin, V. Martini, C. Raffaelli, F.Baroncelli, B.Martini, P.Castoldi, R. Nejabati, D. Simeonidou, Proc. of OFC 2009, March 24-26 2009, San Diego, California USA.

“SIP-based Service Architecture for Application-aware Optical Network”, A. Campi, W. Cerroni, G. Corazza, F. Callegati, B. Martini, F. Baroncelli, V. Martini, P. Castoldi, 2° International Conference on Information and Communication Technology (ICTA), Hammamet, Tunisia, May 2009

2.3.4 Planned Activities for Y3

Within the activities related to the SIP-based service platform, additional tests are planned in case of multiple requests processed in parallel both in case of single resource and multiple resources from different users in different location. Such requests could require a multimedia media transfer for a video-voice application, bidirectional channel or the commitment of additional resources or the release of unused resources.

For multiple requests requiring multiple resources, futher enhancements of testbed software will be implemented for enabling parallel processing of requests.

2.3.4.1 Targeted Publications

“A Service-Oriented Multi-Granular Optical Network Architecture for the Clouds”, G.S. Zervas, V. Martini, Y. Qin, E. Escalona, R. Nejabati, D. Simeonidou, F. Baroncelli, K. Torkmen, P. Castoldi, submitted to Journal of Optical Communications and Networking

A journal paper is in preparation on joint research activity carried out by UniBO and SSSUP.

2.4 JA 4: Joint Optimisation of Grid and Network Resources

Participants: BME, SSSUP, PoliTO

Responsible person: Tibor Cinkler (BME)

2.4.1 Objectives for Y2

The objective of Joint Activity 4 is to consider and optimise ***jointly***, on the one hand, the Services and Applications (S&A), and on the other hand, the optical network resources.

The services and applications are assumed to be GRID services, namely network-wide distributed storage and computing services. Nowadays it seems a preferred strategy to virtualise and share own resources than building huge private supercomputers and data storage centers. This approach is much more efficient for bursty computing requirements, and can be much cheaper for end users to cover slightly higher power-costs, however, having access to practically unlimited parallel computing resources.

The considered networks are assumed to be controlled by the GMPLS protocol family. Different data planes were studied that range from a general packet-switching capable one, to more heterogeneous wavelength-switched and packet-switched grooming-capable one.

The composition of partners is as follows: SSSUP having expertise in CP issues, PoliTo having expertise in Storage issues, BME having competence in Computing GRID issues, and all three partners experienced in related optimization issues.

The activities in Y2 target the following objectives:

1. Using simpler and faster yet efficient heuristics based on greedy algorithm and shortest path searches instead of ILP for optimising GRID ***storage*** resources.
2. Using simpler and faster yet efficient heuristics based on greedy algorithm and shortest path searches instead of ILP for optimising GRID ***computing*** resources.
3. Applying joint optimisation of network resources and GRID storage resources assuming that the network consists of two layers, a WDM wavelength switched layer and a TDM capable layer over it.
4. Considering ***resilience*** issues in joint optimisation of network resources and GRID computing resources. Namely, if a network link fails or a computing node fails a part of results is not received by the initiator. Therefore different scenarios are investigated ranging from those that optimise the speed to those that optimise the cost.
5. Evaluating the GMPLS control plane support for the above methods.

2.4.2 Achieved Results in Y2

To achieve JA4 objectives, the work was organized in sub-activities. The progress of the sub-activity is reported next.

2.4.2.1 Sub-activity 2 of JA4

Heuristic methods for reducing the state-space, thus reducing the time required for optimization, have been proposed. Furthermore, a simple greedy algorithm has been proposed that is based on recursive use of Dijkstra's shortest path algorithm.

2.4.2.2 Sub-activity 3 of JA4

The model of PoliTO for optimizing storage resources over a single layer network is being extended with the multilayer optical beared network model of BME.

2.4.2.3 Sub-activity 4 of JA4

A task distributed between multiple distant GRID computers is considered executed if all subtasks are executed and all the sub-results are collected. Either a network or GRID-computer failure can hinder a task to be executed. Therefore, we consider the following strategies:

1. If network resources fail, in either direction at any time, a new connection will be used or set up either via protection or via restoration, respectively.
2. If any of the GRID computing resources fail the task will be sent out to closest fastest GRID unit to execute it as fast as possible not optimizing for the cost any more, since the result is obtained only if all subtasks are all completed.
3. All the sub-tasks are sent to two geographically distinct nodes connected via physically diverse network resources, the sub-tasks executed twice, and sub-results collected. This is the best from the resilience point of view, since multiple network and computing unit failures can be easily overcame with no delay at all, the accuracy of sub-results can be verified, unfortunately it waists both, computing and transmission resources.

Co-operation with JA7 was discussed. Particularly sub-activity 4 of JA4 that deals with resilience is closely related to JA7. Merging of JA4 and JA7 was also proposed, not yet decided.

2.4.2.4 Sub-activity 5 of JA4

To meet the QoS requirements of global grid-enabled applications, a dynamic and deterministic joint optimisation of both computational and network resources is necessary. In wavelength-switched optical network (WSON), the joint optimisation is needed when bandwidth-greedy applications require the provisioning of lightpaths among grid resources. A way of achieving this joint optimisation is the cooperation between the grid resource manager (GRM) and the network resource manager (NRM). The typical objective of GRM is to achieve the maximisation of the overall amount of grid services successfully established and the minimisation of delivery time subject to the required QoS constraints. However, the lack of network information at the GRM does not guarantee the efficient network resource utilisation and, in turn, the maximisation of the established grid services.

The introduction of the path computation element (PCE) within the control plane of optical networks is an appealing solution to provide efficient utilisation of network resources. The PCEP protocol is proposed here as the new standard interface between the GRM (behaving as PCC) and the NRM (embedding a PCE implementation). Two novel schemes for grid resource selection specifically designed to exploit the standard PCE protocol (PCEP) features are proposed. The schemes resort to the currently defined PCE architecture and PCEP implementation to provide a feedback to the GRM on the expected network resources utilised by alternative choices of computational resources. In this way, the GRM can evaluate in advance the impact on the network performance of different alternatives and, among them, select the one that minimises the grid service delivery time and the overall network resource utilisation. The proposed schemes are suitable in case of grid services requiring large connection bandwidth, high level of connectivity among multiple grid resources (e.g., full mesh of lightpaths) and long duration. These schemes do not require additional control plane extensions or interfaces specific for grid purposes. In addition, the proposed schemes preserve the requested level of confidentiality between GRM and NRM, concerning the detailed network information (i.e., link bandwidth availability and strict routes information are not disclosed by NRM). Therefore, the proposed PCEP-based solution can be implemented at NRMs, e.g., independent internet service providers (ISPs) not belonging to any grid virtual organizations.

Different grid-networking schemes with specific relationship between GRM and NRM are considered. The former two schemes are generally applied, the latter two are specifically proposed for joint optimization:

- *Dg scheme*: GRM selects the least loaded *g* computational resources. NRM is in charge of setting up network connections among them. Lightpaths are computed and provisioned by GMPLS protocols in a distributed way. The PCE is not utilized. Joint optimisation is not achieved.
- *Pg scheme:* GRM selects the least loaded *g* computational resources. NRM acts as PCE having full and detailed visibility of the network and may perform multiple and concurrent path computation. Network resources are better utilized, however joint optimization is not achieved.
- *Pkg scheme*: GRM considers the combination of all the possible least loaded *g* computational resources (assuming *k* equally loaded resources). NRM acts as PCE and is requested to perform multiple concurrent path computation of combinations furnished by GRM. NRM replies providing the GRM with path computation solutions including a TE metric representative of network QoS level. GRM selects the solution minimizing the combined grid-network metric.
- *PGkg scheme*: in an offline PCEP-based communication, GRM requests a non-synchronised independent path computation between all node pairs connecting the grid resources potentially involved. PCE replies with a related static metric value representing a lower bound of the actual metric (i.e., optimistic value, considering the shortest path which is not always available). Such metrics are used by GRM upon new grid requests occurrences, identifying the *k* resources that minimise the considered metrics. Random selection performed in *Pkg* scheme is thus avoided. Successive steps are the same as *Pkg* scheme.

Scheme	*Description*	*GRM*	*NRM*	*Complexity (# path computations)*
Dg	Grid networking scheme not utilising the PCE	g	Distrib.	$p = g(g-1)/2$
Pg	Grid networking scheme utilising the PCE	g	PCE	$p = g(g-1)/2$
Pkg	PCEP-based scheme with feedback information	$k\ (>g)$ g	PCE	$cp = k!/(g!(k-g)!\ g(g-1)/2$ $p = g(g-1)/2$
PGkg	PCEP-based scheme with feedback information and a-priori communication	G $k\ (>g)$ g	PCE	$cp' = k!/(g!(k-g)!\ G(G-1)/2$ $cp = k!/(g!(k-g)!\ g(g-1)/2$ $p = g(g-1)/2$

Table 3: Considered grid networking schemes

The four schemes are summarized in Table 3, together with their associated complexity (i.e., number of required path computations), where *G* is the total number of grid resources in the network.

The considered schemes are evaluated through simulations considering the NSF network topology where each link carries 8 wavelengths and one grid resource per node is considered. The case with $g = 3$, and thus $p = 3$ bidirectional lightpaths is reported. Services are sequentially established and never torn down. The path computation is based on integer linear programming (ILP) formulation. The implemented objective function minimises the overall amount of used network resources and, as secondary objective function, minimises the load of the most loaded link.

Figure 22.1 and Figure 22.2 show the blocking rate versus the load. Results confirm that *Dg* and *Pg* achieve similar results. The two curves are almost completely overlapped, due to the low number of equal cost shortest routes of the topology. Scheme *Pkg* with $k = 4$ and $k = 5$ significantly improves the blocking rate compared to *Dg* and *Pg* schemes. The *Pkg* scheme with $k = 5$ achieves a higher blocking rate compared to the case with $k = 14$. The *PGkg* scheme, which performs upon service request 10 path computations, further improves the *Pkg* performance, closely approximating the optimal performance achieved by *Pkg* or *PGkg* with $k = 14$ (requiring, however, 364 path computations).

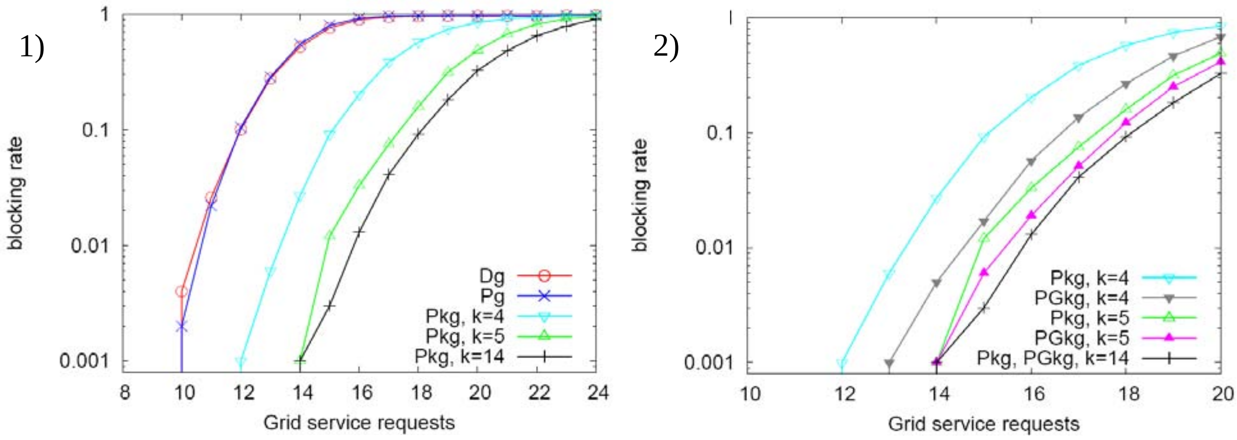


Figure 22: Simulation results

The proposed schemes are experimentally evaluated on the 'JGN2plus's metropolitan area WSON testbed (Figure 23). The GMPLS control plane runs on control PCs exploiting generic routing encapsulation (GRE) connections. The data plane consists of 4 wavelengths generated by gigabit Ethernet interfaces equipped with DWDM media converters. The NRM consists of two sub-blocks: the PCE and the lightpath activator. The PCE is implemented in a Linux PC and is based on C code and ILP formulations. The PCEP module is based on C++ and TCP socket libraries. The five edge nodes are connected to equally least loaded grid resources (i.e., G = 5). The GRM triggers one grid service si, requiring bidirectional lightpaths connecting in a full mesh g = 4 grid resources.

When the Dg scheme is applied, if g grid resources randomly selected includes router R1, the service is rejected since only one wavelength is available on the link connecting router R1 to OXC X1 (i.e., full mesh of lightpaths cannot be supported). If the selection excludes R1, the service may also be rejected because of lack of network resources on the other links (e.g, unless the lightpaths are assigned wavelength $w0$ for lightpath R2-R5 and wavelength $w3$ for lightpath R3-R4). The Dg scheme experiences a service rejection equal to 97.7%. When the Pg scheme is adopted, the latter critical condition affecting Dg is avoided thanks to the concurrent path computation. However, the former critical condition related to the random choice of R1 persists. Service rejection is reduced to 80%. By applying the Pkg and the $PGkg$ schemes with k = 5, the service is successfully established since both previous critical conditions are avoided. The five combinations of p = 6 lightpath computations exclude router R1 and allocate the wavelengths that avoid service rejection.

The figure indicates also the sequence of PCEP message exchanged between the grid user and the NRM. Pkg and $PGkg$ scheme requires an additional PCEP message. However, such additional PCEP does not significantly increase the overall grid service delivery since the overall lightpath activation latency is dominated by the high OXC switching time (e.g., some seconds). For this reason, grid services exploiting the current optical network technologies can be provisioned dynamically but at a limited rate.

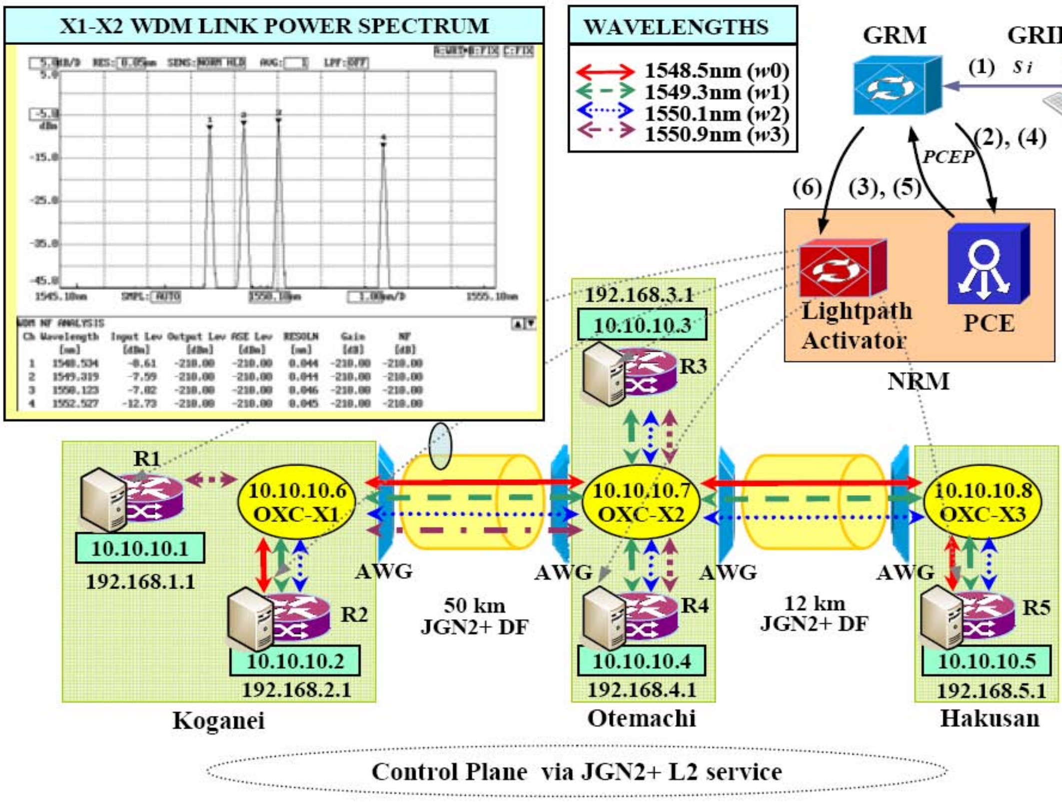


Figure 23: Experimental implementation on the JGN2plus testplan

2.4.3 Published Joint Papers (joint and single partner)

No joint papers published in 2009.

2.4.4 Planned Activities for Y3

During Y3, the partners are planning to complete the sub-activities 3, 4 and 5 started in Y2 (2009) as explained in Section 2.4.2. Also, partners will be committed to disseminate the results, and establish a close co-operation with JA7.

2.4.4.1 Targeted Publications

ICTON 2010, all three JA4 partners involved.

2.5 JA5 - Programmable Service Composition Algorithms for Service Oriented Optical Networks

Participants: IT/Telecom-ParisTech, UESSEX, UoP, RACTI, SSSUP

Responsible person: Chinwe Abosi (UEssex)

2.5.1 Objectives for Y2

Future Internet applications require different type of resources: Computational, Storage, Visualization and Network resources. Depending on the application, it may require access to the corresponding resources. Those resources will be heterogeneous, managed by different administrative entities and not collocated. A service provider will have to aggregate the information coming from all of the resources and make them seem to behave as if they were all similar. These resources may or may not employ virtualization technologies. For example, within a computational cluster, through virtualization, resources can be aggregated or segregated to satisfy a particular application demand. In the same way network resources can be sliced to be rented or sold to different service providers.

However, only through a unified means in which resources can be holistically viewed can the true benefit of virtualization be achieved.

The objective of JA5 is to design programmable algorithms that support effective and efficient implementation of service oriented optical network functionality such that the optical network and IT resources are holistically provisioned. The programmable algorithms developed within this JA will:

- enable unified coordination and interaction between both network and IT resources which in turn enable cross-optimized allocation of the desired resources to satisfy user requirements
- introduce a service differentiation in Grid Services provisioning
- introduce and investigate economical models on Grid Services provisioning.

The results of this JA will contribute to ensuring end-to-end quality of experience, improved scalability and optimized use of resources and economical viability.

In Year 2, JA 5 focused on

- <u>Service differentiation in Grid environment (IT/Telecom-ParisTech)</u>

  Grid services aim to dynamically associate remote computing and storage resources via networking facilities in order to provide a given service to end-users generating job demands. Grid services provisioning require the definition of suitable business models taking into account the expectations of both end-users and service providers. Traditional resource management techniques mainly deal with end-users requirements but neglect the benefit that service providers may expect from the sharing of their equipments. We propose a faire management algorithm defending the interests of both the end-users and the service providers. We consider scheduled jobs in the context of layer-2 Virtual Private Networks (VPN).

- Impact of network topology abstraction on resource virtualization (IT/Telecom-ParisTech/UEssex)

  The objective is to numerically evaluate via simulation the impact of network abstraction strategies developed at UEssex on the network resources virtualization techniques developed at Telecom ParisTech, a subsidiary of GET (now Institut Telecom).

- Service-oriented resource provisioning through resource co-selection, co-reservation &co-allocation (UEssex)

  Service-oriented resource provisioning has to be performed within a single uniform, integrated platform that combines and orchestrates heterogeneous (network and IT) resources such that they can be selected in a single step. The effect of a service plane platform on different allocation granularities of the underlying network resources is investigated. The impact of the service plane on migrating traffic patterns as the requirements of emerging future Internet applications transform

- Investigation of the position of the service plane which implements the service composition algorithms in relation to the infrastructure provider (UEssex)

  An entity that is able to efficiently coordinate optical bandwidth, processing and storage capacities owned by different infrastructure providers will most likely be an independent service provider. This independent service provider needs means to protect the infrastructure provider from divulging details of its resources. The aim of this sub-activity is to evaluated several abstraction schemes and evaluate performance trade-offs faced by the infrastructure providers and users when (a) the service plane is the same entity as the infrastructure provider, i.e. no need for infrastructure abstraction and (b) the service plane is a third party entity.

- Definition of a mapping framework between connectivity requests issued by applications in terms of end-host addresses and perceived QoS parameters and technology-dependent network directives (SSSUP).

  The objective is to enable an automatic and self-consistent node configuration across the network for on-demand connectivity set-up while avoiding error-prone correlation activities carried out by human operators. Thus, signalling architecture are required for configuring the network. Network configuration based on distributed entities, that cooperate by means of a dedicated signalling scheme and correlate network configuration parameters for a self-consistent and automatic configuration of MPLS-based VPN, is investigated.

2.5.2 Achieved Results in Y2

To achieve JA5 objectives, the work was organized in sub-activities. The progress of the sub-activity is reported next.

2.5.2.1 Service Differentiation in Grid environment

Grid services aim to dynamically associate remote computing and storage resources via networking facilities in order to provide a given service to end-users generating job demands. It is important to specify for a Grid Service Provider (GSP) business models taking into account the expectations of both end-users and service providers (network providers, storage providers and computing resources providers). Traditional resource management techniques mainly deal with end-users requirements but neglect the benefit that service providers may expect from the sharing of their equipments.

A fairer management algorithm defending the interests of both the end-users and the service providers is proposed. For that purpose, we consider scheduled jobs in the context of layer-2 Virtual Private Networks (VPN). In practice, the instant of activation of a scheduled job may vary over a certain time window. Depending on the requirements and the budgets of end-users, the size of this time window may fluctuate, the job computing delay being fixed *a priori*. Our motivation is two-fold. First, we can expect to serve a higher number of end-users while respecting each user’s job utility. Second, we aim to increase as best as possible the benefit of service providers.

For that purpose, we introduce a weighted cost function enabling a service differentiation relying on time constraints disparity. Two approaches are considered to solve this problem: an exact approach based on an Integer Linear Programming (ILP) formulation and a meta-heuristic inspired from the Simulated Annealing algorithm. The obtained numerical results outline the impact of the flexible time window on the system performance. System performance can be evaluated by means of the number of executed jobs and the gain of the service provider.

2.5.2.2 Impact of Network Topology Abstraction on Resource Virtualization

A mobility action has been established between UEssex and GET in order to investigate the economic viability of resource infrastructure abstractions in Grid networks. In the context of this mobility action, Miss Rosy AOUN, PhD student at GET, has been hosted by UEssex from 04/04/2009 to 05/05/2009. Similarly, Miss Chinwe Abosi from UEssex has been hosted at Telecom ParisTech during the month of June 2009.

The joint collaboration is currently focusing on:

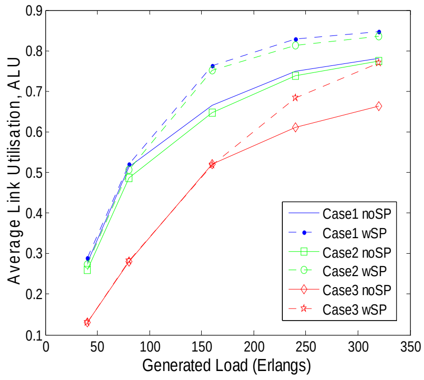
- Study and comprehension of infrastructure abstractions and economic approach in resource virtualization and job scheduling in optical-grid environments. For each approach, investigation of the programming tools used is carried out.
- Impact of the infrastructure abstractions on the proposed ENST economic cost algorithm. In particular, ENST algorithm is extended to include ESSEX job model
- Study and investigation of the accuracy (in terms of accepted demands) and viability of resource infrastructure abstractions to clients, service and resource providers.

2.5.2.3 Service-oriented Resource Provisioning through Resource Co-selection, Co-Reservation and Co-allocation

An investigation of the introduction of the service plane as a uniform, integrated platform for provisioning of heterogeneous resources was performed. The service plane exploits the concept of resource co-selection and co-reservation mechanisms.

We first investigated the impact of the service plane over two types of underlying network granularity: (I) fine, where bandwidth can be provisioned at a low granularity as long as capacity exists within the link. (II) coarse – a single request is provisioned over a full wavelength.

Figure 24 measures the performance of the service plane in terms of a) average link utilisation b) success ratio; as traffic load characteristics transform. The figure shows that although fine granularity allocation performs better overall, the introduction of a service plane improves the performance of both cases. This shows that our service plane can results in overall performance regardless of the restrictions placed by the granularity of the underlying network allocation.



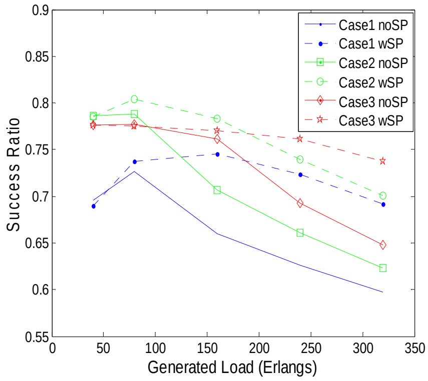


Figure 24: Network performance (average link utilization (a) and success ratio (b)) with and without service plane (SP) for different levels of network granularity

The impact of the service plane for future Internet applications was also investigated. Future Internet applications are typically distributed and place a high demand of bandwidth and IT resources. Results in Figure 24 show that regardless of the load contribution of the Internet traffic, the introduction of the service plane achieves an overall increase of the average resource utilization for resource owners as well as increase of the demand acceptance for the users, when compared to the traditional two-step provisioning approaches. The results show that resource co-selection and co-reservation mechanism presented which forms a central part of the service plane improves the performance of service provisioning for future Internet applications.

2.5.2.4 *The Effect of the Position of the Service Plane in Relation to the Infrastructure Provider*

The performance of the service plane were investigated under two conditions: (a) the service plane is the same entity as the resource provider and (b) the service plane is a third party entity. Three scenarios were identified: Case I: integrated; Case II: trusted-separate and Case III: separate entities.

The performance trade-off faced by infrastructure providers and users under fully dynamic conditions was studied. The impact of the service plane as an integrated entity vs. trusted or separate entity is evaluated by measuring the success ratio (i.e., the fraction of accepted requests) and the uncertainty factor (i.e., fraction of requests accepted by service provider but rejected by infrastructure provider). Performance results on the success rate and uncertainty factor are illustrated in Figure 25 and Figure 26, respectively. The results indicate that a third-party service plane entity can perform comparably to an integrated system with an effective update policy.

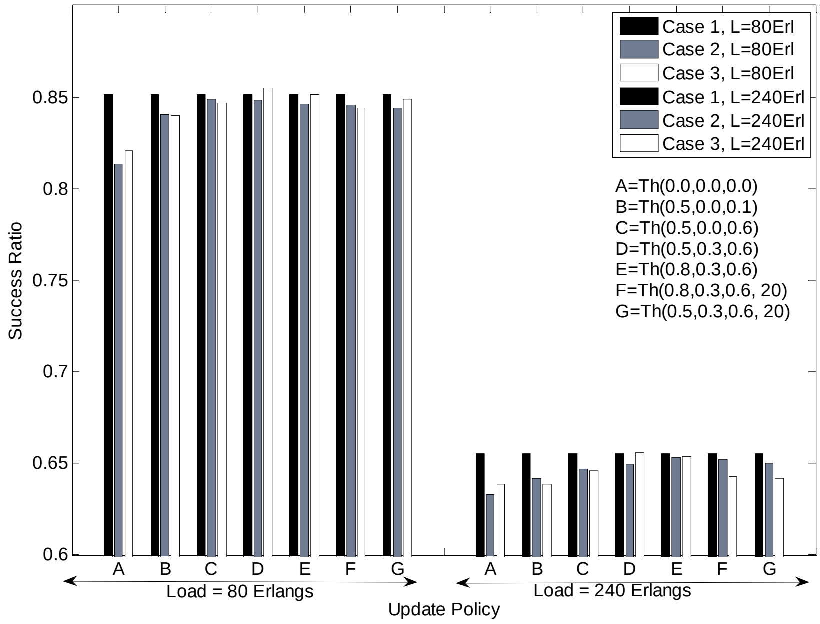


Figure 25: Success ratio

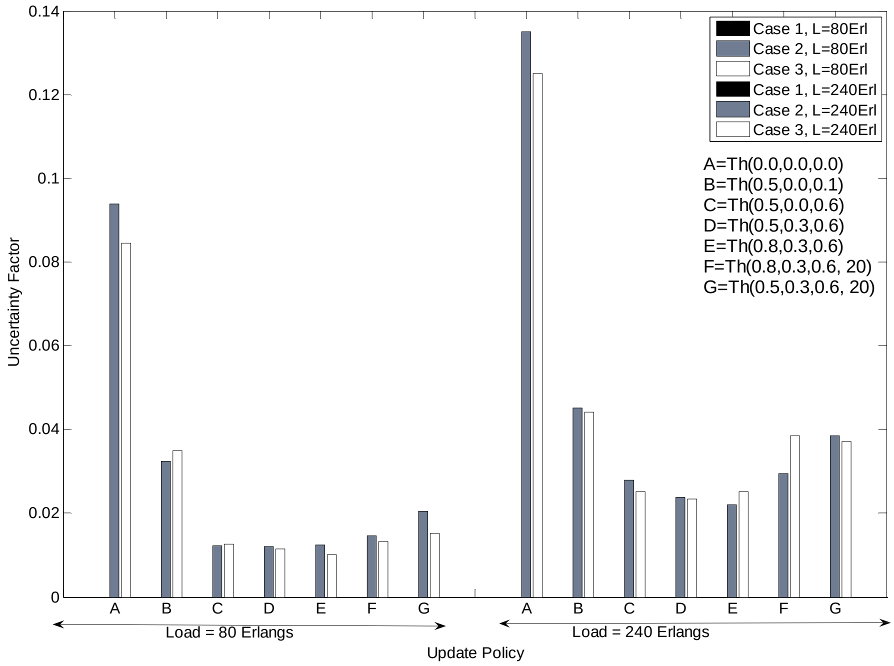


Figure 26: Uncertainty Factor

2.5.2.5 *Definition of a Mapping framework between Connectivity Requests Issued by Applications in Terms of end-host Addresses and Perceived QoS Parameters and Technology-dependent Network Directives*

The testbed shown in Figure 27 has been built to validate the proposed architecture, and to evaluate the performance of the distributed/centralized service element (DSE/CSE) prototype, developed by the authors, during the delivering of a VPN. Virtual Private LAN Service (VPLS) and Border Gateway Protocol (BGP)/MPLS IP VPN are respectively the L2 and L3 MPLS-based VPN considered by the authors within the scope of this work. VPLS is a L2 VPN technique that emulates the full functionality of a traditional multipoint Ethernet LAN by interconnecting several LAN segments across a Packet Switched Network (PSN). In VPLS, PEs emulates bridges performing MAC learning and Ethernet frame forwarding. Since no route is exchanged between the provider edge (PE) and the customer edge (CE), VPLS allows a complete separation between the provider and the customer networks. BGP/MPLS IP VPN is a L3 VPN technique that provides a private multipoint IP-based connectivity across an IP/MPLS network. In this case, PEs need to exchange routes with CEs to acquire reachability information about the customer networks involved in a VPN.

Both VPLS and BGP/MPLS IP VPN require specific mechanisms to assure different VPN traffic distinction to discover the PEs involved in the VPN provisioning and advertise the involved PEs about their reciprocal identities, to create tunnels among the PEs.

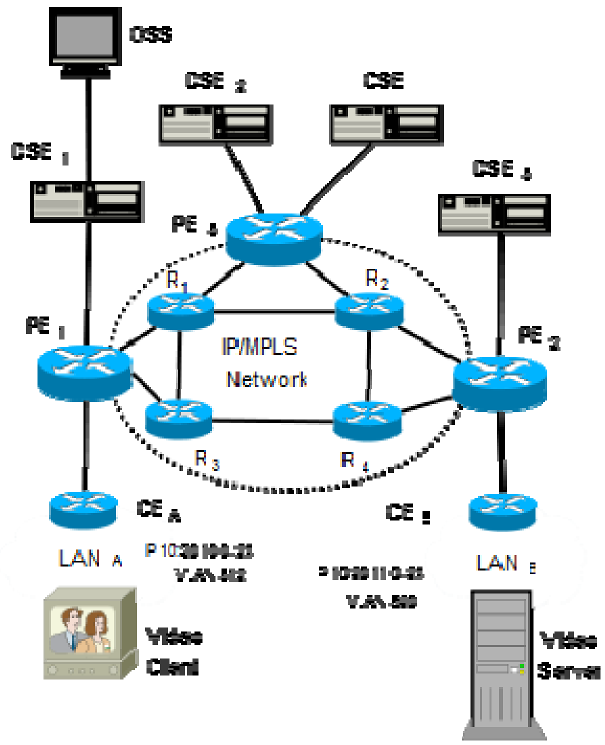


Figure 27: Testbed with distributed and centralized service elements

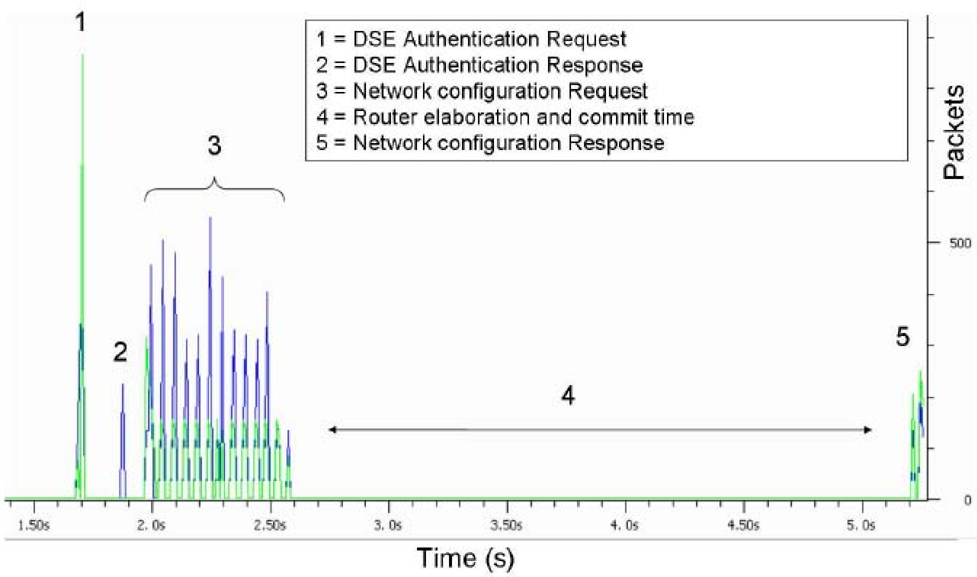


Figure 28: Captured signaling packets between DSE and PE

The principal result is that the time needed to provide a VPN is independent from the VPN type. In fact, in commercial routers the time needed to setup a VPLS or a BGP/MPLS IP VPN are comparable since these two VPNs present similar configuration operations. The overall time needed for the provisioning of a VPN is the sum of the elaboration and transmission times presented in Table 4 and Table 5 (approximately 5.5 sec). We observed that the Video Client starts to receive video traffic 10 sec after the last router ACK independently of the type of VPN provided. This time is substantially due to the time needed by the PEs for processing the set of configuration commands. In particular, the most time-consuming router action is the exchange of the routing tables among the PEs. Since a BGP instance was previously enabled within the PEs before the beginning of the test, the time necessary to setup a BGP signaling instance among the PEs is not considered in this work. The DSEs configure the PEs in parallel, i.e. approximately at the same time, thus the number of PEs involved in the VPN does not affect the VPN provisioning time.

Average Time (sec)	**Src/Dst**	**Message Name**	**Msg. size (Kbyte)**
0.1	OSS/DSE	VPN Client Req	2.6
0.1	DSE/DSE	Service Correlation Req	2.7
0.7	DSE/PE	Network Configuration Req	3.6
2.6	PE/DSE	Network Configuration Resp	0.2
0.1	DSE/DSE	Service Correlation Resp	0.2
0.1	DSE/OSS	VPN Client Resp	1.8

Table 4: Transmission time of the different signaling messages

Average Time (sec)	Task description
0.4	Creation of the Master F SM
0.4	Elaboration of the Service Correlation request for the other DSEs
0.5	Elaboration of the Network Configuration response from PE
0.4	Elaboration of the Service Correlation response to OSS

Table 5: Elaboration time of the different signaling messages

2.5.3 Published Joint Papers (joint and single partner)

1) R. Aoun, M. Gagnaire, "Service differentiation based on flexible time constraints in market-oriented grids", IEEE Globecom conference, Honnolulu-USA, November 30-December 4, 1999.
2) R. Aoun, and M. Gagnaire, "An exact optimization tool for market-oriented grid middleware", in Proceedings of the 14th international IEEE Communication Quality & Reliability (CQR) Workshop, Florida, USA, May 2009.
3) R. Aoun, and M. Gagnaire, "Towards a fairer benefit distribution in grid environments", in Proceedings of the first international workshop on grid Computing (GRIDCOM), Rabat, Morocco, May 2009.
4) C. E. Abosi, R. Nejabati and D. Simeonidou, "A Service-Oriented Resource Provisioning for Future Optical Internet Applications" submitted to ICC 2010.
5) C. E. Abosi, R. Nejabati and D. Simeonidou, "Performance Evaluation of a Novel Service Provisioning Mechanism for Future Optical Internet Infrastructure" submitted to OFC 2010.
6) F. Baroncelli, B. Martini, V. Martini, P. Castoldi "A cooperative approach for the automatic configuration of MPLS-based VPNs", International Journal of Grid Computing and Multi Agent Systems (GCMAS), to be published.

2.5.4 Planned Activities for Y3

The following activities are planned for year 3:

1) Enrich our economical models for Grid services in considering a negotiation between end-users and service providers.
2) Adapt market-oriented grids specifications to the abstraction model. Through study of its impact on service providers who now have the choice of buying a guaranteed abstracted topology, or an over-dimensioned abstracted topology, or a physical topology.
3) Enhance the service allocation algorithms to optimize resource selection and allocation.

4) Investigate the implications of the service provisioning platform on underlying network and IT resource virtualization.
5) Enlarge the set of network services provisioned to applications to include requirements from Service Delivery Platform and allow the on-demand establishment of channels for transfer of media streams such as bidirectional channel, compound channel supporting contemporary video, voice, chat sessions.

2.5.4.1 Planned publications

R. Aoun, E. Doumith, M. Gagnaire, C. Abosi, R. Nejabati, D. Simeonidou, “An exact approach to an abstracted topology design in WDM Networks” , to be submitted to IEEE Transactions on Services Computing (TSC) journal.

R. Aoun, and M. Gagnaire, “Grid Service Provider - Network Operator Relationship: Impact on Grid Economy”, to be submitted to the IEEE Journal of Lightwave Technology (JLT).

R. Aoun, and M. Gagnaire, “Fair economical models based on sliding time window mechanisms for grid environments”, to be submitted to the Journal of Future Generation Computer Systems, Elsevier Editor.

C. Abosi, R. Nejabati, D. Simeonidou, R. Aoun, E. Doumith, M. Gagnaire, “ An efficient heuristic approximation for abstracted topology design in Service-Oriented Optical Networks”, to be submitted to IEEE/OSA Journal of Optical Network and Communication journal

2.5.4.2 Targeted Publications

IEEE Transactions on Service Computing, Globecom 2011, ICTON 2010, Journal of Optical Communication and Networks

2.6 JA6 - UNI Extensions for Service Oriented Optical Networks

Participants: UESSEX, RACTI, AIT

Responsible person: Eduard Escalona (UEssex)

2.6.1 Objectives for Y2

This Joint Activity aims to define a set of interoperable procedures between the service and network layers to allow a seamless integration of next generation services and optical networks. This should be achieved by building a Service Oriented architecture identifying the requirements imposed by the service layer and specifying the functionalities that should be supported in the interface between layers such as resource discovery, characterization, allocation, and management services.

The objectives for Y2 include the definition of the architecture to support the proposed advanced functionalities and the identification of use-cases to help on the development of the interface. This work was carried out in the context of the OGF Network Service Interface (NSI) working group.

2.6.2 Achieved Results in Y2

The achieved results are mainly related to the participation and contributions to the OGF NSI working group, whose main objective is to define and implement a recommendation for a network service interface to support advanced network connectivity requests from heterogeneous network users and providers. This recommendation describes the architectural context for use of the interface. The components described in the architecture include the transport resources, the Network Service Agent (NSA) controlling these resources, the Requestor Agent (RA) and finally the Network Service Interface (NSI).

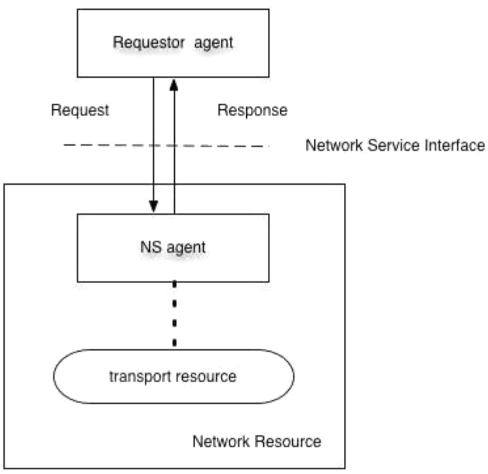


Figure 29: : NSI architecture

The term *Network Resource* refers to a transport resource with an agent that instantiates, schedules, tracks and announces its resources over a service oriented interface. The transport services supported by the NSI are basically connections between ports.

The NSA is an agent which has control over the Transport Resources. The control functions supported by this NSA may include resource allocation and scheduling, implementation policy for resource allocation, and communication with other agents. This recommendation describes a method of communication between a requestor and a network agent.

The NSI recommendation proposes a new paradigm for creating connection oriented services in which a set of network resources interconnect at both data and control level. This concept allows services to directly request network connections from one or more network resources directly. In this way this concept supports both the Grid approach which treats networks as resources and the network approach which treats networks as collaborating entities that provision a service to requestors. These approaches have been unified in the NSI recommendation, allowing network resource infrastructure to be deployed and configured in a variety of ways.

Transport resources, supported by different agents, may be represented in the form of a transport topology. This topology describes transport elements and their interconnections and can be used to compute possible end-to-end connections. The concept of a service plane is introduced to describe the mesh of Agents interconnected by NSI interfaces. The recommendation describes a number of service plane infrastructure architectures that might be deployed.

In this architecture a Network Resource is an atomic element. That is, the agent and the transport resource combine to form a single entity known as the Network Resource. The way in which services are provided by the transport resource is not mandated in this recommendation. The ability to create connections may also require the interoperation between network agents which may be facilitated by some sort of federated inter-agent trust configuration.

It is expected that the NSI interface will be implemented using Service-Oriented Applications (SOA) and web services messaging. In addition, the architecture is assumed to be implemented by RSVP messaging where appropriate.

The goal of the recommendation is to provide a mechanism for requesting connection oriented networking. Such deployment will depend on additional services such as path computation, attribute authorization, topology sharing, and scheduling. The NSI architecture will be supported by an abstract message definition.

Apart from the architecture, the work has also focused on the preparation of the use-case deliverable also in for the NSI WG. The deliverable has been created using the input received from a questionnaire sent to different users, applications and service providers that could benefit from a standardized interface.

2.6.3 Published Joint Papers (joint and single partner)

Unofficial draft version of OGF NSI user-case deliverable (work in progress).

2.6.4 Planned Activities for Y3

The remainder of the work to be carried out in this activity will define the abstract messaging to support an efficient communication between layers and the identification and study of the most suitable protocol architecture to be considered as a possible implementation.

2.6.4.1 Targeted Publications

IEEE Communications Magazine

2.7 JA7 - Photonic Grid Dimensioning and Resilience

Participants: IBBT, AIT, RACTI, AGH, KTH

Responsible person: Chris Develder (IBBT)

2.7.1 Objectives for Y2

The overall objectives of this joint activity are to address these fundamental and challenging research questions in such an optical Grid supporting anycast services:

• Dimensioning: Network dimensioning algorithms capable of optimizing network topology, resource capacity and resource placement will be designed and evaluated. Processing and storage resources dimensioning strategies will be also examined. In judging the relative optimality of the different approaches, simulations will be used. Various optical switching paradigms can be considered.

• Resilience schemes: To study and propose algorithms and strategies for providing reliable Grid operation. Based on a general analysis to provide resiliency in optical Grid environments, we will define protection and restoration strategies, and present specific case studies.

Outcome of the joint research activity: The major outcome of the joint work will be novel dimensioning and resilience strategies, which we aim to disseminate via publications (both in journals and international conferences). The complementarity of the partners involved will allow for a deeper insight in all issues involved (ranging from scheduling algorithms and networking protocols to physical layer issues) than possibly achievable by each of the groups by themselves.

Where the dimensioning approaches have been the main focus in Y1 (see BONE deliverable D21.1), the second year activities pertained mainly to resilience studies. These are summarized below.

2.7.2 Achieved results in Y2:

The results are summarized in the following studies, which are discussed in more detailed in the following subsections:

- Exploiting relocation to reduce network dimensions of resilient optical Grids.
- The multicost approach for path and link protection
- Differentiated BER ICBR & Resilience
- Resilience differentiation for OBS-based grid networks

2.7.2.1 Exploiting relocation to reduce network dimensions for resilient optical Grids

We propose to exploit the anycast principle in Grid networking to reduce the network capacity needed to provide network resiliency in Grids. In contrast to classical network protection schemes, we will not necessarily provide a back-up path between the source and the original destination. Instead, we will try to relocate the job to another server location if this means that we can provide a backup path which comprises fewer hops (and hence fewer reserved wavelengths) than the one the traditional scheme would suggest. This relocation is possible because of the Grid specific anycast principle: a user generally does not care where his job is executed and is only interested in its results. We present ILP formulations for both resilience schemes and we evaluate them in a case study on a European network topology. This principle is illustrated in Fig. 1.

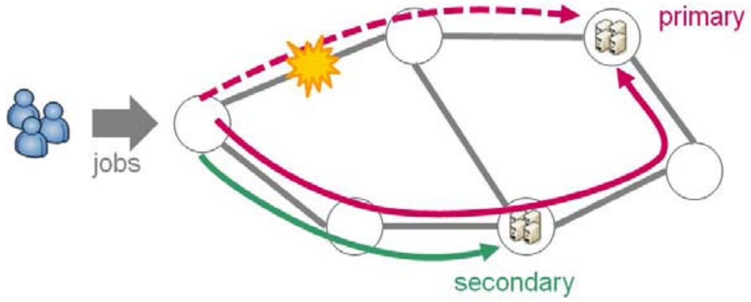


Fig. 1 – Illustration of the relocation principle: instead of routing jobs to the original (primary) destination site via a long backup path, in case of a network failure, we send them to another site that is more nearby.

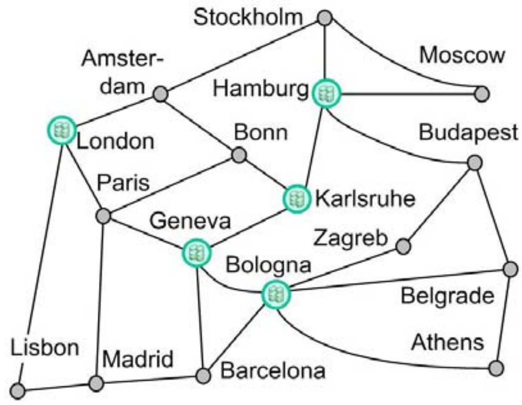


Fig. 2 – The European network topology used in a case study [Develder2009].

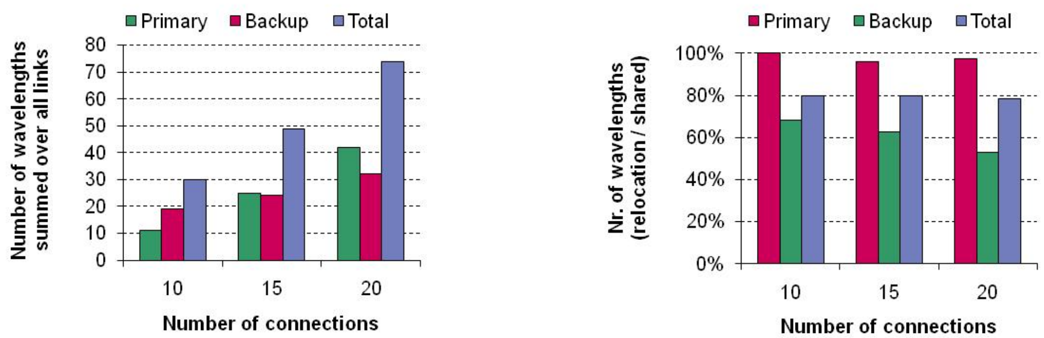


Fig. 3 – Sample results (from [Develder2009]) showing the number of wavelengths required, summed over all links: (a) in absolute values for the classical shared protection case, (b) relative values for the relocation case (in % of the wavelengths for classical shared protection). By exploiting relocation we save about 20%.

To investigate the possible savings, we considered a network dimensioning study for shared path protection schemes:

Given:

- Network topology, in terms of nodes and links (see e.g. Fig. 2)
- Candidate Grid server locations (e.g. indicated with green circles in Fig. 2)
- Job arrivals at each of the sites (i.e. the network nodes), expressed as a total of required number of wavelength paths.

Find:

- The (minimal) number of wavelengths required on each of the links to carry the traffic

Such that:

- The network can survive single link failures

We formulated this problem as an Integer Linear Program (ILP), and considered both the classical shared path protection case and a relocation scheme, as explained above. On European network topologies (see e.g. Fig. 2), we performed a couple of quantitative case studies. Sample results are shown in Fig. 3. From these studies, we concluded that exploiting relocation can achieve around 20% reduction of total wavelength capacity in the network.

For the ILP formulations, and detailed discussion of results, we refer to [Develder2009] and [Buysse2009].

2.7.2.2 The multicost approach for path and link protection

Our focus is on protection schemes that can be implemented by extending the multicost approach and used for protecting paths and links connecting Grid resources.

In the multicost approach a vector of cost parameters, which are related to different quality of service (QoS) parameters is assigned to each link. Then, by defining appropriate operations on these link cost parameters, we can calculate the cost vector of a path. We are mainly interested in showing that through extensions to the multicost approach, protection can be easily provided and also examine the effects (e.g., in terms of network blocking) of its application. Moreover, the multicost approach offers inherently quality of service (QoS) differentiation features, since it is possible to apply different optimization functions to select a lightpath from the set of candidate lightpaths, so as to serve differently the connection requests belonging to different users. Similarly, different approaches can be used for selecting the primary and the backup lightpath(s) of different connections, e.g., by applying different optimization functions so as to provide a different level of protection to users. First we describe the multicost approach without the protection scheme and then we extend the multicost algorithm to provide protection capabilities.

The multicost approach

We assume that we have a WDM network with N nodes and L links, each of which uses m wavelengths. The WDM network employs no wavelength conversion. We assume that the node where the algorithm is executed has knowledge of the physical topology of the network.

The multicost routing and wavelength assignment algorithm we propose consists of two phases: given a source-destination pair (S and D, respectively), the set P_{n-d} of non-dominated paths between them is calculated first, and then an optimization function $f(V_p)$ is applied to the cost vector of each path $p \in P_{n-d}$ to select the optimal one.

Phase 1: Algorithm for computing the set of non-dominated paths P_{n-d}

The algorithm that computes the non-dominated paths from a given source to all network nodes (including the destination D) can be viewed as a generalization of Dijkstra's algorithm that only considers scalar link costs. The basic difference with Dijkstra's algorithm is that instead of a single path, a set of non-dominated paths between the origin and each node is obtained. Thus a node for which one path has already been found is not finalized (as in the Dijkstra case) since we can find more "non-dominated" paths to that node later.

Phase 2: Choosing the optimal path and wavelength from the set P_{n-d}

In the second phase of the proposed algorithm we apply an optimization function $f(V_p)$ to the cost vector, V_p, of each path $p \in P_{n-d}$. The function f yields a scalar cost per path and wavelength in order to select the optimal one. The function f can be different for different tasks, depending on their QoS requirements. Note that the optimization function f applied to the cost vector of a p has to be monotonic in each of the cost components. For example, it is natural to assume that it is increasing with respect to delay, etc. The final step is to choose from P_{n-d} the path and wavelength that minimizes $f(V_p)$.

Multicost algorithm enabled with path and link protection

The multicost fault-tolerance scheme to be discussed can provide both *link* protection and *path* protection.

Link protection of a link of the primary lightpath can be provided using two additional operations. Following the creation of the set of candidate lightpaths, these are grouped into subsets of lightpaths that do not use a specific link and as a result they offer protection against a probable failure of that specific link. Next, using these subsets we can find for a selected primary lightpath, the backup lightpaths that protect against failures of specific links of the primary lightpath. In the link protection fault tolerance mechanism, there are several policies that can be used when choosing the primary and the backup lightpath(s). These policies may differ not only in the exact optimization functions applied for the actual selection, but also on the number of the links protected, how these links are selected, and other factors.

In the path protection mechanism, from the set of candidate lightpaths a primary lightpath is selected to serve the connection request, while a second link- (or node-) disjoint lightpath, from the same set is selected as backup, so as to serve the connection in case of link or node failures in the primary. As a result, in order to provide dedicated path protection, the multicost schemes must be augmented by two operations. Initially, the set of candidate lightpaths calculated by the first phase of the multicost algorithm must be searched so as to find disjoint *pairs* of lightpaths. One of these pairs is then selected according to some optimization function or policy to form the primary and the backup lightpaths. It is evident, that more than one backup lightpath can also be selected.

We have performed simulation experiments **(**Figure 30**)** evaluating different optimization policies (Optim1, Optim2) for selecting the primary and the backup lightpaths, measuring the number of blocked connections due to limited number of wavelengths. We should note that a connection request is blocked when no primary or (disjoint) backup lightpath can be found. In our simulations, we observed the importance of the selected optimization policy to the performance of the algorithm. In addition, the number of disjoint paths found for each primary lightpath depends on the number of candidate lightpaths produced by the first phase of the multicost approach.

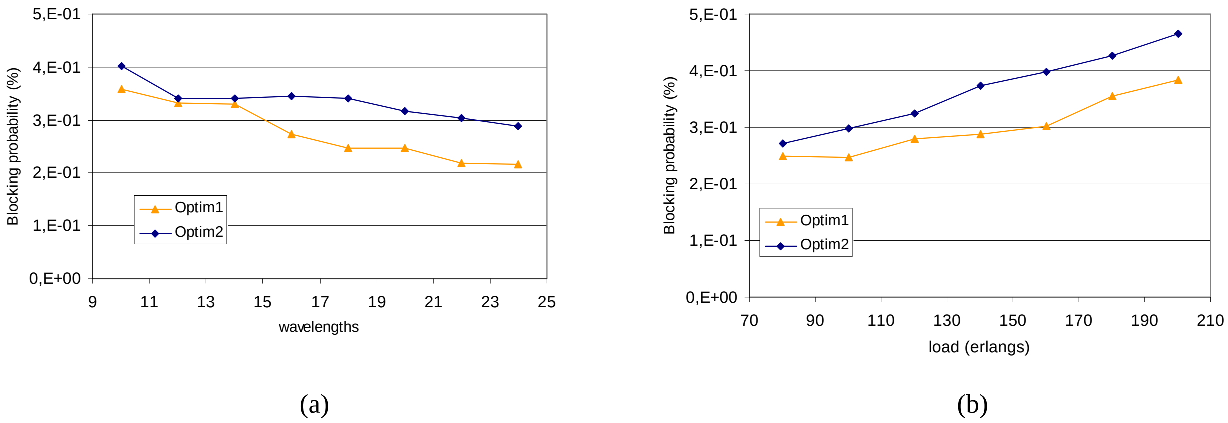


Figure 30 Sample results (from ongoing work) showing the blocking probability as a function of (a) the number of available wavelengths for fixed load, (b) the load in the network for a fixed number of wavelengths (W=20 wavelengths), for different optimization policies of the multicost approach.

2.7.2.3 Resilient provisioning of BER-level differentiated services in optical Grids

Optical Grid applications pose variable Quality of Service (QoS) requirements that need to be fulfilled to ensure liveliness, dependability and user satisfaction (in case of user-to-machine applications). This joint activity applies service differentiation at the physical layer, namely in maximum Bit Error Rate that can be tolerated by a lightpath, both to working and protection resources and evaluates the benefit of such an approach in terms of improved utilization of resources, compared to flat (non-differentiated) service provisioning. Although our approach is generally applicable to connection provisioning in WDM networks, the case of interest for the proposed approach in the context of optical Grid applications is obvious, namely having applications that have stringent BER requirements (e.g. real-time applications such as 3D-imaging or weather visualization applications) and elastic-traffic applications (e.g. bulk data transfers as part of distributed computations), all operated over the same WDM optical network.

In contrast to conventional Impairment-Constrained Based Routing (ICBR), where a single (flat) BER threshold is applied to check the sanity of working/protection resources, our scheme proposes the use of multiple BER thresholds, matching the minimum variable requirements of applications in terms of bit error rate. Unlike conventional ICBR, where a connection is blocked, if its computed BER exceeds the BER threshold, our scheme associates a BER threshold with each incoming grid application request. By doing so, and assuming for the sake of simplicity two BER classes of service with thresholds 10^{-15} and 10^{-9} respectively, a connection will be assigned a lightpath (if available) that satisfies its minimum requirements. As such, if a request corresponding to an elastic Grid application that can tolerate a BER of 10^{-9} can be served via alternative lightpaths that exhibit varying error rates, the lightpath with maximum BER will be picked among the alternatives, both for provisioning the working as well as the protection path. This avoids unjustified waste of resources and improves resource utilization, as shown in the results presented below.

To quantify the improvement brought by our approach, we used online simulation of connection requests over the Pan European network topology (shown in Fig. 4). Connection requests are randomly generated following a Poisson distribution and are served sequentially. The source/destination pair of each request is randomly chosen with uniform probability and the connection holding time is exponentially distributed. Each simulation experiment runs until the 90%-confidence level becomes less than 10% of the sampled mean. We consider two classes of connection requests with regards to signal quality requirement, namely requesting BER less than 10^{-15} (Class-1) and 10^{-9} (Class-2). We consider two cases regarding the compensation of Class-1 vs. Class-2 requests to the overall traffic load: a) 30%-70% and b)50%-50%. A Class-1 request is blocked, if there is no lightpath connecting the two endpoints of the request that exhibits BER less than 10^{-15}; whereas a Class-2 request is blocked if there is no lightpath with BER less than 10^{-9}. Moreover, in our algorithm the lightpath with the maximum BER that satisfies the signal quality requirement of the request is selected. This scheme is compared against the following conventional approaches:

- Shortest-path routing in routing primary/protection paths.
- Non-differentiated ICBR with selection of the Best path (lowest BER) among alternative paths for a single requests (IABP)
- Non-differentiated ICBR with selection of the worst path (maximum BER) among alternative paths for a single requests (ICBR Max BER)

In fact, in evaluating all conventional approaches connection requests of either Class-1 or Class-2 are blocked if there is no route with BER less than 10^{-15}. Note that we use First-fit for allocation of wavelengths to candidate lightpaths, while wavelength conversion is not an option.

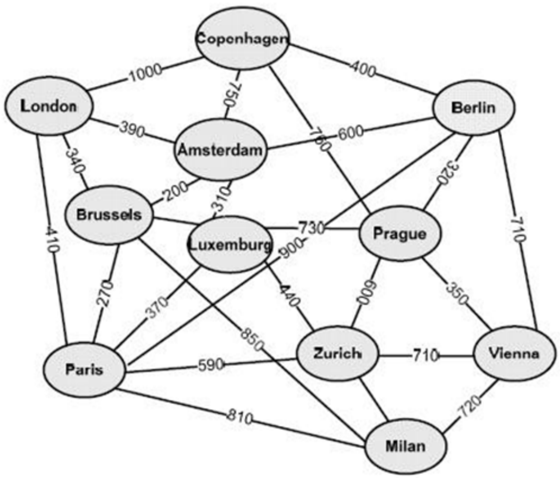


Fig. 5 – Pan European test network topology (COST 239)

Fig. 6 presents the evaluation results that exhibit the superior utilization of network resources achieved by our approach compared to the rest of the routing methods, both with and without support for sharing of protection resources by distinct connections. Further details on the specification of the proposed algorithm and the interpretation of results can be found in [Jiratti2009].

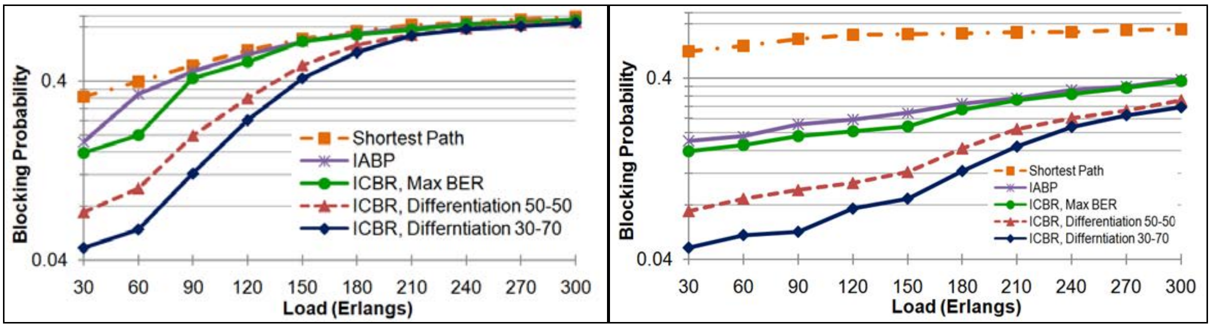


Fig. 7 – Blocking probability with BER-level service differentiation compared to conventional approaches without sharing (left) of protection resources and with sharing (right).

2.7.2.4 Resilience differentiation for OBS-based grid networks

OBS-based grid networking

AGH performed some studies focused on reliability differentiation in OBS networks. Obtained results can be adopted to OBS-based grid network characteristics. By this term we consider possibility of using different resilience actions after failure in order to provide specific/dedicated quality to a few classes of traffic. The exact number of considered traffic classes concerning resilience and also used resilience procedures are for further studies.

Studies were performed with the use of a modified version of a E-OBS simulation tool developed originally at the UPC in Barcelona. The simulator is a tool to study a variant of OBS concept called Offset-Time Emulated OBS (E-OBS). The investigations were stimulated by one mobility action held during 2008 at UPC.

Current trends in core network design are focused on Optical Packet Switching (OPS) as a promising long-term solution for future core networks. OPS is a technological approach where user information is transferred in packets being kept only in optical domain inside the network. Advantages of such technology include ability to handle bursty traffic, adaptability and effective resource utilization. Unfortunately current technology state does not make it possible to analyze signals in optical domain. What is more, optical buffering possibilities are very limited. That is why so far OPS networks can’t be implemented.

OBS networks are currently seen as intermediate phase in networking towards OPS. In OBS ingress node is performing burst assembling. OBS brings as near to OPS as possible considering actual technology state. In this type of networks, clients’ data are transmitted in optical domain without any O/E/O conversion. Carried investigations were focused on resilience feature and they restored end-to-end connectivity (p-2-p traffic). Further step will be to investigate also point-to-anycast feature of grid networks.

Data incoming from any type of network are aggregated at the edge of the OBS domain into variable size data units called bursts. When the burst is ready a Burst Control Packet (BCP) is sent using signaling channel (one of the wavelengths in fiber) to reserve resources for incoming burst. BCP undergoes O/E/O conversion in each node on the way in order to be analyzed and set switching matrix for incoming burst. The burst is sent after the Offset time, when the resources should already be reserved. It can be seen from above description that OBS paradigm fits well the target functionality of grid networking because entire grid task could be sent to its destination as a single burst, without fragmentation.

Resilience differentiation for OBS-based grid networks

In OBS networks the path for a burst is likely to be determined on the hop-by-hop basis at every intermediate node, and the established connections will be relatively short. That is why — as many studies conclude — the restoration mechanisms would perform better than protection schemes in this type of networks.

That is why our studies focused on two different restoration mechanisms:

- *Local deflection* works in the following manner: The node directly connected to a faulty link is responsible for rerouting all traffic that is supposed to take a faulty link as a next hop. The important advantage of such a solution is it's short reaction time. The restoration mechanisms can be in operation immediately after the failure detection. On the other hand the drawback of such a mechanism is the fact, that it may cause a congestion of links directly connected to the node (focused congestion), which is responsible for rerouting, and therefore result in excessive data dropping or delays.
- *Global Routing Table Update*: This mechanism provides more efficient traffic distribution by informing all nodes inside the network about the failure and forcing them to recalculate routing tables. Very significant drawback of such a solution is a relatively long time, which is required to propagate failure notification throughout the entire network and to recalculate routing tables.

We performed our simulation based on an existing topology of NSFNet network presented in Fig. 8

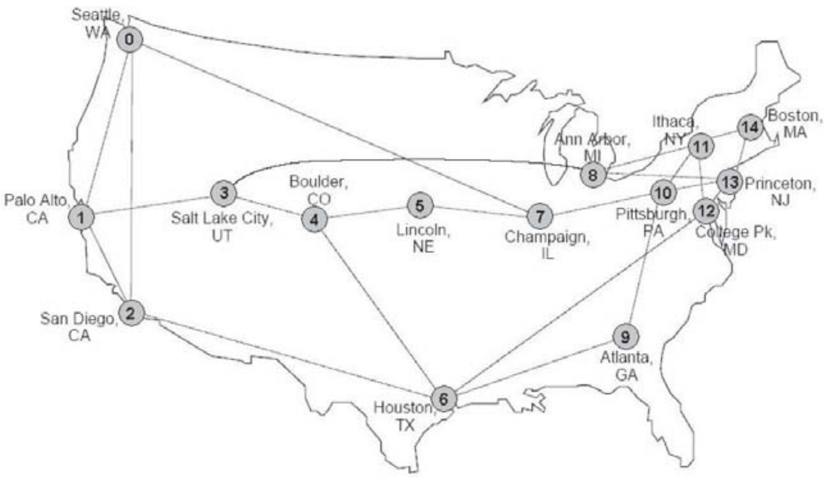


Fig. 8 – NSFNet topology

The obtained results are presented in Fig. 9 and compared with a scenario where network failure occurs.

The main criterion for scenario evaluation was the Burst Loss Ratio achieved under different network loads. As clearly shown the Global Routing Table Update (GRU) in comparison with a Local Deflection (LD) decreases significantly data loss ratio by providing more efficient traffic distribution over the network. The reason is that in LD scenario links directly connected to the node that performs deflection are more likely to get congested (local effect), while the GRU mechanism causes traffic (which was supposed to pass through failed link) more uniformly distributed among network links. During each simulation run 1 000 000 bursts were sent in the E-OBS network. Each link had 32 wavelengths with data rate 10Gbps. The delay introduced in each node was set to 10µs, no FDL was used.

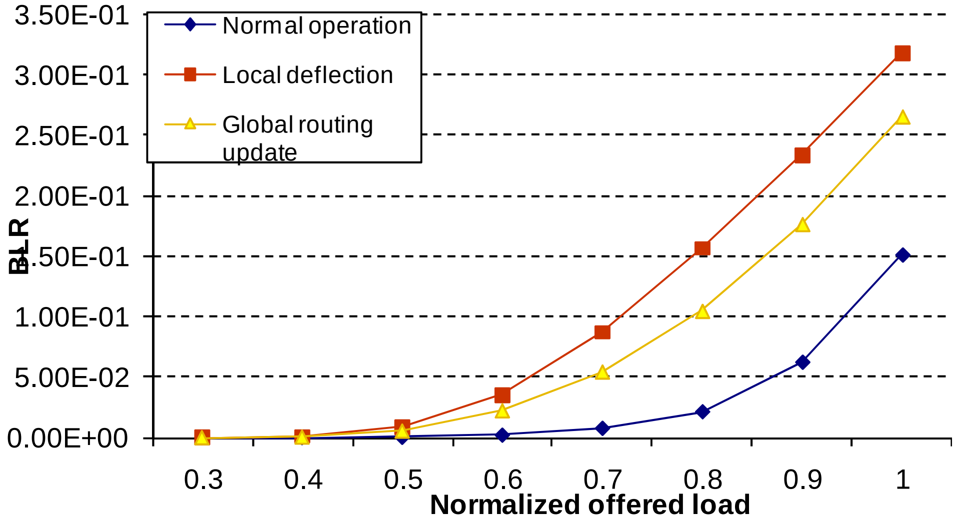


Fig. 9 – NSFNet topology with a Bypass Routing (BP) for three investigated scenarios.

2.7.3 Published Joint Papers (joint and single partner)

The **joint papers** are indicated with a "*".

***[Jiratti2009]** A. Jirattigalachote, L. Wosinska, P. Monti, K. Katrinis and A. Tzanakaki, *"Impairment Constraint Based Routing (ICBR) with Service Differentiation in Survivable WDM Networks"*, in Proceedings of the 35th European Conference and Exhibition on Optical Communication (ECOC 2009), Vienna, Austria, September 2009. (AIT+KTH)

[Varvarigos2008] E. Varvarigos, V. Sourlas, K. Christodoulopoulos, "*Routing and Scheduling Connections in Networks that Support Advance reservations*", Computer Networks (58), 2008. (RACTI)

[Develder2009b] C. Develder, B. Dhoedt, B. Mukherjee, P. Demeester, *"On dimensioning optical grids and the impact of scheduling,"* Photonic Netw. Commun., vol. 17, no. 3, pp. 255–265, Jun. 2009. (IBBT)

***[Stevens2009]** T. Stevens, M. De Leenheer, C. Develder, B. Dhoedt, K. Christodoulopoulos, P. Kokkinos, and E. Varvarigos, *"Multi-cost job routing and scheduling in grid networks,"* Futur. Gener. Comp. Syst., vol. 25, no. 8, pp. 912–925, Sep. 2009. (IBBT+RACTI)

[Doulamis2008] N. Doulamis, P. Kokkinos, E. A. Varvarigos, *"Spectral Clustering Scheduling Techniques for Tasks with Strict QoS Requirements"*, Proc. 14th Int. Conf. on

Parallel and Distributed Computing (Euro-Par 2008), Las Palmas de Gran Canaria, Spain, 26-29 Aug. 2008, pp. 478-488. (RACTI)

[Kokkinos2008] P. Kokkinos, K. Christodoulopoulos, A. Kretsis, E. A. Varvarigos, *"Data Consolidation: A Task Scheduling and Data Migration Technique for Grid Networks"*, Proc. 8th IEEE Int. Symp. on Cluster Computing and the Grid (CCGRID 2008), 19-22 May 2008, pp. 722-727. (RACTI)

[Markidis2008] G. Markidis and A. Tzanakaki, *"Network Performance Improvement through Differentiated Survivability Services in WDM networks"*, J. of Optical Networking, June 2008, vol. 7, no 6, pp. 564-572.. (AIT)

2.7.4 Planned Activities for Y3:

The JA will wrap up its activities in the final project year, by elaborating further the resiliency issues in the following complementary studies:

- IBBT: The resilient grid dimensioning study will be extended to larger case studies, for which we will develop a heuristic approach (in addition to the ILP methodology studied previously). Also, where the resiliency scheme (exploiting relocation) studied so far mainly focused on protecting against network (link) failures, we will expand our focus to resiliency strategies also protecting against resource failures. Therefore, we will also include resource dimensioning (as opposed to previous evaluations focusing on minimizing network capacity only).
- RACTI: We will examine various parameters affecting the quality of transmission of the selected primary and backup lightpaths.
- AIT & IBBT: Assess potential advantage of considering different resilience schemes for different tasks belonging to a workflow
- AGH plans to further develop OBS simulator, to test resilience differentiation methods and to study control plane issues for supporting enhanced resilience in OBS-based grids. Besides using reported OBS simulator, AGH is working towards implementation of GridSim tool.

Targeted publications

- Journal publication on exploiting relocation in providing Grid resiliency (IBBT)
- Journal and conference publications of the achieved results in path and link protection in optical Grids (RACTI)
- Journal submission on the joint ICBR resilience work taking into consideration differentiated grid services. (AIT+KTH)
- Conference and/or journal submission on workflow resilience (IBBT+AIT)
- Conference submission on resilience differentiation in OBS-based grid network.

3. Conclusions

This report summarized the status of all joint activities currently operating within VCE-WP12, including a few joint activities that were moved from former WP21 Topical Project on Service Oriented Optical Networks.

Twelve partners were involved in this workpackage and seven joint activities were and are currently running. During the second year, the collaboration between partners increased and helped to set-up three joint testbeds for experimental evaluation of service-oriented platforms for optical networks, based on GMPLS or OBS. Dissemination of the joint activities in international conferences and journals and in the Open Grid Forum standardization body has been actively pursued.

Research progress were achieved in the field of strategy and approaches to improve and guarantee the service availability. In particular, the relocation of services was proposed as a way to guarantee resilience and was able to reduce the capacity requirements of the network by 20%. Also, a multicost approach that support multiple QoS metric was extended to ensure reliability of the services. In addition, the impact of the resilience scheme on the reliability of the services was evaluated in OBS networks. The results indicated that in case of failure a global update of network information for updating forwarding tables is preferable to a local deflection routing. Moreover, an approach to guarantee availability in terms of BER and evaluated for optical grids. Thanks to the proposed approach, not only the BER is guaranteed, but also the blocking probability experienced by the provisioned grid serviced can be drastically reduced. Finally, evaluation of the reliability of wireless services on an integrated wireless and wired network was carried out and showed a tradeoff between reliability and number of handovers experienced by the service.

Algorithms, strategies, and mechanisms for implementing service abstraction and resource virtualization were proposed and investigated. Mechanisms for the co-selection and co-reservation of the resources in the service plane were shown to improve the performance of service provisioning for future Internet applications. Also, the impact of the service plane as an integrated entity vs. trusted or separate entity was evaluated. Also, the configuration of the service layer and network virtual platforms to support cloud computing have been proposed and discussed. Finally, the experimental activity demonstrated the feasibility of the service plane and service-oriented optical networks. During the experimental assessment, the service set-up latency experienced in a service-plane testbed was evaluated and was shown to meet ITU-T recommendations. Finally, PCE-based control plane enabled with functionality for the setup of grid services has been tested on an optical network with commercial equipment and validated the simulative results.

Planned activities for next year aim at consolidating the collaborative effort carried out so far, by further investigating or extending the proposed concepts and approaches. Also, partners of JA4 and JA7 plan to further strengthen the collaboration on the similar research topics.

4. Acronyms

AO-M	Application-Oriented Module
ASN	Access Service Network
ASN-GW	Access Service Network Gateway
B2B	Back-to-Back
BC	Binding Cache
BCH	Burst Control Header
BCP	Burst Control Packet
BER	Bit Error Ratio
BGP	Border Gateway Protocol
BS	Base Station
CE	Customer Edge
CMS	Closest MAP Selection
CN	Correspondent Node
CoA	Care of Address
CSD	Call Set-up Delay
CSE	Centralized Service Element
CSN	Connectivity Service Network
CSPF	Constraint Shortest Path First
DoS	Denial of Service
DSE	Distributed Service Element
DS-TE	DiffServ-aware Traffic Engineering
E-OBS	Offset-Time Emulated OBS
ER	Edge Router
FA	Forward Adjacent
FER	Frame Error Ratio
FMS	Farthest MAP Selection
FPGA	Field Programmable Gate Array
(G)MPLS	(Generalized) Multi-Protocol Label Switched Path
GRACF	Admission Control Function for Grid Services
GRM	Grid Resource Manager
HA	Home Agent
HMIPv6	Hierarchical Mobile IPv6
HMS	Half-Way MAP Selection
HoA	Home Address
IaaS	Infrastructure as a Service
IAT	Inter-Arrival Time
ILP	Integer Linear Programming
ICBR	Impairment Constraint Based Routing
JA	Joint Activity
LCoA	Link Care of Address
LSP	Label Switched Path
LSR	Label Switched Router
NaaS	Network as a Service

NVP	Network Virtualization Platform
MAP	Mobility Anchor Point
MEMS	Micro-Electro-Mechanical System
MG-OXC	Multi-Granular Optical cross-connect
MIPv6	Mobile IPv6
MS	Mobile Station
NE	Network Element
NR-DB	Network Resource Data Base
NRDL	Network Resource Specifications
NRM	Network Resource Manager
NSA	Network Service Agent
NSI	Network Service Interface
OGF	Open Grid Forum
OBS	Optical Burst Switching
OPS	Optical Packet Switching
OBT	Optical Burst Transport
PCE	Path Computation Element
PCEP	Path Computation Element Protocol
PCC	Path Computation Client
PCS	Path Computation Scheduler
PCReq	Path Computation Request
PS	Proxy Server
PSN	Packet Switched Network
QoS	Quality of Service
RA	Requestor Agent
RCoA	Regional Care of Address
SaaA	Software as a Service
SCF	Service Control Functions
SIP	Session Initiated Protocol
SOON	Service Oriented Optical Network
SO-MGON	Service-Oriented Multi-Granular Optical Network
SP	Service Plane
SR	Service Request
SRLG	Shared Risk Link Group
TED	Traffic Engineering Database
UAC	User Agent Client
UAS	User Agent Server
VPLS	Virtual Private LAN Service
WG	Working Group
WP	Work Package
WSON	Wavelength-Switched Optical Network
Y1	First Year
Y2	Second Year