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 /D21.2

Contractual Date of Deliverable: 30/11/08

Actual Date of Delivery: 24/12/08

Title of Deliverable: D21.2: Report on year 1 and updated plan for activities

Work package contributing to the Deliverable: WP21: Topical Project on Service Oriented Optical Network Architectures

Nature of the Deliverable R

Dissemination level of Deliverable PU

Editors: UEssex / Dimitra Simeonidou
IBBT / Mario Pickavet

Authors:

Mario Pickavet, Chris Develder, Marc De Leenheer (IBBT)

Kostas Katrinis, Anna Tzanakaki (AIT)

Panagiotis Kokkinos, Kostas Christodoulopoulos, Kyriakos Vlachos, Emmanouel Varvarigos, Apostolis Siokis (RACTI)

Tanya Politi, Alexandros Stavdas (UoP)

Guido Alejandro Gavilanes Castillo, Fabio Neri (PoliTo)

Carla Raffaelli (UniBo)

Dimitra Simeonidou, Chinwe Abosi, Eduard Escalona (UEssex)

Abstract:

This document is the second deliverable of the WP21 "Topical Project on Service Oriented Optical Network Architectures". This report gives an overview of the WP21 activities during the first ten months of the project and the plans for the second year.

There are 10 partners involved in this work-package, which is structured in five joint activities.

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 second deliverable of the WP21 "Topical Project on Service Oriented Optical Network Architectures". This report gives an overview of the WP21 activities during the first ten months of the project and the plans for the second year.

This workpackage concentrates on service-oriented architectures that will be enabled by an underlying optical transport network capable of dynamic and agile resource allocation. To allow efficient and intensive collaboration between BONE partners, the activities were split up in five joint activities, each of them including a limited number of collaborating and contributing partners. This allows for a limited group of partners (3 to 5) to collaborate on a common research topic and to have close collaborations and interactions. This is also stimulated by regular physical workpackage meetings (twice a year) and mobility actions (one per joint activity is targetted), where a researcher from one institute is spending some time abroad at the other institute to work together on a daily basis.

In the first two joint activities, service composition techniques and user-to-network interfaces are defined for service oriented optical networks. The study of grid services over optical networks is tackled in joint activity 3. The impact of services on the switch architectures and control mechanisms is studied in joint activity 4. The last activity concentrates on the energy-efficient provisioning of ICT services, with both estimating studies of the ICT footprint and some potential energy reducing measures that can be taken.

2. Participating partners

In this workpackage TP21, there are currently 10 partners actively collaborating, as shown in Table 1. This group of partners brings together the main expertise on service oriented optical network architectures.

Partner No	Member
1	IBBT
19	AIT
21	RACTI
23	UOP
27	FUB
30	POLITO
32	DEISUNIBO
38	AGH
41	KTH
47	UEssex

Table 1: Current work package participants in the WP21 joint activities

3. Overview of joint activities

To organize the integration and research activities in an efficient way, the optimal structure was discussed at the meeting in Athens (June 2008) together with the related workpackages. As a result, the workpackage TP21 tasks were divided over 5 joint activities, each of them including a limited number of collaborating and contributing partners. The following table shows the key information about these joint activities:

No	Joint Activity Title	Responsible person	Participants	Mobility Action	Deadline
1	Programmable Service Composition Algorithms for Service Oriented Optical Networks	Chinwe Abosi (UEssex)	UESSEX, UoP, RACTI	Yes	M24
2	UNI extensions for Service Oriented Optical Networks	Eduard Escalona (UEssex)	UESSEX, RACTI, AIT	Yes	M24
3	Photonic Grid Dimensioning & Resilience	Chris Develder (IBBT)	IBBT, AIT, RACTI, (AGH)	Yes	M24
4	Impact of services on optical switch architectures and control	Carla Raffaelli (DEIS-UniBO)	UniBo, UEssex, UoP	Yes	M24
5	Green optical networking	Mario Pickavet (IBBT)	IBBT, PoliTo, UEssex, (KTH, FUB)	Yes	M24

Table 2: Summary list of the planned joint activities

As it is depicted in the above table, five joint activities with at least five mobility actions are planned for this work package. The duration of the joint activities covers the two years of the project. Each joint activity incorporates 3 to 5 partners, leading to intense and efficient collaborating projects. The partners joining later have been indicated between brackets, their contributions will be reported upon in the next deliverable.

4. Research and integration results: detailed description

This section provides a detailed description of the integration and collaboration activities in the five joint activities. An overview of the main research results obtained in the first ten months (January 2008 – October 2008) is presented, together with an indication of the planned activities from November 2008 on.

4.1 Joint Activity 1: Programmable Service Composition Algorithms for Service Oriented Optical Networks

4.1.1 Architecture for programmable service composition

A new architecture for programmable service composition was proposed and designed. The networking paradigm introduces service awareness in the core optical network, by creating self-organized islands of service transparency. A service island consists of a single (non-network) resource and a group of networking nodes that constitute the shortest path towards that resource. This group of nodes is service transparent and thus upon a service request, end-users' data are transparently forwarded, to the island's resource and not outside it. The proposed architecture and particularly the service islands are self managed entities in the sense that core nodes are self-organized in an ad-hoc fashion, based on multi-criteria path selection algorithms, thus adapting themselves to updated networking or non- networking conditions. During the 1st year, the proposed architecture was detailed, its basic notations and metrics were defined, as well as how the networking peers (nodes) interact to each other to form self-managed ad-hoc entities. Node interaction is based on resource discovery algorithms that "discover" information from both network and non-network resource providers and facilitates end-to-end service creation. In general, each service is composed of service attributes that can be different for different services and user requests. Potential implementation using currently proposed GMPLS extensions were also defined. Figure 1 shows the concept of self-organization. In this example, two kind of services are offered, denoted as service A and B. Each one possesses two set of resources denoted as resource #1 and resource #2, connected to specific egress nodes. 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.

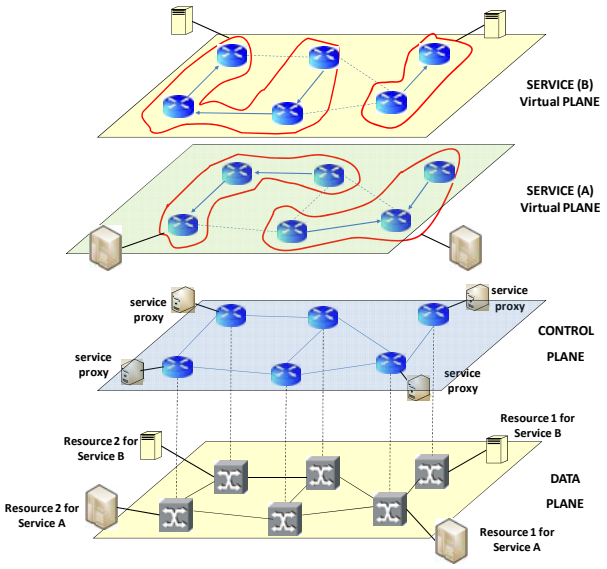

Figure 1: network paradigm for programmable service composition employing service transparent optical islands.

Figure 2 and Figure 3 show an example of a grid network and how network nodes are organized after routing table establishment. For the self-organization process, an optimization function F_s F_s has been used for path (island selection) as follows:

$$F_{s\hat{p}} = a_1 \frac{B_{R_i} - \mathrm{B}_{\min}}{\mathrm{B}_{\max} - \mathrm{B}_{\min}} + a_2 \left(\frac{h_{R_i \to s\hat{p}} - h_{\min}}{h_{\max} - h_{\min}} \right)^{-1} + a_3 \left(\frac{D_{R_i \to s\hat{p}} - D_{\min}}{D_{\max} - D_{\min}} \right)^{-1} + a_4 \frac{CPU_{s\hat{p}} - \mathrm{CPU}_{\min}}{CPU_{\max} - CPU_{\min}}$$

where the min, max refer to the corresponding min, max values among the set of all candidate islands in the vicinity of node R_i. In the specific example of node organization, shown in Figure 3, the optimization function was

$F_s = 50\% \frac{B_{R_i} - \mathrm{B}_{\min}}{\mathrm{B}_{\max} - \mathrm{B}_{\min}} + 50\% \frac{CPU_i - \mathrm{CPU}_{\min}}{CPU_{\max} - CPU_{\min}}$, subject to $D \leq 80 \cdot h, h \leq 3, ST \geq 40\ (a.u)$

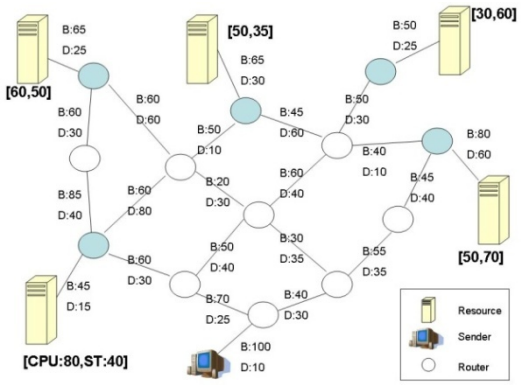

Figure 2: Grid network example consisting of five (non-network) resources and thirteen networking nodes. Numbers on links and resources denote the available bandwidth, delay, CPU and storage respectively.

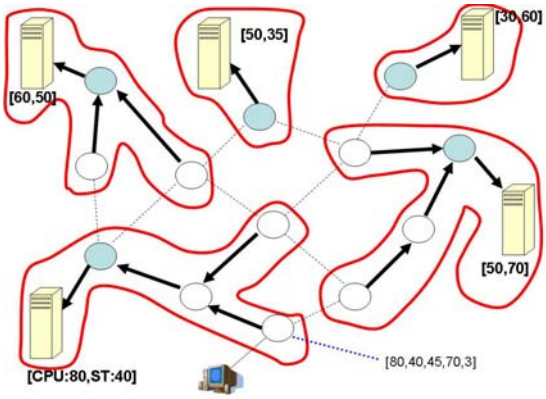

Figure 3: Network example after routing table establishment.

4.1.2 Resource discovery algorithms

Resource discovery algorithms enable resource state information discovery and dissemination by maintaining a resource state database of a resource domain (Figure 2) that facilitates service creation. A mechanism that provides an accurate representation of network and IT resources in an abstracted form with low signalling overheads under dynamic conditions was implemented. The abstracted representation is to allow resource providers to hide internal details of their resource (network and IT) while exposing enough information for successful service compositions. This gives resource providers the opportunity to participate in service creation at whatever level of abstraction they desire. At the highest level of abstraction, service oriented resource discovery requires information about service-end nodes and the connectivity between them. Service end-nodes are nodes that make and/or execute requests for services.

The challenge is to find possible ways to accurately represent resources without exposing too much information about the resource details so that the following objectives are achieved: a) maximise the number of requests accepted, b) minimise the number of possible requests accepted by the service creation process, which cannot be satisfied by the resource provider c) minimise the number of resource-state update messages propagated to the service plane. To meet the above conditions (a, b and c) the following issues were addressed: (i) Highly accurate abstract structural representation of the resources (ii) Accurate approximation of resource metrics (iii) Adequate monitoring of resources to provide accurate information to the service plane (iv) Implementation of a mechanism to trigger resource updates.

Network resource abstraction consists of two subsequent steps: (I) Structural abstraction reduces the physical topology to a full-mesh connected logical topology consisting of only service end-nodes. (II) Parametric abstraction associates the connectivity between service-end nodes by service oriented performance metrics such as bandwidth, delay,

wavelength availability etc. Four parametric associations are compared: (i) Best Case (SILK-BC) (ii) Worst Case (SILK-WC) (iii) Average Case (SILK-AVG) (iv) Modified Korkmaz-Krunz (SILK-KK), against two novel algorithms (v) CON (vi) Extended Korkmaz-Krunz (SILK-EKK). IT resource abstraction mechanism presents and computes resources levels according to a provider defined resource abstraction level. Depending on the provider, it can be computed based on physical capacities of resources (number of processors, amount of memory etc), functional capabilities (MIPS or MFLOPS of CPU etc), or a combination. The abstraction considered is based on physical capacities. Three levels of abstraction are identified: (i) Detailed (D), (ii) Multiple aggregate (MA), (iii) Single aggregate (SA). Resource-state update algorithms adequately monitor resource-state to provide accurate information to the service plane. A novel hybrid based (HB) update algorithm computes updates resource-state using a combination of event based and timer based update algorithms.

The algorithms were implemented on a variety of network topologies including GEANT European Topology and the US NSFNET under static and dynamic conditions. Under "static" conditions, the algorithm is implemented without taking into account any changes in the resource state, i.e. the resource state is assumed to be constant during the simulation. Static conditions measure the accuracy of the abstraction algorithm. Under dynamic conditions, the resource state is changed according to an updating policy. Dynamic conditions measure the abstraction algorithms under more realistic circumstances.

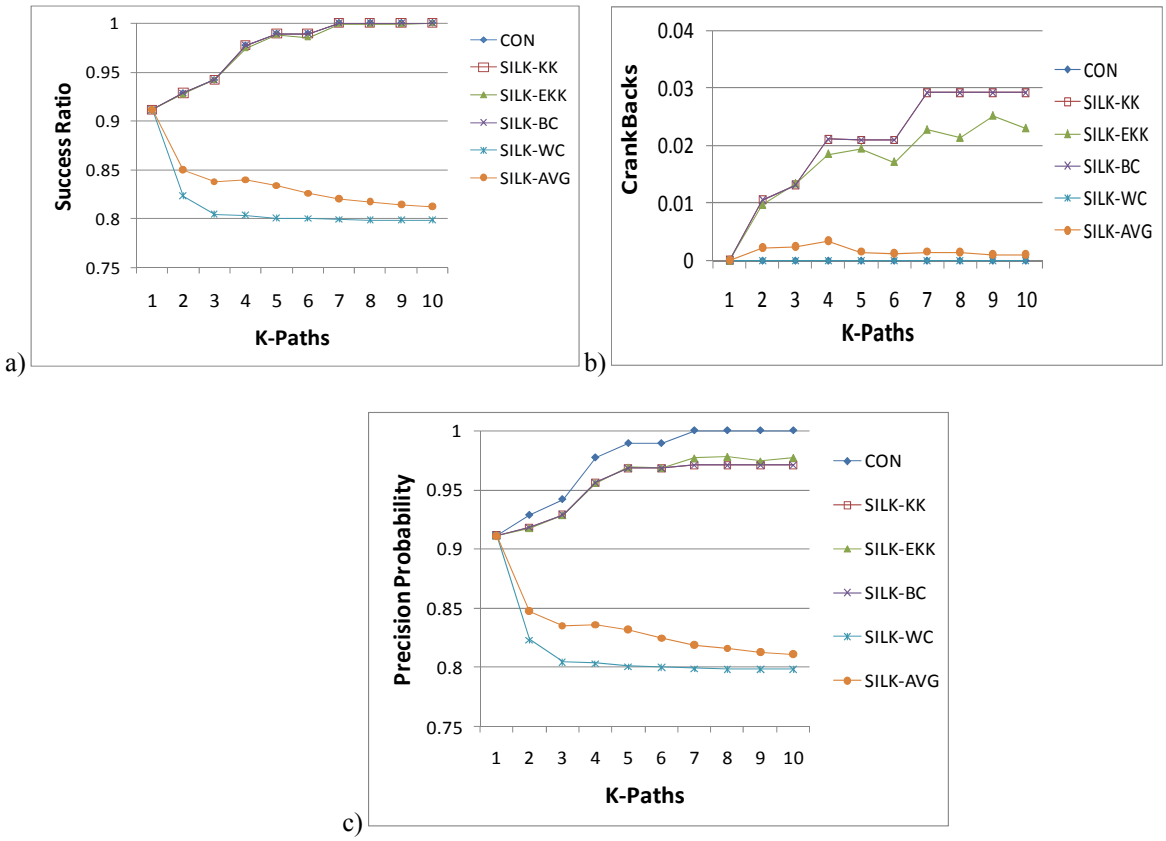

Figure 4 – Static Case: The effect of the number of paths used to calculate the connectivity between service-end nodes (a) success ratio (b) crankbacks (c) precision probability

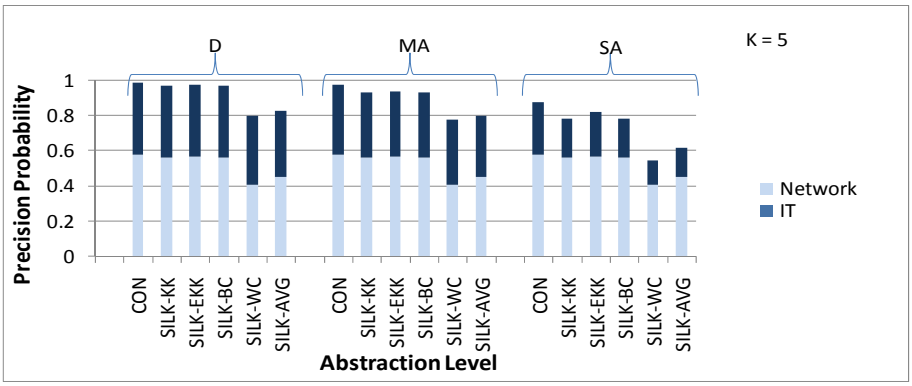

Figure 5 – Static Case: Impact of IT resource abstraction and network abstractions on the precision probability

Under static conditions, SILK-EKK and CON perform best on all evaluation metrics. SILK-BC and SILK-KK perform comparatively. This is because these abstraction models overestimate the network state information by advertising the best delay and bandwidth values. This results in a high probability of accepting feasible paths in the abstracted models. Contrarily, SILK-AVG and SILK-WC have the worst performance in all evaluation metrics. On the interactions between network resource abstractions and IT resource abstractions, it is observed that at lower abstraction (D) levels, the overall performance of IT resources is better than higher levels. This decrease in performance is due to the nature of constraints placed by clients.

Under dynamic conditions, SILK-WC and SILK-AVG are excluded from the results because of their poor performance. SILK-BC and SILK-KK are excluded because they exhibit analogous success ratio performance, which is slightly worse than SILK-EKK. The threshold values chosen are shown in Table 3. Update schemes A – D use the Hybrid algorithm (HB), E - F use the Time-Based algorithm (TB), and G - H use the Event-Based algorithm (EB). The range (max and min values) over all abstraction models was observed.

	A	B	C	D	E	F	G	H
Timer	3000	1500	3000	1500	3000	1500		
UpperThreshold1	16	16	14	14				
UpperThreshold2	14	14	12	12				
LowerThreshold1	4	4	6	6				
LowerThreshold2	2	2	4	4				
Number λ status change							2	16

Table 3 - Threshold Parameters for Update Algorithm

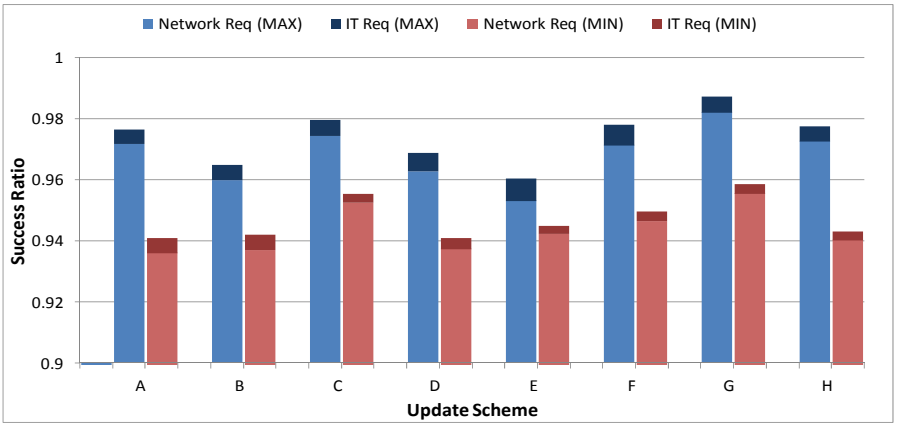

Figure 6 – Dynamic Case: Measure of success ratio on abstraction models using the Hybrid Update Algorithm

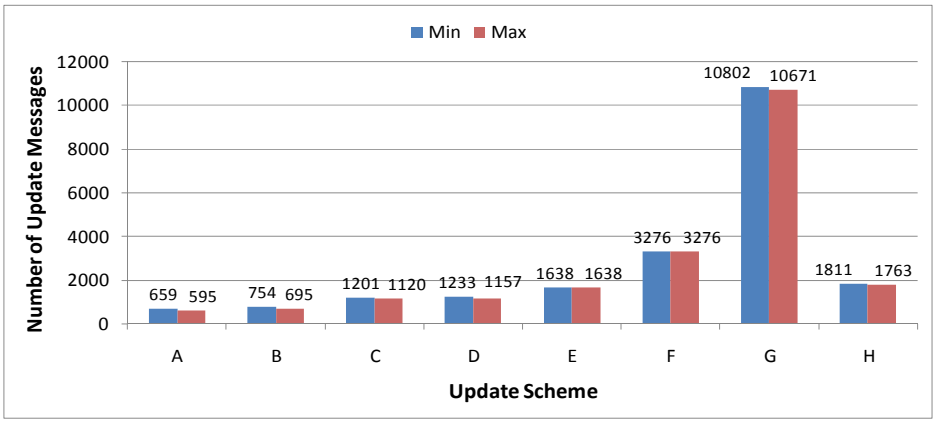

Figure 7 –Dynamic Case: Number of Update Messages

The results obtained from the simulation show that our resource abstraction and update policy algorithms disseminate resources-status information with high accuracy and precision and low signalling overheads. This is an important consideration in the introduction of a service plane since the resource discovery mechanism is responsible for the retrieval of information about the resources that are used to compose services. This in turn determines the efficiency and scalability of the service creation process and the service plane.

In the 2nd year of this joint activity, the resource discovery algorithms and the service composition algorithms developed within the framework of this joint activity will be incorporated. In addition, service proxy operation for end-to-end service connectivity will be detailed.

4.2 Joint Activity 2: UNI Extensions for Service Oriented Optical Networks

The technological evolution drives the need of a number of distinct layered architectural models across geographical organizational boundaries, heterogeneous environments with different policies, service provisioning systems, control and transport planes as well as security standards. In order to take advantage of IT resources (computers, instruments, sensors, data resources), network resources (bandwidth, wavelengths, switches, ports), as well as storage resources, in a joint and seamless way, Grid and network provisioning systems must be interfaced via a generic interface. Moreover, the advent of new applications that require higher processing and storage capacity necessitates the development of interoperable procedures for requesting and establishing dynamic network and non-network services between clients, applications and resources, all connected by the transport network. Optical networks have shown that its capacity is able to cope with the bandwidth requirements imposed by these applications and services; but to take profit of this capacity a user to network interface has to be properly designed. The development of such interface has to support advanced network services requests such as advance reservations or auto-discovery mechanisms.

This emerging scenario requires a closer cooperation between the service layer and the transport layer. The convergence can be carried out by extending current solutions or by defining a new interface that considers the characteristics of both layers keeping it abstract and general enough to support future technologies.

The main objectives of this Joint Activity are to identify the requirements imposed by the service layer and define interoperable procedures for an efficient communication between services and network which will help on the development of extensions to existing protocols or the definition of a new interface. Different types of services such as resource discovery, characterization, allocation, and management services (middleware and connectivity services) need to be supported. All of these issues must be addressed by protocols and mechanisms that build an architectural model.

OIF UNI 1.0 focuses primarily on the ability to create and delete on-demand point-to-point transport network connections according to bandwidth, signal type and routing constraints. These connections can be SONET services of payload bandwidth STS-1 and higher, or SDH services of payload bandwidth VC-3 and higher whereas their properties are negotiated during the connection establishment phase.

The procedures involved through messaging between a UNI-C and a UNI-N entity (i.e. UNI signalling) for the invocation of transport network services are:

- **Connection creation**: Permits the creation of a connection under specific network constrains and security procedures.
- **Connection deletion**: Identifies the disposal of an existing connection.
- **Connection status enquiry**: Represents the exchange of specific connection attributes allowing the connection status discover.
- **Neighbour Discovery (Optional)**: Neighbour discovery procedure allows Transport Network Elements (TNEs) to discover their identities as well as the identities of remote ports to which their local ports are connected and can be characterized as an essential procedure required for performing interface mapping between a client and a TNE.
- **Service Discovery (Optional)**: Service discovery is invoked after neighbour discovery is complete to allow a UNI-C to expose to the transport network the client device capabilities and to acquire from the UNI-N, transport network service information (i.e. the signalling protocols used and UNI versions supported, client port-level service attributes, transparency service support, and network routing diversity support).
- **Signalling Control Channel Maintenance**: OIF UNI 1.0 supports procedures for maintenance of the control channel under different configuration possibilities. These procedures allow signalling peers to continuously monitor and maintain control channel connectivity for UNI signalling.

OIF work has been continued towards UNI 2.0, under which a connection can be a SONET service of bandwidth VT1.5 and higher, or SDH service of bandwidth VC-11 and higher, an Ethernet service, or a G.709 service and the properties of the connection are defined by the attributes specified during connection establishment. The following features are added in UNI 2.0 [OIF-UNI2.0-COMM]:

- Separation of Call and Connection Controllers
- Multi- and Dual- Homing for Diverse Routing
- Non-Disruptive Connection Modification
- 1: N Signalled Protection
- Sub STS-1 Rate Connections
- Transport of Ethernet Services
- Transport of G.709 Interfaces
- Enhanced Security

OIF UNI 2.0 is officially a working document in OIF Architecture and Signalling Working Group, but it has a consolidated position towards becoming a de-facto Implementation Agreement, because of the wide vendor support and the many interoperability events demonstrated worldwide in conferences and research projects).

Some of the possible extended functionalities that a service oriented network interface should include are:

- **Network Topology Enquiry and Restoration:** the interface should be able to provide network topology information to users/services and implement various protection and restoration schemes to deal with the diverse nature of transport network failures.
- **Network Resource Availability:** All available network resources (i.e., amount of bandwidth, links, cross-connects, etc.) should be discovered and facilitated to the service layer.
- **Network Resource Capability:** For supporting different bandwidth demands for users and services the interface should introduce flexible allocation mechanisms providing multiple wavelengths, single wavelength or sub-wavelength bandwidth.
- **Network Advance Reservation:** Through automatic service discovery the interface should provide users the capability to automatically schedule, provision and set-up light-paths across the network, facilitating thus network advance reservations.
- **Traffic classification and shaping:** The interface should be able to map the user data traffic into the used transmission entity (e.g., timeslot, wavelength etc.) performing traffic classification and aggregation.
- **Data plane security:** A security mechanism able to support security credentials and policy information provided by any agreement provider should be also facilitated.

Some of these new functionalities and the standard ones should be handled by the exchange of protocol messages. Each protocol uses different ways of encapsulating the required parameters inside the messages. Thus, an abstract message definition is required. In the table below, “UNI-C” and “UNI-N” are used to identify the entities at the two ends of a network service that initiates and terminates a given action (C: client, N: network). *Table 4* depicts messages related to network reservation establishment, and service discovery.

Message No.	Abstract Message Name	Message Direction
1.	NS Create Request	UNI-C ➔ UNI-N UNI-N ➔ UNI-C
2.	NS Create Response	UNI-N ➔ UNI-C UNI-C ➔ UNI-N
3.	NS Create Confirmation	UNI-C ➔ UNI-N UNI-N ➔ UNI-C
4.	NS Delete Request	UNI-C ➔ UNI-N UNI-N ➔ UNI-C
5.	NS Delete Response	UNI-N ➔ UNI-C UNI-C ➔ UNI-N
6.	NS Status Enquiry	UNI-C ➔ UNI-N UNI-N ➔ UNI-C

Message No.	Abstract Message Name	Message Direction
7.	NS Status Response	UNI-N ➔ UNI-C UNI-C ➔ UNI-N
8.	NS Notification	UNI-N ➔ UNI-C
9.	Network resource availability	UNI-N ➔ UNI-C
10.	Network topology information	UNI-N ➔ UNI-C
11.	Network end-point assigned addresses (TNAs)	UNI-N ➔ UNI-C

Table 4: Interface abstract messages

Part of this work is also being done in the recently created NSI workgroup in OGF. This workgroup aims to facilitate interoperation between Grid users, applications and network infrastructures spanning different service domains, via the development of abstract messaging and protocols. The NSI WG must provide a general and open definition independent of implementation of provisioning systems or control planes. It should be flexible, modular and scalable to facilitate future enhancements.

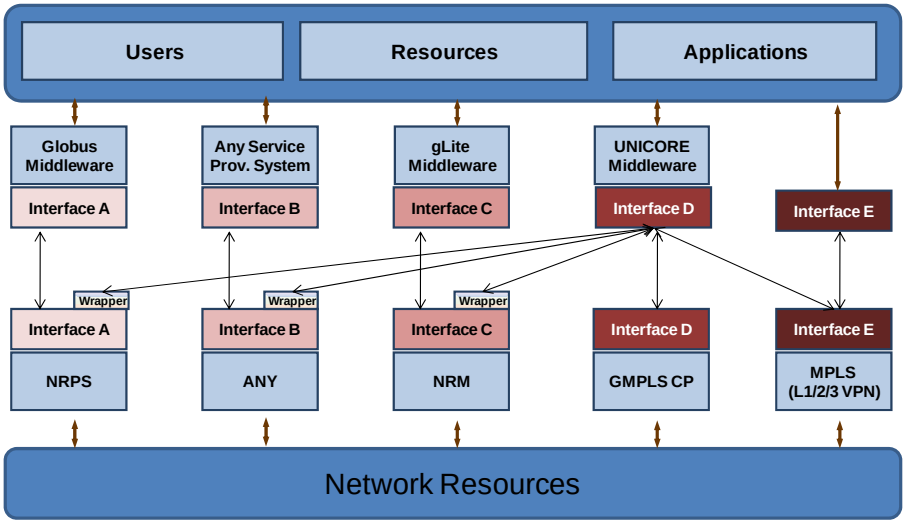

Figure 8: Interoperability between service and transport layers

Currently, interoperability between applications, middleware and network provisioning systems is implementation specific and depending on the solution used the interfaces between them may "speak" different languages. When trying to interoperate different systems, wrappers are needed in order to "translate" from one language to the other (*Figure 8*). Thus, the NSI WG will define an abstract interface which will support the features required by the different solutions in a standard and common language to be adopted by the service layer and network provisioning systems.

The work carried out within the NSI WG involves the elaboration of a use-case deliverable which will depict different real scenarios in which a service oriented user to network interface is needed. These use-cases will expose the requirements and functionalities that will enable the definition and implementation of the new interface protocols.

The remainder of the work for this Joint Activity will be focused on further identifying advanced functionalities to be supported by the user to network interface which will define a general use-case for its development. The two different approaches (extensions to UNI and definition of a new extended interface) will be studied and solutions will be proposed for both implementations.

4.3 Joint activity 3: Photonic Grid Dimensioning and Resilience

4.3.1 Main JA research results

Introduction

During the first project year, this Joint Activity (JA) mainly focused on Grid dimensioning issues. Next to that, work on resilience has been initiated and will be the main focus of the second project year as described in Section 4.3.2.

The results of the dimensioning studies as described in the following sections mainly explore three areas:

- Joined network & resource dimensioning addresses the basic problem of dimensioning Grids. In more traditional networks, only the amount and location of network resources (fibers, switches…) needs to be determined, but in Grids also the grid resources (for computation, storage…) need dimensioning. An additional complexity arises by the fact that grid jobs have no pre-determined target location, i.e. the grid can freely choose where it ends up being executed (cf. anycast routing), hence there is no so-called traffic matrix fixed a priori. A sequential approach solving the Grid dimensioning problem is summarized in the following.
- Resource dimensioning: Even disregarding the network dimensioning sub-problem, determining the amount of required grid resources is complex. For resource dimensioning we propose the Spectral Clustering Scheduling (SCS) algorithm that finds the minimum number of computational resources required for task scheduling, and the corresponding task-to-resource assignments, so as to increase the utilization efficiency and the percentage of tasks served by the Grid without violating their Quality of Service (QoS) requirements. With respect to storage resources, we consider the Data Consolidation (DC) problem which arises when a task (i.e. a Grid job) needs multiple pieces of data (which can be taken from replicas intelligently distributed over the network).
- Physical layer aware dimensioning: In optical networks, physical layer characteristics can have an important impact on the overall network design: physical impairments may degrade the quality of the signal carried by the optical network, leading to either excess overhead to higher layers or worst-case to practically unusable paths. In any case, this type of effects could be detrimental to a Grid application. To overcome this issue the incorporation of impairment-awareness to the problem of network dimensioning becomes critical.

A phased approach to dimensioning optical Grids

When deploying Grid infrastructure, the problem of dimensioning arises: how many servers to provide, where to place them, and which network to install for interconnecting server sites and users generating Grid jobs? In contrast to classical optical network design problems, it is typical of optical Grids that the destination of traffic (jobs) is not known beforehand. This leads to so-called anycast routing of jobs. For network dimensioning, this implies the absence of a clearly defined (source,destination)-based traffic matrix, since only the origin of Grid jobs (and their data) is known, but not their destination. The latter depends not only on the state of Grid resources, including network, storage, and computational resources, but also the Grid scheduling algorithm used. We present a phased solution approach to dimension all these resources, and use it to evaluate various scheduling algorithms in two European network case studies. Results show that the Grid scheduling algorithm has a substantial impact on the required network capacity. This capacity can be minimized by appropriately choosing a (reasonably small) number of server site locations: an optimal balance can be found, in between the single server site case requiring a lot of network traffic to this single location, and an overly fragmented distribution of server capacity over too many sites without much statistical multiplexing opportunities, and hence a relatively large probability of not finding free servers at nearby sites.

A classical network design problem is dimensioning: figuring out how much capacity is needed for the network to be able to transport a given amount of traffic. Typically, this traffic is specified in a traffic matrix: for each source site i and destination site j, the amount of traffic flowing from site i to j is given by as a number T_{ij} (say in Mbit/s). A broad range of dimensioning algorithms is available, either based on heuristics or exact solution methods using for example Integer Linear Programming (ILP). The algorithms vary depending on the network technologies and topologies, design criteria, single or multi-period planning, etc. Yet, if we want to apply any of these approaches for dimensioning grids, the problem arises of accurately estimating the traffic matrix. Indeed, given the anycast principle typical of Grids, the destination of the traffic (Grid jobs) is not given a priori.

In this work, we focus on a ‘clean slate’ or greenfield Grid dimensioning problem finding the complete Grid capacity required to meet a given Grid job arrival pattern. Also, we assume fully flexible scheduling strategies without any knowledge of probabilities for selecting a given destination site. This presents a viable dimensioning methodology, and assesses the impact of the scheduling algorithm on Grid network dimensions. Yet, since development of scheduling algorithms as such is not this paper’s primary concern, we will assume fairly straightforward scheduling strategies,

based on a single all-knowing scheduler, finding a free server for every arriving job based solely on the job's arrival time and duration, and server processing speed and occupation.

We will take an iterative dimensioning approach, starting with an algorithm for choosing appropriate server site locations: not every grid site will necessarily be a server site. Next we will calculate the amount of servers needed (and distribute them amongst the chosen server site locations). Subsequently the inter-site job rates are determined, and hence required bandwidth. In the work presented, we focus on computational Grids, where jobs consist of a single unit of work submitted to the Grid, characterized by a data size, and a computational requirement (say, expressed in number of floating point operations, FLOP). An accurate problem statement is the following:

- Given:
  - Graph representing the network topology (nodes representing Grid sites and switches, links the optical fibers interconnecting them),
  - Arrival process of jobs originating at each site,
  - Job processing capacity of a single server CPU (an average of λ jobs/s), and
  - Target maximum job loss rate,
- Find:
  - Locations of the server sites,
  - Amount of Grid server CPUs at each site, and
  - Amount of link bandwidth to install,
  - While meeting the maximum job loss rate criterion and minimizing network capacity.

Given the complexity of the problem (such as the dependence of the network capacities on the choices of server locations and capacities), we opt for a phased solution approach comprising subsequent steps. The first step will be to find K server locations (out of the N grid sites), while a second step finds the server capacities at each of the K chosen sites. The third step will determine the amount of jobs exchanged between the grid sites and the server locations. The final fourth step will be to calculate the actual network dimensions, that is link bandwidth. Each of these steps is discussed in detail in [8].

The proposed step-wise scheme to dimension both server CPU and network resources for a computational Grid was applied in a realistic case study, to highlight the importance of choosing an appropriate scheduling and server CPU placement algorithm when trying to limit the network resource requirements. Again, we refer to [8] for detailed results and discussion.

From a network perspective, the most important criterion is network bandwidth. To establish the network dimensions in the considered case study, we would need to choose a particular network technology (OBS versus OCS, or hybrids). Yet, these relate to the traffic matrix stating the amount of bandwidth exchanged between each node pair. A useful measure to comprehensively summarize this information in the assumed Grid context is the average hop count a job needs to traverse to reach the server it will be executed on. Hence, we will use the average job hop count as a measure to judge the network capacity requirements. We summarize the average job hop count results for varying number of server sites K in *Figure 9*.

Comparing the various combinations of dimensioning and scheduling alternatives, the relative influence of the scheduling algorithm seems to be important. The reasonably large fraction of traffic sent to non-closest server sites—recall the 'local processing rates' from the previous section—has an important influence on the network load. Hence, by adopting shortest path driven scheduling (SP), the lowest average job hob count is reached.

The influence of the dimensioning strategy is also obvious, though less significant. Especially for a larger number of server sites K, it pays off to intelligently distribute server capacity: prop and lloss (which hardly differ in resulting average job hop count) result in lower network load than straightforward uniform server distribution.

Comparing the European backbone case with the EGEE/Geant2 case, we note that the major qualitative difference lies in the curves for the uniform dimensioning strategy: this curve is mainly increasing for larger values of K in the EGEE/Geant2 case. This can be explained by the larger relative variations in shortest path hop counts in this network: the penalty of unintelligently distributing server capacity is more pronounced.

With respect to the choice of the number of server sites K, we observe there is an optimal choice, which tends to lie around K = 5 in the studied cases, ranging between 1/7 and 1/4 of the total number of sites. Note that the optimum depends on both the dimensioning strategy and the scheduling approach. When too much server sites are installed, the total server capacity is fragmented too much, resulting not only in low 'local processing rates' (i.e. the chance a job has to find the closest server free) but also a lot of jobs sent to remote servers. Indeed, for larger number of server sites, the opportunities of statistical multiplexing diminish and with it the probability of finding a free server at the closest server site for a particular job. This apparently outweighs the fact of lowering the average distance to a server site.

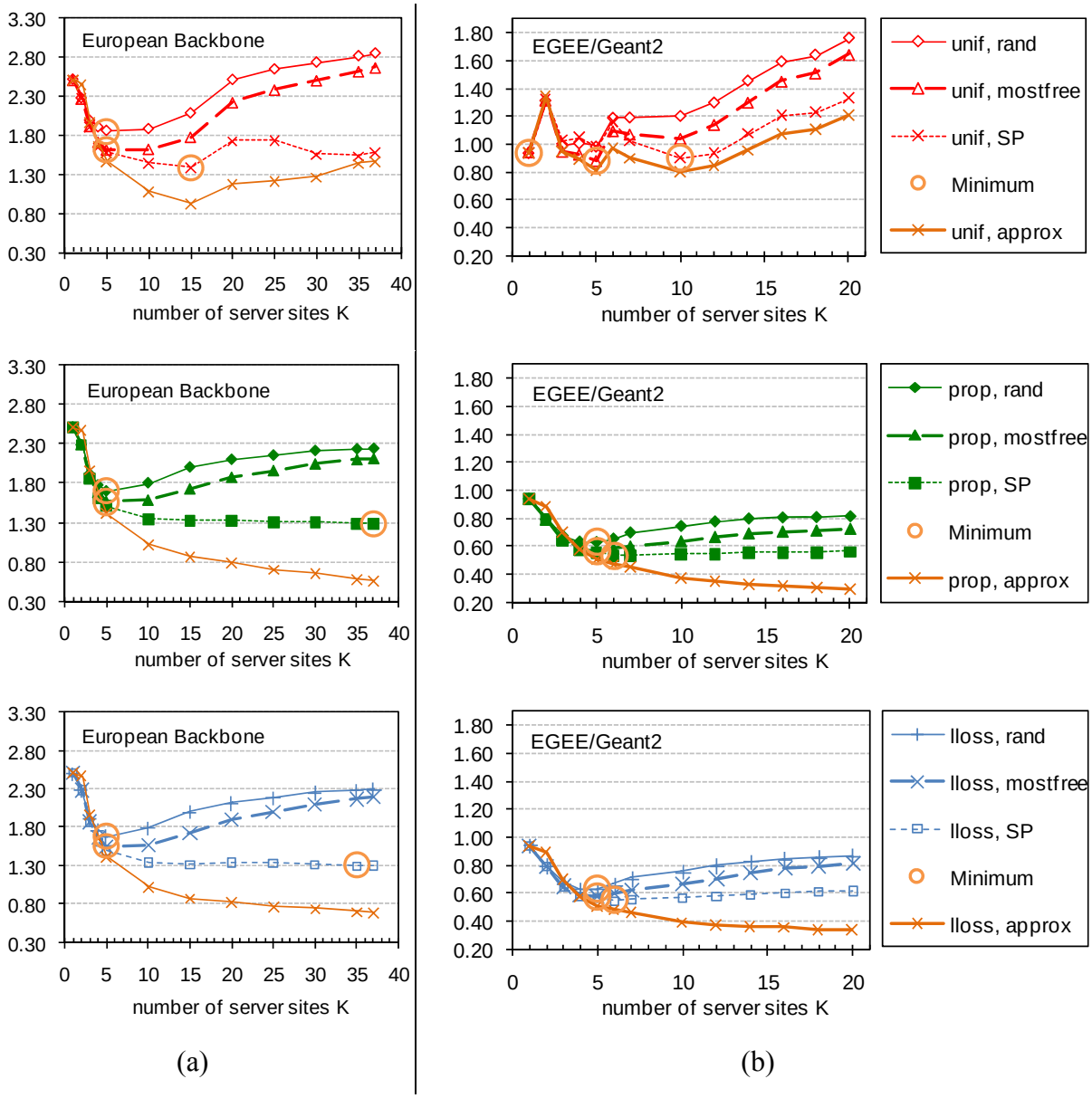

Figure 9: The required network capacity, which is proportional to the average job hop count, is minimized by adopting shortest path routing and scheduling (SP), intelligently positioning server capacity (prop vs unif)*, and deploying a reasonable number of server locations: average hop count over all jobs for (a) the European backbone network, (b) the Geant2 network. The analytical fixed point approximation (approx) deviates from simulation results because it does not accurately capture site inter-dependencies.*

In conclusion, we proposed a dimensioning methodology fully taking into account the anycast routing principle, that is, without presuming a priori knowledge of (source, destination)-based traffic. The proposed step-wise methodology is suitable for dimensioning both server and network capacities. We used the methodology to evaluate various scheduling algorithms and server dimensioning options with respect to the required network capacity. From two case studies on European topologies, we concluded that placing server capacity where a lot of jobs arrive is important to minimize network bandwidth requirements: the prop dimensioning strategy, placing a number of servers proportional to the job arrival rates at its closest Grid sites is most beneficial. With respect to Grid scheduling, a simple shortest path (SP) strategy, preferring closer server sites, led to the lowest bandwidth demands. With respect to choosing an appropriate number of server sites K (ranging from 1 to the total number of Grid sites N), we found that there is an optimal value. For a larger number of server sites, the total server capacity gets fragmented, reducing opportunities for statistical multiplexing, whereas for smaller server site counts K the average distance jobs need to travel is too large. That optimum of K depends on the scheduling algorithm and server site dimensioning strategy, and in the considered case studies was about 1/7 to 1/4 of the total number of sites.

Network Design for Photonic Grids

The network and Grid resource dimensioning approach as outlined above makes some simplifying assumptions, and takes an abstract view of the network. However, when dealing with *optical* Grid networks, there are a number of physical layer issues to consider. In this context, this JA (esp. AIT) focuses on the network design aspects of photonic Grids, which aim at providing cost and resource efficient delivery of network services. Although the problem of dimensioning a transport network is not new we are concentrating on the impact of the physical layer characteristics on the network design and the additional constraints and requirements it introduces. Also the implications associated with any resilience requirements that may exist due to specific service requests will be taken into consideration.

The process of designing a networked Grid entails the solution of a network dimensioning problem including specification of the capacity that needs to be supported by the network in order to seamlessly carry the estimated volume of traffic generated by the Grid sites. In this context we have developed a network dimensioning model suitable for photonic Grid network design that given an input physical topology and an amount of traffic requests in the form of a traffic matrix can dimension the following network resources:

- Number of fibers that need to be installed connecting any pair of optical nodes.
- Number of wavelengths that need to be installed on each fiber.
- Number of input/output ports of the switching nodes.

The problem of dimensioning a Grid network can be also seen as an optimization problem that involves the identification of capacity requirements of the above elements (number of fibers, number of wavelengths & node dimensions) in a way so that the total capital and operational cost of the Grid network is minimized, while all traffic demands are deterministically served. The topic of network design and the associated cost optimization problem have been widely addressed in the literature following formulations relating to the fundamentals of network optimization theory. However, certain particularities of the optical networks such as e.g. its physical layer characteristics can have an important impact on the overall network design: physical impairments may degrade the quality of the signal carried by the optical network, leading to either excess overhead to higher layers or worst-case to practically unusable paths. In any case, this type of effects could be detrimental to a Grid application. To overcome this issue the incorporation of impairment-awareness to the problem of network dimensioning becomes critical. A solution that has been proposed in the literature aiming at solving the physical layer impairment related issues is the use of selective optoelectronic regeneration at particular network locations. In this context the following points become very important:

- Specifying the optimal point in the network design process, where regenerators should be placed.
- Devising a network dimensioning method that guarantees acceptable signal quality and at the same time being tolerant to the heterogeneity of the optical components deployed throughout an optical network.

An optimal network dimensioning method has been developed that incorporates selective regeneration based on signal quality. To quantify the effectiveness of the proposed approach compared to the distance-based regenerator placement, the problem of joint dimensioning and regeneration location is formulated as an integer linear program (ILP). The linear model presented in [14] was adopted; however several amendments were introduced in order a) to incorporate the cost of regeneration into the objective function and b) to regenerate paths according to the placement criterion used. Through initial simulations it has been shown that cross-layer design of optical networks introduces substantial reduction in regenerator requirements and therefore cost savings compared to approaches employing traditional distance-based regeneration location.

An additional aspect relating to optical networks that we are considered is resilience. Due to the enormous bandwidth offered by these networks and the increasing number of “mission critical” services, survivability is becoming an essential network design aspect and should aiming at providing service resilience in an efficient and cost effective manner. In this context resource efficient network design schemes that take into consideration both working and protection traffic are important to study. Regarding grid network resilience there is a variety of resilience mechanisms available and the level of service survivability is effectively dictated by the resilience scheme(s) implemented in the network. Ideally it would be desirable to provide a 100% resilience guarantee to all services supported, but this may be unnecessary and wasteful in terms of resource utilization resulting in cost inefficiencies as different services may not require the same level of resilience. Thus, a more efficient resilience scheme suitable for networks supporting a variety of applications would be a scheme that provides different levels of network survivability to different service types in accordance with the respective service level specifications (SLSs) maximizing the network utilization [37]. Therefore in a photonic grid network environment supporting a variety of services an important requirement will be to provide

differentiated survivability services to different types of traffic enabling higher priority demands to exploit higher network availability [19, 25]. Some of our work has focused on providing resilience in WDM optical networks supporting differentiated survivability traffic requirements and more specific on fault-tolerance for high priority, high resilience traffic through the backup multiplexing technique [20]. Our work is based on the backup multiplexing technique in order to facilitate efficient resource sharing and investigates different routing and wavelength assignment schemes that considerably enhance the spare capacity utilization. A simple approach that can be used to assign different classes of service supporting varying restoration requirements is proposed and significant network performance improvement has been demonstrated through relevant simulations [12].

Related Work

The literature on dimensioning the network infrastructure in support of an optical Grid is limited. The applicability of established optimization methods in the context of Grid network dimensioning has been proposed using Divisable Load Theory towards designing tractable algorithms and it has shown incurred benefit through simulations [33]. Otherwise, past work on regenerator placement and design of translucent optical networks [27] is closely related to the problem of dimensioning a Grid network. Many of the aspects and trade-offs for designing core optical networks optimally are addressed in [32], where among others general guidelines for selective regeneration are given. Part of the associated literature focuses on studying the concept of selective regeneration and its benefits without taking into consideration the details of the physical layer. The work in [34] demonstrates the trade-off between mass of regeneration and lightpath capacity in real networks relating mostly to routing rather than network design, [36] presents solutions for both network design and connection provisioning sub-problems. [18] addresses the problem of minimum regeneration cost network design in two deployment scenarios (all nodes candidates for regeneration vs. “transparency islands”) and [31, 35] provide approximate solutions to related problems. The problem of designing translucent networks with transparency islands owned by more than a single carrier is dealt with in [16] through a game-theoretic approach. [15, 35] approach the problem of regenerator placement as a connected dominating set instance and show reduction in total regeneration cost. All this work has provided exact and approximate solutions to the design optical networks employing selective regeneration. As all this work does not consider the details of the physical layer it is based on the simplifying assumption that the physical distance covered by a lightpath is a good approximation of signal quality degradation. While this is realistic for limited hop-count lightpaths and for networks exhibiting high homogeneity in terms of deployed optical equipment, there are still various realistic scenarios, where the approach of designing translucent optical networks using a single optical reach threshold as regeneration criterion is not accurate.

[24] tests the correlation between optical reach and network performance, using analytical modeling of impairments. Although not taking a minimum cost approach concerning network design, still the aforementioned work models and evaluates the impact of a group of physical impairments to optical network design. Last, the impact of PMD in the network design is studied in [13] and a) it presents an explicit model and analytical formulation of a network design problem that takes impairments into consideration and b) it indicates that optical equipment heterogeneity is an important aspect of network design that should not to be neglected.

Network survivability can be defined as the capability of the network to provide continuous services in the presence of failures resulting in lightpath disruptions. Due to the large amount of traffic at the lightpath level, resilience mechanisms are highly critical and many schemes have been proposed to address this issue [20]. Survivability can be broadly classified into two types, dynamic restoration and predesigned protection. In dynamic restoration [28, 22] the backup lightpath discovery procedure is initiated after a primary lightpath fails. This procedure might not result in a backup lightpath identification due to a lack of network spare capacity, and therefore this method does not guarantee successful recovery. In predesigned protection [29, 10, 23] on the other hand a backup lightpath is computed and wavelength channels are reserved for it at the time of establishing the primary lightpath. If a backup lightpath cannot be found under the current network conditions, the connection request is blocked. A database of restoration paths for this method can be populated by dynamic restoration. The advantages offered by the predesigned protection method that compare with dynamic restoration are the shorter restoration times and the 100% restoration guarantee under the assumption of a single failure. A further classification of the predesigned protection method is performed based on link or path protection schemes. In the link-based method the failed link is replaced a new path, which is merged with the unaffected portion of the primary path, to constitute the backup path. This method constrains the choice of the backup paths and requires more spare resources than the path-based method [21], which computes a complete end-to-end backup path from the source to the destination of the failed primary path. In the path-based method, wavelength channels on the backup path can be either dedicated or shared. If dedicated the wavelength channels assigned to a specific backup path cannot be assigned to other backup paths, whereas in the shared method, backup paths can share wavelength channels under the single link failure assumption if their primary paths are link-disjoint. This is known as backup multiplexing and provides improved resource utilization [20]. Specifically in [26], it was shown that the total resource requirement for the dedicated backup method is 260%–265% of the requirement without lightpath protection, and it can be reduced to 186%–195% by considering backup multiplexing.

Resource Dimensioning

Whereas the previous section dealt with optical *network* specifics, here we specifically address the dimensioning issues related to the computational and the storage *Grid resources*. A number of different dimensioning issues can be considered:

1. The in-advance knowledge of the number of computational and storage resources needed to satisfy users requirements is helpful for planning/expanding the Grid Network and is a key issue when offering new services. An important goal is also the maximization of the resource utilization, which leads to the minimization of the number of computation and storage resources needed for task execution.

2. The proper placement in the network of the computational and storage resources is another dimensioning issue. For example, such an issue may arise for data intensive applications. The proper placement of the storage resources in the network (and of course their capacity) influences the applications execution time. A similar problem is the efficient placement of datasets and their replicas in the storage resources of the Grid Network. This is very important for applications requiring more than one piece of data and for data redundancy in case of failures.

3. The type of computational and storage resources used is also a very important issue. These kinds of resource can be categorized in various ways. For example by using Flash-based storage devices users can immediately increase application performance and save on energy costs compared to traditional Fibre Channel storage systems. Furthermore, computation resources can be categorized based on the type of users they serve or the priority they give to each type of users.

4. Dimensioning also relates to the effect of various computation or storage related parameters on the performance of the Grid Network and the related algorithms.

For resource dimensioning we propose the Spectral Clustering Scheduling (SCS) algorithm that finds the minimum number of computational resources required for task scheduling (Issue 1), and the corresponding task-to-resource assignments, so as to increase the utilization efficiency and the percentage of tasks served by the Grid without violating their Quality of Service (QoS) requirements (Issue 4) [10]. The tasks QoS requirements are given in the form of a desired start and finish time. In general, a number of task-to-resource scheduling algorithms have been proposed that try to maximize overall system performance (resource utilization efficiency), while missing to satisfy users requirements and vice versa. The proposed Spectral Clustering Scheduling (SCS) scheme exploits concepts derived from spectral clustering, and groups tasks together for assignment to a computing resource so as to (a) minimize the time overlapping of the tasks assigned to a given resource and (b) maximize the time overlapping among tasks assigned to different resources. The above two objectives are transformed into a matrix representation. The proposed scheduling scheme uses the notions of generalized eigenvalues and the Ky-Fan theorem to obtain an algorithm of polynomial order. The Ky-Fan theorem states that an optimal schedule can be derived as a solution of the largest eigenvectors of the matrix that represent the objectives of interest.
The SCS algorithm can help the capacity planning of the Grid infrastructure since it allows advance estimation of the number of computational resources required in order to satisfy the tasks' requirements. This becomes more relevant with the emergence of cloud computing. Knowing in advance the number of resources needed to satisfy the requirements of the users, could be useful in capacity planning and in achieving predictability in such systems. Finally, we can argue that the same concepts and ideas used by the proposed algorithm in order to find the minimum number of computational resource required for task execution, could also be used for finding the minimum number of storage resources needed for serving the large storage needs of the Grid applications.

We perform a number of experiments, in which we compare SCS with the Greedy and the Min Cut Tree algorithms. The proposed Spectral Clustering Scheduling (SCS) algorithm is recursively applied assuming different number of computational resources in the Grid. Then, we select the minimum number of resources that provide no task overlapping (no violation of the tasks' QoS). The Greedy algorithm assigns each task to a resource, so that no task overlapping is encountered, by taking into account the current load of the resources. Each time a newly considered task overlaps with the already assigned tasks, a new resource is activated and this task is assigned to this resource. The main performance metric is the minimum number of computational resources required in order to schedule all the tasks without violating their QoS requirements. Moreover, we also look at the case where the number of computational resources is given and compare the percentage of QoS violations that occur under the different scheduling policies considered. Experimental results show that the proposed SCS algorithm outperforms the Greedy and the Min Cut Tree algorithms for different values of the granularity and the load of the submitted tasks.

We also examine a task scheduling and data migration problem for Grid Networks, which we refer to as the Data Consolidation (DC) problem [11]. DC arises when a task needs for its execution multiple pieces of data, possibly

scattered throughout the Grid Network (*Figure 10*). The DC problem consists of three interrelated sub-problems: (i) the selection of the replica of each dataset (i.e., the data repository site from which to obtain the dataset) that will be used by the task, (ii) the selection of the site where these pieces of data will be gathered and the task will be executed and (iii) the selection of the paths the datasets will follow to arrive at the data consolidating site. Furthermore, the delay required for transferring the output data files to the originating user (or to a site specified by him) should also be accounted for. The algorithms or policies for selecting the data replicas, the data consolidating site and the corresponding paths comprise a Data Consolidation scheme. The investigation of the DC problem relates strongly to the placement of the storage resources and the datasets in the Grid Network (Issues 2).

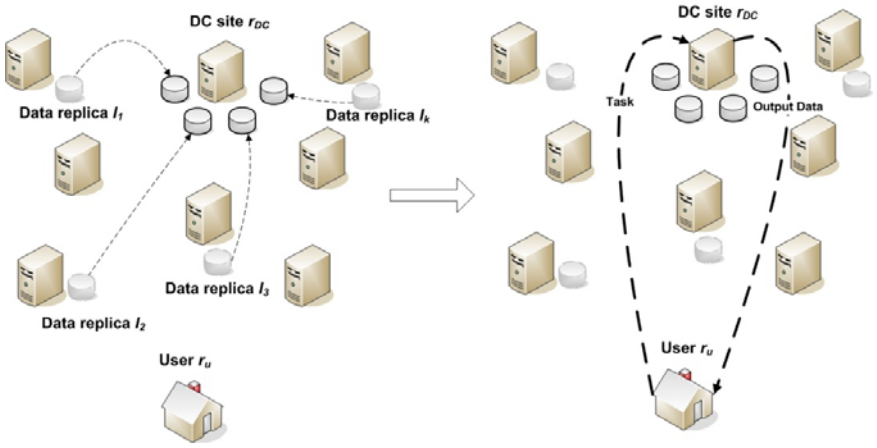

Figure 10: A Data Consolidation scenario: Initially the datasets a task requires are transferred to a single Data Consolidation site r_{DC}. After all data transfers have been completed, the task itself is also transferred to the site, where it is executed. Finally, the task's output data are transferred back to the task originating user r_u.

Generally, a number of algorithms or policies can be used for solving these three sub-problems either separately or jointly. Moreover, the order in which these sub-problems are handled may be different from the order they are presented, while the performance optimization criteria used may also vary. We propose a number of DC schemes that consider only the data consolidation (ConsCost) or only the computation (ExecCost) or both kinds (TotalCost) of task requirements. Specifically, the Consolidation-Cost (ConsCost) algorithm selects the replicas and the DC site that minimize the data consolidation time (Dcons). The Execution-Cost (ExecCost) algorithm selects the DC site that minimizes the task's execution time. The Total-Cost (TotalCost) algorithm selects the replicas and the DC site that minimize the total task delay. This delay includes the time needed for transferring the datasets to the DC site, the task's execution time, and the time needed for the output data to be transferred to the task's originating user. This algorithm is the combination of ConsCost and ExecCost algorithms. Moreover, for comparison purposes we also use the Rand algorithm, where both the replica sites and the DC site are randomly chosen. In all the algorithms examined the paths are constructed using the Dijkstra algorithm.

We experimentally evaluated the proposed DC schemes, and examined their performance by altering various parameters, such as the number of datasets, the number of storage resources, the requirements of the tasks (computation vs data intensive tasks), etc (Issue 4). We considered a P2P (opaque) network consisting of 11 nodes and 16 links, of capacities equal to 10Gbps, where the delay for transmitting between two nodes includes the propagation, queuing and transmission delays at intermediate nodes. Only one transmission is possible at a time over a link, so a queue exists at every node to hold the data waiting for transmission. We use the following metrics to measure the performance of the algorithms examined:

1. The average task delay, which is the time that elapses between the creation of a task and the time its execution is completed at a site.
2. The average load per task imposed to the network, which is the product of the size of datasets transferred and the number of hops these datasets traverse.
3. The Data Consolidation (DC) probability, which is the probability that the selected DC site will not have all the datasets required by a task and as a results DC will be necessary.

Our simulation results brace our belief that Data Consolidation (DC) is an important problem that needs to be addressed in the design of Data Grids. If Data Consolidation is performed efficiently, benefits are provided in terms of task delay, network load and other performance related parameters. Specifically, we observed the effects of the increased number of task requested datasets L, on the measured metrics and on the evaluated algorithms. Generally, given our assumption

that for any value of *L* the total size of data a task requests is the same (equal to S), we believe that in order for the Grid Network to better handle the case of increased *L*, it needs more sites equipped with storage resources than storage resources of increased capacity. This way the probability that a site holding a needed dataset is close, is increased. The higher the number *L* of datasets a task requests, the higher is the probability that these datasets will not be located at the DC site, given that the size of datasets a site can hold is limited. We observe that the algorithms that take the data consolidation delay into account (namely, the ConsCost and TotalCost algorithms) behave better than the algorithms that do not consider this parameter (that is, the Rand and ExecCost algorithms), in terms of the task delay. However, as the number *L* of datasets a task requires increases, the average task delays of all the algorithms converge. We observed that the algorithms that do not take into account the data consolidation delay (that is, the Rand and ExecCost algorithms) induce, on average, a larger load on the network than the algorithms that do take this into account (ConsCost and TotalCost algorithms). This is because the former algorithms transfer on average more data, over longer paths. Moreover, the decisions made by these algorithms are not affected by the dataset sizes *I* or their number *L*, and as a result they induce on average the same network load. We also observe that as the number of sites in the network increases the average task delay and the load induced to the network also increases. This is because having a larger number of sites increases the chance that the data will not be located at the DC site or at sites close to that. Finally, the TotalCost algorithm performs better in all cases, as tasks become more cpu- rather than data-intensive.

4.3.2 Planning for future research

First of all, we see some opportunities in further exploring joined network and resource dimensioning (in a more integrated way than the currently proposed sequential scheme of Section 0). Yet, the major efforts in the remainder of this JA will now mainly consider Grid resilience: how to make a Grid resilient against possible failures? Here, failures can be resource failures (e.g. broken hard disk causing server crash) or network failures (e.g. fiber cut disconnecting two switches). Both types of failures will be addressed in our studies.

For *network resilience*, we will build on existing literature and our developed tools (e.g. for grid dimensioning) to study the efficiency of certain network resiliency techniques in the Grid context. Two main approaches exist for fault management: path protection and restoration. The failure of a single fiber link is the most frequent cause of network failure and may cause all lightpaths using the faulty fiber to fail. In restoration, after a failure occurs, we try to find alternate routes for all disrupted lightpaths, using any excess unused resources. In path protection, for each logical edge we set up two lightpaths, the primary and the restoration lightpath, which uses a path that is fiber disjoint with respect to the path used by the primary lightpath. The resources for the restoration lightpath are allocated at design time, before any failure has occurred. When there is a failure due to a fiber cut, communication that uses the disrupted primary lightpaths, is resumed using the corresponding restoration lightpaths.

The major issue to assess the “cost” of providing resilience, is to determine the amount of resources it requires (in addition to the resources in a failure-free scenario). Hence, this boils down to a dimensioning study, for example: Given the requested connections (offline traffic) and the characteristics of the network, our objective is to minimize the congestion of the network, by determining the appropriate wavelengths and routes through the physical topology for a primary and a restoration lightpath.

In this context, we plan the following activities:

- RACTI plans to design and implement various algorithms for 1+1 protection and 1:1 restoration based on LP relaxation formulations.
- AIT will further study the impact of service resiliency requirements: network survivability will be directly associated with service level resilience requirements and as such will be examined as a service specific parameter, where service resilience can be variable. (It should be noted that some constraints are specific to optical networks (such as e.g. physical layer performance), which will be treated accordingly in order to identify optimum solutions and practical approaches for photonic grid network design.)
- IBBT will consider a Grid-specific resiliency scheme, offering an alternative to pure network resilience: since Grid jobs in general can be executed at another server, we will consider the possibility providing each Grid job flow with a backup path that ends at a different destination than the primary. Through a dimensioning study, its potential will be evaluated.
- IBBT and AIT will also consider the potential advantage of considering different resilience schemes for different tasks belonging to a workflow (e.g. protecting tasks with a lot of dependent following tasks, but re-running less critical tasks without compromising the total time of a complete workflow even in case of failures).

For *resource* resilience, RACTI plans to consider Data Consolidation (DC) and resiliency. It is evident that the site were the datasets consolidate becomes a critical point for the operation of the Grid Network. In case this site fails in any

way, then the Data Consolidation operation needs to be repeated, increasing the task delay and the network load. For this reason we plan to add to the proposed Data Consolidation (DC) techniques, resilience features in order to provide fault tolerance to the DC operation. We will consider two relative simple resiliency techniques. In the first technique, called Double Site, we select two Data Consolidation sites, the first and the second "best" according to the corresponding DC technique used and we transfer data to both sites. The task is transferred only in the first site, while the other is used as a backup in case the first site fails. In the second technique, called Half Data, we again select in the same way two DC sites, however in the second-"best" site we transfer only half of the data needed by the task. This way we reduce the load induced to the network, but also we also reduce the resiliency efficiency, since in a case of a failure in the first DC site, the rest of the data need to be transferred to the second DC site. Furthermore, we plan to propose new DC techniques in order to cope with the increase in the network load, due to the resilience techniques applied. For these reason we will concentrate on the third sub-problem of the DC, that is the selection of the paths the datasets follow in order to arrive at the DC site, investigating a number of tree-based DC schemes. Intuitively this seems like the right thing to do, since in DC we have many repository sites (the leaves, or intermediate nodes of a tree) whose data are transferred to a DC site (the root). Since we want these transfers to occur concurrently and efficiently, we try to select the tree using as optimization criterion either the time for transferring the data over a link or the load of a link, as measured by the size of data queued or under transmission at it.

IBBT will investigate a combination of checkpointing and replication. In checkpointing, the intermediate state of the running Grid jobs is saved (externally to the server it is actually running on), so as to allow quick resumption in case of a failing server. In replication, multiple instances of the same job are executed on different servers.

4.4 Joint Activity 4: Impact of Services on Optical Switch Architectures and Control

The objective of this joint activity is to investigate which options are available and feasible in an optical packet switch to support service profile and requirements as can be foreseen in future internet.

Switch architectures based on hybrid technology were studied to analyze how switch characteristics can match traffic requirements. Hybrid technology could be meant as a mix of electronic and optical technology or a mix of different optical components to implement optical gates, such as MEMS and SOAs, with different speed characteristics.

A switch architecture with wavelength converters and electronic buffers was investigated and described in [38]. A slotted context has been studied with fixed size optical payloads. When packet arrives on incoming optical channels, their header is processed to find the correct output interface. If contention for internal or external resources arises, it is solved by changing the encoding wavelength of the packet by wavelength conversion or by delaying the packet forwarding to a subsequent time slot. This latter operation is obtained by storing the packet in an electronic queue, after O/E conversion. The two option are not equivalent from a time transparency point of view, which turns into different forwarding delay. So it is important to analyze, in relation to switch dimensioning, i.e. number of input and output interfaces, number of wavelength channels, number of wavelength converters, number and size of electronic queue, which part of the optical packet traffic will pass the switch transparently and which part, instead, will need O/E/O conversion. This evaluation has been done by assuming Bernoulli input traffic and by defining suitable scheduling algorithm for packet forwarding. The related works which describes the approach are [38], [39], [40]. Sample figure from [38] are reported here.

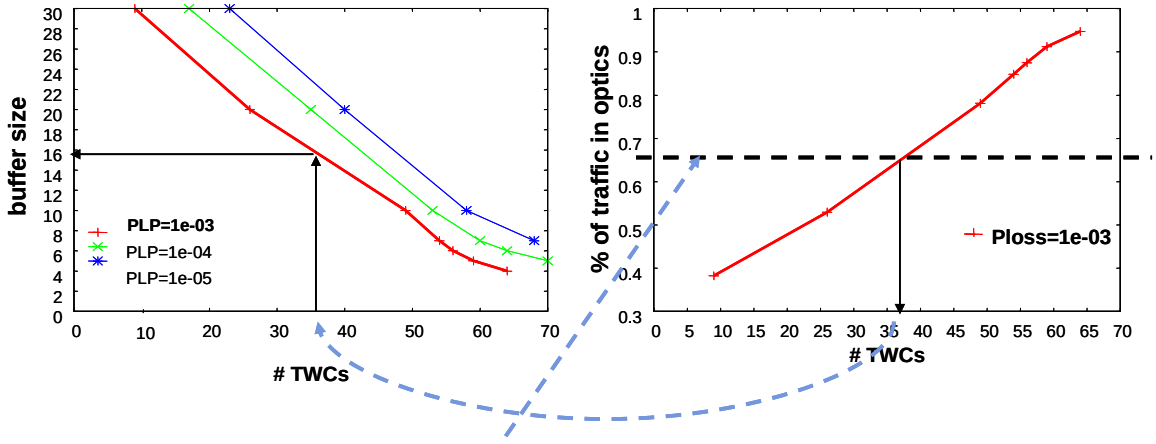

Figure 11 – Simulation results for buffer size @ different packet loss probabilities and for % of traffic in optics as a function of the number of wavelength converters, for N= 16 input and output interfaces, M=16 wavelength channels per interface, p=0.8 as a parameter of Bernoulli traffic [1].

These figures show how to define the buffer ize and the number of wavelength converters to obtain a minimum fixed value of the percentage of traffic in optics. Having fixed the packet loss probability PLP @1.e-3 and the percentage of traffic in optics, from the second plot the number of wavelength converters is known and, form the first plot, the buffer size can be found. The percentage of traffic in optics is the delay sensitive traffic present in the input traffic mix.

From the physical layer viewpoint, different cases arise depending on the contention resolution scheme that is required for each application hence physical layer modeling was adapted to investigate the different cases. UoP developed an analytical evaluation of the physical performance of the hybrid switch architecture by assuming that existing technology is utilized for the implementation, i.e. SOAs and OE-O TWCs. The evaluation resulted that packets switched through the bypass configuration (hence with no time delay) after maximum three nodes (depending on the capacity) should be converted through the multi-stage configuration so that signal regeneration is simultaneously performed. Buffered packets, although they suffer from delays, due to the regenerative properties of the buffer blocks and converters can go through a number of nodes. It is worth mentioning that from the physical layer viewpoint there is not any significant degradation due to the excess coupling loss when the fully equipped node is utilized.

Sample study has been applied to a multi-stage switch as described in [42]. The reference switch is a hybrid switch which combines multiple contention resolution scheme. Optimization of the data path, in terms of achieved throughput, and of the control plane, in terms of scheduling algorithms, and physical properties of the optical path are considered.

The other aspects of hybrid technology is represented by the presence of different type of components used for switch implementation. This study started at the University of Essex during the mobility of a researcher from UNIBO. The problem of how to map service requests on fast and slow switch sub-systems has been considered with reference to an Optical Burst Switched networks. The algorithms which controls the wavelength assignment to fast and slow connections have been defined at the edge and core nodes. A papers related to this joint activity has been submitted to OFC [41].

The plan of this joint activity for future months is to study how the hybrid technology availability can be dynamically exploited to achieve high efficiency in switch resource usage.

The plan for the next year is to design a unified control plane that can address different services. More specifically in the next year we will investigate the requirements in terms of resource allocation, traffic management and control plane extensions of unified optical transport infrastructures, under specific applications and service models. The following steps are planned towards modelling and evaluation a set of networking scenarios:

a) Select and characterize traffic patterns of evolving applications like grid networks, broadcast multimedia, streaming peer-to-peer etc. in terms of temporal and spatial characteristics
b) Define a unified control plane that can integrate the above requirements, while being interoperable with proposed technologies for wavelength routing, optical flow and dynamic burst switching
    a. Propose appropriate signalling and resource reservation procedures taking into account physical, control and service layer interfaces
    b. Propose appropriate resource allocation strategies and scheduling algorithms for specific optical switch architectures
c) Develop a simulation model to evaluate the impact of control plane protocols and resource allocation mechanisms on
    a. Overall network design cost
    b. Resource utilization
    c. Delay performance
    d. Loss performance

4.5 Joint Activity 5: Green Optical Networking

Today's ICT provides many environmentally friendly solutions. Some typical examples are:

- Providing alternative solutions for flights and road traffic by tele-working, phone and video conferencing, etc.
- Sensoring systems allowing to reduce/optimize the heating in buildings

However, ICT on its own also represents a major energy consumption factor. And this energy footprint is expected to grow significantly in the coming years, mainly driven by the steeply growing file sizes and information flows.

The goal of this joint activity is to point out the necessity of more energy efficient solutions and to investigate some possible solutions in more detail. The first step (section 4.5.1) is to accurately and objectively estimate the footprint of ICT and to identify the main contributing factors. This is a crucial step, because it will allow us to distinguish key issues from details, and to evaluate research proposals on their real benefit in terms of footprint reduction. Based on this study, some particular pain points and possible solutions to that are identified and investigated in the next sections 4.5.2 and 4.5.3.

4.5.1 Estimating the footprint of ICT worldwide and identifying main contributors

The current image of ICT is rather environmentally friendly. This is largely correct since the worldwide communication via datacom and telecom networks has transformed society and created opportunities to reduce the human impact on nature.

There is however a downside to ICT. The ubiquitousness of ICT in daily life (both private and professionally) has the drawback that the energy consumption of computers and network equipment is a significant part of the global energy consumption. It is to be expected that this share will largely increase in the coming years. Together with a growing energy price (due to shortage in fossil fuels) and the increasing awareness of the green house effect which will be translated in government policies the energy footprint of ICT will very soon be under pressure and will stimulate the demand for energy efficient solutions.

When assessing the impact of ICT on the worldwide energy production and consumption it is crucial to get a good overview of the major energy consumption factors. Firstly we give an overview of the energy consumption in the use phase of different equipment types. Secondly we give an estimate for the manufacturing phase. Finally we try to forecast the electricity consumption in the coming years.

Electricity consumption during use

When estimating the total electricity consumption the following five categories were distinguished:

1. Data centers: Servers, storage devices and network equipment, but also: cooling, backup power infrastructure (e.g. UPS systems) …
2. PCs
3. Network equipment (excluding network equipment inside data centers and PCs)
4. TV sets (including video and DVD players)
5. Other ICT equipment: All equipment not contained in the first 4 categories. (audio equipment, telephone handsets, gaming consoles, printers …)

Based on various sources we estimated the following numbers given in *Table 5*. The total power consumed by ICT is about 156 GW which is more than 8% of the global electricity consumption.

Equipment Type	Est. Consumption 2007 (GW)	Est. Annual growth rate
Data centers	26	12%
PCs	28	7.5%
Network Equipment	22	12%
TVs	40	9%

Others	40	5%
Total	156	

Table 5: Energy consumption for various equipment types

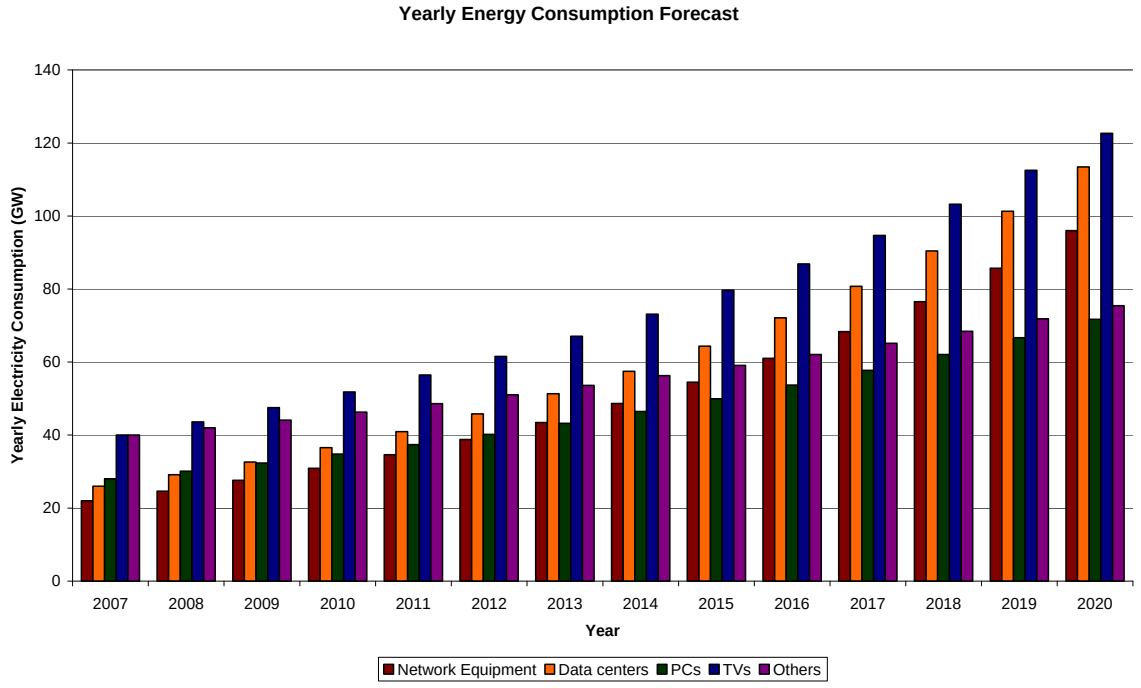

Figure 12: Electricity consumption forecast of ICT equipment during use.

When taking the annual growth rate into account we can forecast how the numbers will evolve in the next years. This is summarized in *Figure 12*. One sees that the power consumption is growing from 156 GW in 2007 to 430 GW in 2020. If we assume an energy consumption growth of 3% for all other equipment [1] this will result in a relative contribution of 14% of ICT to the worldwide energy consumption in 2020. Note that this is even excluding the manufacturing energy cost.

One should also take into account that the consumed electrical energy needs to be produced in a power plant. Currently the limited yield factor of power plants implies that 1J of electrical energy corresponds (on average) with 2.5J of primary energy.

Energy consumption for manufacturing of ICT equipment

Besides the electricity consumption in the use phase of the equipment one needs to take the manufacturing process into account as well. In [2] a modeling exercise has been carried out in detail leading to an average of 1550 MJ of electrical energy and 4850 MJ of non-electrical primary energy sources to manufacture a typical PC configuration.

With the electricity production yield factor of 40% the 1550 MJ of electrical energy requires 3875 MJ of primary energy. This leads to a total of about 8700 MJ per produced PC.

When assuming an economical lifetime of 4 years for a PC the average energy consumption during the use phase is about 8800 MJ. This means that the energy needs during the manufacturing phase are comparable to the needs during the use phase. This conclusion depends obviously on the type of equipment. More specifically it depends on the energy intensity of the manufacturing process and the expected economical lifetime. It will be important to take these factors into consideration when designing new energy efficient solutions.

4.5.2 Energy saving potential by selective turning off of network elements

In our work, we consider a wide area network scenario. Given the network topology and a traffic demand, we evaluate the possibility of turning off some elements (nodes and links) under connectivity and Quality of Service (QoS) constraints. The goal is to minimize the total power consumption of a large network, in which usually resource overprovisioning is large, by using some simple optimization algorithms.

Problem formulation

An informal description of the design problem studied is the following: Given i) a physical network topology comprising routers and links, in which links have a known capacity, ii) the knowledge of the average amount of traffic exchanged by any source/destination node pair, iii) the maximum link utilization that can be supported, iv) the power consumption of each link and node; Find the set of routers and links that must be powered on so that the total power consumption is minimized; Subject to flow conservation and maximum link utilization constraints.

As we detailed in [43], the problem can be formulated using an Integer Linear Programming (ILP) methodology. Unfortunately, solving the ILP is not viable, since it falls into the multi-commodity flow class, which is known to belong to the NP-hard class. Exact solutions can be found only for some trivial cases. We therefore propose some simple heuristics in order to solve the problem also for large networks.

Algorithms

The algorithms we propose consider a network in which all elements are powered on, so that $x_{ij} = 1 \forall\, i, j$ and $y_i = 1 \forall\, i$, being x_{ij} the state of link from node i to node j, and y_i the state of node i. Each algorithm then iteratively tries to switch off each element (either a node or a link).

At each step, traffic is then rerouted on the shortest path for each (s, d) pair to verify a connectivity constraint, i.e. each (s,d) has to be connected through a path of on devices. We impose also a QoS constraint, i.e. the link utilization must not exceed a given threshold. If no violation is present, then the selected element is powered off. Fig. 1 reports a schematic description of the algorithms.

We implemented two different kinds of algorithms: node-oriented and link-oriented heuristics. We expect that it is more difficult to turn off a node than a single link, but the energy saving introduced in the former case is much larger, as reported in [1]. The two heuristic approaches are therefore combined so that the nodes are checked first, and then links are possibly powered off at a second stage.

Several policies can be adopted to iterate through the node set. We implemented the following ones:

- random (R)
- least-link (LL)
- least-flow (LF)

According to each heuristic, the node set is first sorted considering a given rule before iterating through all the nodes. In particular, the least-link heuristic sorts the nodes according to the number of links that are sourced and sinked at each node, so that nodes with a smaller number of links are checked first.

The least-flow heuristic takes instead into account first the nodes with the smallest amount of information flowing through them.

```
// node optimization
sort_nodes(vectnodes);
for (i=0; i<N; i++) {
                disable_node(vect_nodes[i]);
                compute_all_shortest_path();
                compute_all_link_flow();
          if (check_paths() == false) {
                    enable_node(vect_nodes[i]);
                    continue;
                }
                if (check_flows() == false) {
                    enable_node(vect_nodes[i]);
                    continue;
                }
}
```

```
// link optimization
sort_links(vect_links);
for (j=0;j<L; j++) {
                disable_link(vect_links[j]);
                compute_all_shortest_path();
                compute_all_link_flow();
          if (check_paths() == false) {
                    enable_link(vect_links[j])
                    ;
                    continue;
                }
                if (check_flows() == false) {
                enable_link(vect_links[j]);
                continue;
                }
}
```

Figure 13. The pseudo-code description of the proposed algorithms.

Finally, the random heuristic sorts nodes in random order.
Similarly, considering link heuristics, we implemented two algorithms:

- least-flow (LF)
- random (R)

which leverage on the same intuition as the corresponding node sorting heuristics: the least-flow policy sorts links in increasing order of carried flow.
All possible node/link sorting combinations have been studied. Besides these heuristics, we also tested the corresponding ones in which a decreasing order is adopted. Since they all perform consistently worse, we decide not to consider them in this paper.

Performance comparison
In order to assess the performance of the proposed heuristics, we consider a simple Wide Area Network scenario. The goal is to show that, for a given (static) traffic demand, it is possible to power off some network elements, and to still guarantee full connectivity between sources and destination, while enforcing that the link utilization remains smaller than a QoS threshold.
We suppose the network follows a hierarchical topology, which is typical of WANs. All links are supposed to be bidirectional links, so that if link (i, j) exists, then link (j, i) exists as well. Three levels of nodes are considered: core, edge and aggregation nodes.
The network core is composed by few nodes that are highly interconnected by means of high-capacity links. Each link connects nodes which may be also geographically far away, e.g., optical links connecting different cities.
The edge nodes are instead used to interconnect aggregation nodes to the core nodes. Links have middle-range capacity, i.e., smaller capacity than the one of links interconnecting core nodes. Each edge node is connected to some of the closest core nodes, and to other edge nodes. One or more edge nodes can be present in cities, and they collect traffic from aggregation nodes spread within the city boundaries.
The last level of nodes is composed by the aggregation nodes, to which users are directly connected. A Digital Subscriber Line Access Multiplexer (DSLAM) is a typical example of aggregation node. Each node is dual-homed, i.e., it is connected to the closest pair of edge nodes (to guarantee alternate paths in case of failure). The links that connect aggregation nodes to edge nodes have low capacity, i.e, smaller capacity than the one of links interconnecting edge nodes.
Considering the link capacity assignment policy, three classes of links are defined: high, middle-range and low capacity links. Each class has a minimum capacity c_{ij}^{min} constraint, that was selected to be 15, 5, and 1 units of traffic respectively. Minimum link capacities are also used as link routing weights, so that the routing cost is inversely proportional to the link capacity. This is commonly adopted to force the traffic to be routed through the edge and the core nodes, rather than through aggregation nodes (which are connected by means of low capacity links). A simple minimum cost path is considered as routing algorithm, similarly to what is commonly adopted in the Internet. Furthermore, a QoS constraint is considered, that forces the total traffic flowing through a link to be smaller than an over provisioning factor $\beta = 0.5$. Therefore, after routing all traffic, link capacities are finally assigned so that:

$$c_{ij} = max(f_{ij} / \beta , c_{ij}^{min})$$

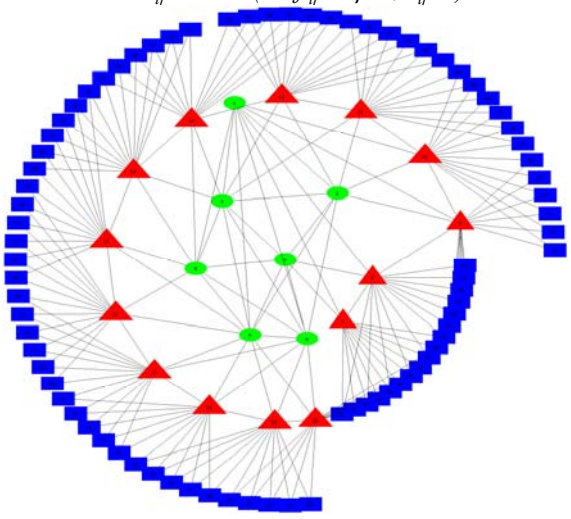

Figure 14: An example of random topology

Results proposed in our work have been obtained considering randomly generated topologies in which 160 nodes are considered. In particular, 10 core nodes, 30 edge nodes, and 120 aggregation nodes are considered. Nodes are assumed to be placed on a plane. Core nodes are randomly connected to other core nodes with probability $p = 0.5$. Each edge node is then connected to the two closest core nodes and to another randomly selected close edge node. Finally, aggregation nodes are connected to the two closest edge nodes. An example of the topology obtained is presented in Figure 2. Aggregation, edge and core nodes are represented by squares, triangles and circles respectively.
Only aggregation nodes are traffic sources and sinks. For the sake of simplicity, we consider a uniform traffic pattern, so that $t^{sd} = U[\,0.5\,,1.5\,]$ units of traffic if s, d are aggregation nodes; $t^{sd} = 0$ otherwise.

A. Simulation Results

For each considered heuristics, we collected the percentage of links and nodes that are turned off, η_L and η_N respectively. This test was repeated on 20 randomly generated topologies and traffic patterns. *Figure 15* show the comparison of the different heuristics by reporting η_L and η_N respectively. Bars report mean values, while the error bars show the standard deviation. Labels on the x-axis report the node-link heuristic combination. A maximum link load factor $\alpha = 0.8$ was considered. We consider the same traffic demand, so that the network must guarantee to transport the same amount of traffic.
We report also an upper-bound obtained by relaxing the maximum link utilization constraint, so that only the flow conservation constraint is imposed. This is equivalent to find the minimum set of nodes and links that permit to route all the offered flows. This allows to better assess the impact of the QoS constraint, and the quality of the solutions generated by the proposed heuristics.
Considering the links off (left part of *Figure 15*), we can see that it is possible to actually turn off about 25% of links in the considered network and traffic scenario. All node selection heuristics show very similar results, while a larger impact of the link heuristics is shown. Indeed, random link selection heuristics (R-R, LF-R, LL-R) show consistently worse results compared to the least flow selection policy (LF-LF, LL-LF, R-LF). Notice also that the best performing algorithm is only 7 percentage points below the upper bound, which shows that little improvement is possible.

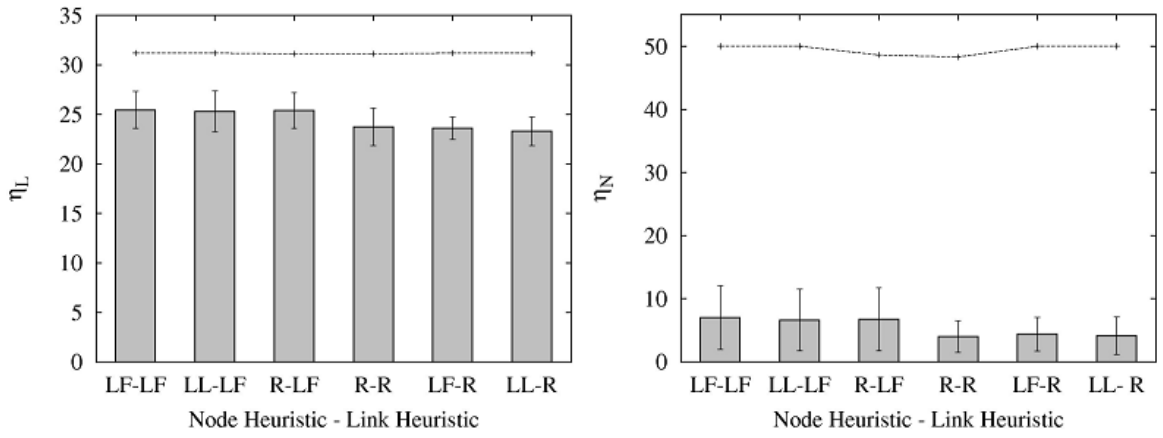

Figure 15: *Comparison of the percentage of links (left) and nodes (right) switched off considering different algorithms*

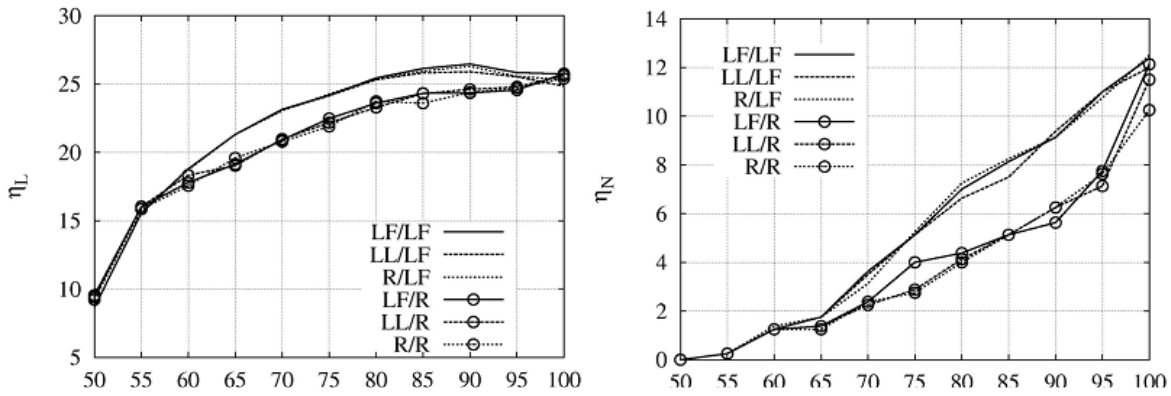

Figure 16: *Percentage of links (left) and nodes (right) versus* α .

Figure 15 on the right reports instead the average number of nodes that different heuristics are able to switch off: in this case, only 5-10% of nodes can actually be turned off, since a node can be powered off if all its links are switched off. Also in this case there is little impact on the node selection policy, while the LF link selection policies generally perform better. Notice that the upper bound is much higher than any admissible solution, suggesting that the QoS constraint cannot be relaxed.

B. Parameter Impact

We performed a study on the impact of the α parameter, in order to observe the possible range of network elements that can be successfully switched off while guaranteeing a maximum offered load on links. For sake of simplicity, only mean values are reported for each of heuristic combination.

Figure 16 on the left reports the number of links switched off for $\alpha \in [0.5, 1]$ in the considered scenario. All algorithms show large improvements for α up to 0.8; after that, little improvement is noticeable, and a final minor decrease in the average percentage of links that can be turned off is observed for values of $\alpha > 0.8$. This is due to the fact that when α is higher, a larger number of nodes can actually be switched off (see *Figure 16* on the right). This reduces the freedom of turning off other links, since not many alternate paths remain available. *Figure 16* on the right shows the average percentage of nodes that are switched off for different algorithms. Similar considerations hold also in this case.

Conclusions and future work

In our work we faced a network design problem. We deviated from the traditional formulations of the problem, in which the objective function is to minimize cost or maximize performance, by considering the minimization of the total power consumed by the network as objective function, while connectivity and maximum link utilization are taken as constraints.

Simple heuristics have been proposed, and their performance assessed considering some simple yet realistic traffic and network scenarios. Results (although dependent in absolute values from the chosen scenario) show that it is possible to switch off both full nodes and links, so that the total network power consumption can be reduced.

As future work, we plan to evaluate the power saving that can be achieved during off-peak hour, in which the traffic demand is much smaller, so that is possible to reroute traffic on the spare capacity and switch off a large number of nodes and links.

4.5.3 Green Routing Protocol

This work considers energy efficiency as one of the dominant factors in the routing and path computation process. Routing algorithms have been extended to include this factor and find the optimum path in terms of energy consumption. The proposed protocol takes into consideration several factors such as geographical location of nodes and links, instant load on the nodes, length of the links as well as data bitrate, attenuation and other physical layer impairment requirements.

Geographical location of nodes and links: Here we consider the geographical location of network nodes, links and end IT devices as this can play a significant role in determining if these nodes can be powered by green energy resources. The amount of power that can be obtained from green resources for each link and each node depends by their geographical location; the remaining required power can be supplied through the traditional non-green energy resources. The amount of power that is provided through green sources determines the "Green Coefficient" of each node and link. As an example if a node can supply all of its required power through renewable source of energy will have a "Green Coefficient" of 100% or 1.

Instant node load: The amount of load that is applied to switches and routers can vary in time. A switch will use a larger number of interfaces when is highly loaded. In the case of electronic switches the number of active O-E-Os ports will be a determining factor concerning the power consumption of a node.

Length of the link: The length of the link is a dominant factor in determining the amount of power that an optical link may consume. Increasing the link length will increase the number of active components need to be placed on the link.

Algorithm

The proposed "Green routing protocol" is a global routing protocol. Each node has the graph topology of the network that is used for route computation. It is also a dynamic routing protocol because it is able to update topology and the load changes. The mathematical modeling which has been considered in the proposed Green protocol calculates the instant cost for nodes and links and then applies this cost to Dijkstra K shortest path algorithm to find the first k candidate paths. K shortest path algorithm is based on Dijkstra algorithm but, offers some flexibility such as calculating arbitrary number of paths that may have different accumulative costs; so it is possible to apply a specific policy on the calculated paths. Calculated paths are in the order from minimum cumulative cost to maximum cumulative cost [45]. Each of the nodes in the network may be partially provided with green energy resources. Depending on the load, a node can be powered fully or partially by green power resources. The calculated non-green power for each node is directly associated with the applied load or the number of interfaces (WDM transponders) which utilized in the network. Therefore the cost will vary in time with the network load on the network or the number of utilized wavelengths.

Each node has its maximum power consumption when it is fully loaded and all its associated wavelengths or O-E-O s are utilised. The following equations have been considered to approximately calculate the non-green power consumption for each node:

Node instant power consumption = (Number of utilized wavelengths of a given node / Total number of wavelengths of a given node)*max power consumption for a given node

To calculate the non-green power consumption of a node, A will be calculated as following:

A = Node instant power consumption – (Green coefficient for the given node * max power consumption of the given node)

If calculated "A" has a positive value then Node instant Non Green Power consumption is equal to "A" otherwise is equal to zero.

In order to calculate non-green power consumption this algorithm consider following assumptions:

For each link, the amount of power consumption depends on the number of active elements such as optical amplifiers and dispersion compensators. Considering ΔL = 80 Km (which is a typical figure in telecom networks) [50] as the maximum distance between each optical amplifier/dispersion compensator the number of active elements will be obtained. The non green power consumption for each link will be calculated as following:

Link Non Green Power Consumption = (1 – Green coefficient for a given link) * total power consumption for that link.

The Dijkstra algorithm just considers non negative cost for each link to find the k shortest paths, but as it has been mentioned in this algorithm both links and nodes have their own cost which is equal to the amount of non green power consumption in any of them. In order to solve this problem each node`s Non Green Power Consumption should apply to all the links which are connected to it.

Before calculating a new path, the "Node instant Non Green Power consumption" will be calculated for each node. The cost which for each link is:

Link cost = Link Non Green Power Consumption + (first connected Node instant Non Green Power consumption /2) + (second connected Node instant Non Green Power consumption /2)

Network Model

The network under study is an optical agile opaque WDM network, which consists of 16 nodes that are connected by 25 optical fibers and follows the NSFNET topology with the channel bit-rate of 10 Gbit/s. Fig. 1 illustrates the network topology. In this long haul opaque network, a unique node model is designed and employed for all the excising nodes in the network using the Opnet Modeler software. The node model generates request packets, assign destination addresses, and process received packets. So each node can act as a source/destination or as a router.

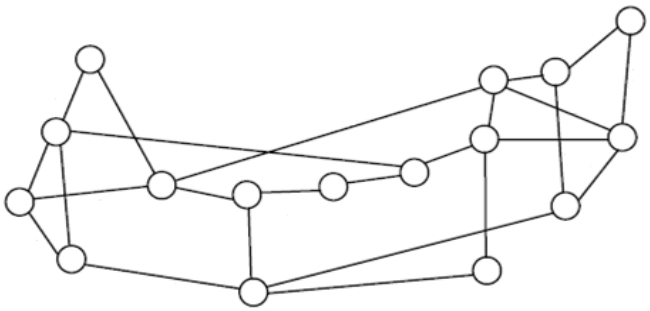

Figure 17: NSFNET topology with 16 nodes and 25 links [7]

O-E-O conversion is employed in all the nodes in order to regenerate and retransmit signals to the next hop.

This network provides the ability to set up and tear down light paths as required. The optical amplifiers that are placed on the links are backward-pumped distributed Raman amplifiers. 12 km of dispersion compensation fiber

is also deployed in each link. The nodes in the above network are placed in long distance from each other (worse case 2815.041 Km link length). Wavelength convertors at every node allow virtual lightpath functionality and eliminates wavelength blocking in the network.

Scenarios

Three different scenarios have been defined; and the focus is on comparing the proposed green routing protocol with Dijkstra minimum hop routing protocol in terms of non green power consumption in a high speed network with nodes and links that are fully or partially provided with renewable resources of energy. A green coefficient which is set for each node and each link determines how much of their required power is going to supplied through renewable resources of energy. In the first, and third scenarios 31 percent of the network`s total required power could be supplied through green resources, while in the second scenario only 10 percent could be supplied via green energy resources. In all the scenarios 34% of the nodes` total required power and 29% of links` total required power is supplied through green energy resources. The difference between each of these scenarios relates to the green coefficients that are going to be set for each of the links and nodes.

This simulation compares Green and minimum hop protocols for different amount of load on the network.

All the scenarios are executed for k = 1 and k = 2, the value of k indicates the number of paths that are going to be calculated for a given source and destination.

Scenario 1

In this scenario we try to allocate green energy resources to the links and nodes which are more likely to be used by the minimum hop protocol for a given source and destination when a total of 30% green energy resources are provided. As it can be observed from *Figure 18* and *Figure 19* there is a remarkable difference between the amount of non green power consumption of the network for Green and minimum hop routing protocols. This scenario can offer 62% improvement for 20% load. Green protocol can even offer 37% improvement, for high 70% load in this scenario; which is quite remarkable achievement for high loads. Quantitatively, the best achievement is 4.4 MW (for 70% load). In addition, in 10% load Green protocol offers 711.1 KW improvements which is again a considerable achievement for low loads. In lower loads (10% - 30%) the performance of the proposed Green protocol is quite similar for k = 1 and k = 2 in all defined scenarios. This is due to the fact that for lower loads there are available channels in the network to select an optimum path and therefore there is no need to use an alternative path.

Scenario 2

This scenario is designed in a way to make the same assumption with scenario 1 (in green coefficients setting); when a total of 10% green energy resources are provided. As can be seen in *Figure 20* and *Figure 21* the difference between non green power consumption for min hop and green protocol is relatively smaller than scenario 1 .The Green protocol in this scenario can offer a maximum of 41% improvement for 10% load. It also provides minimum 8% and 9.2 % improvement for 70% load when k is one and two relatively. In terms of quantity, the Green protocol offers maximum 1.5 MW and 1.3 MW for 70% load when k is one and two relatively. It presents 630.7 KW improvements for 10% load as well.

Scenario 3

This scenario is trying to provide the nodes and links that are more likely to be selected by minimum hop protocol with renewable resource of energy. *Figure 22* and *Figure 23* Show that the difference between the amount of non green power consumption for minimum hop and Green routing protocol is considerably decreased. The highest improvements that have been attained are 4.7% and 4 % for 30% and 20% load respectively (for both k = 1 and k = 2). Quantitatively, in 70% load for k = 1 and k= 2, results show 24.1 KW and 154 KW improvement respectively. The better achievement of the Green protocol for k = 2 is at the expense of 280.126 KW more non green power consumption due to the fact that for k = 2 Green protocol cannot offer its optimum performance because the router may use the second calculated path as an alternative when the first path is busy. Needless to say, even though k = 2 consumes more non green power it can offer an alternative path when optimum path is busy. Therefore it is possible to employ the Green protocol to examine network

topology and find the links and nodes need to be supplied with renewable energy sources in order to make the performance of minimum hop or even other protocols close to Green routing protocol.

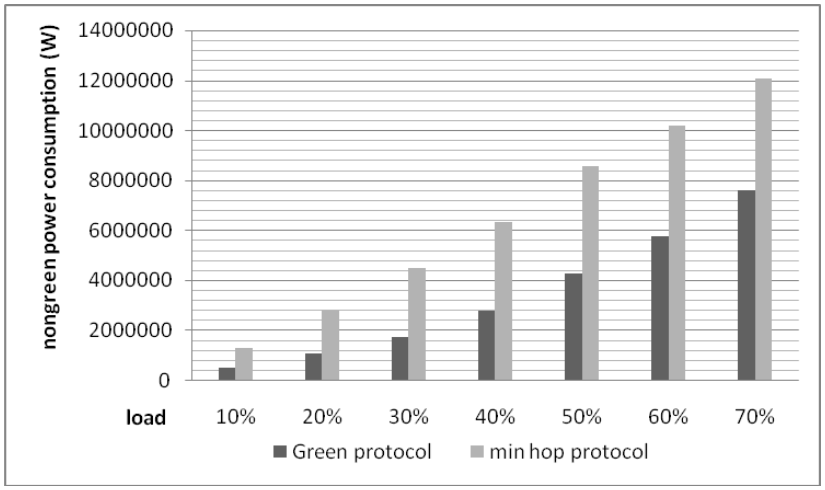

Figure 18: Non green power consumption for Green and minimum hop protocol for k = 1 in scenario 1

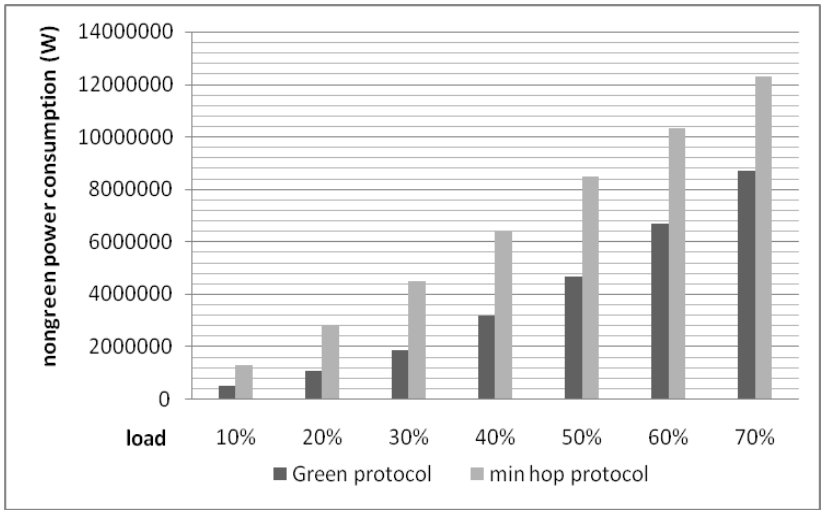

Figure 19: Non green power consumption for Green and minimum hop protocol for k = 2 in scenario 1

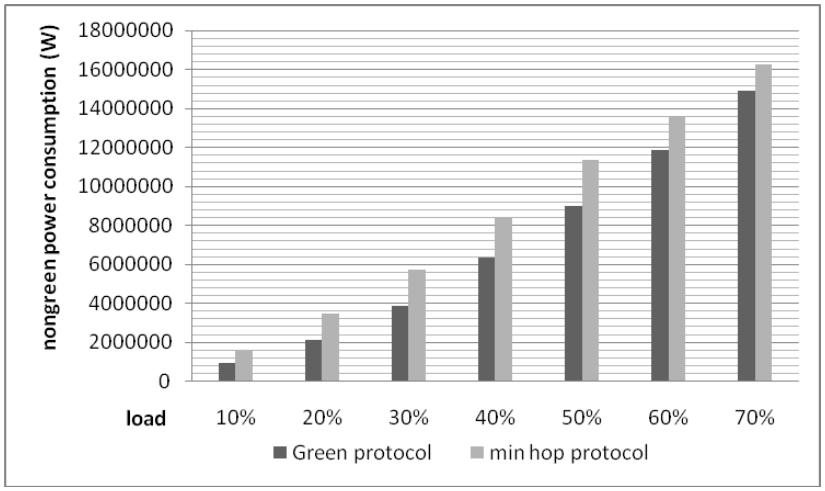

Figure 20: Non green power consumption for Green and minimum hop protocol for k = 1 in scenario 2

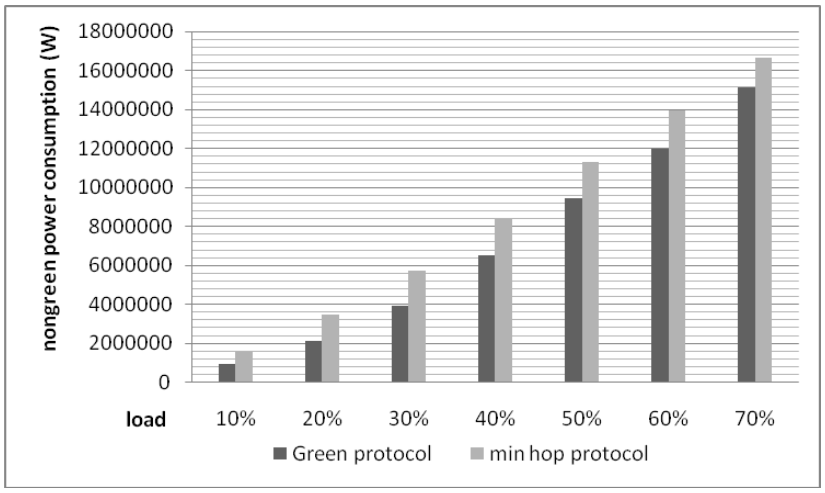

Figure 21: Non green power consumption for Green and minimum hop protocol for k = 2 in scenario 2

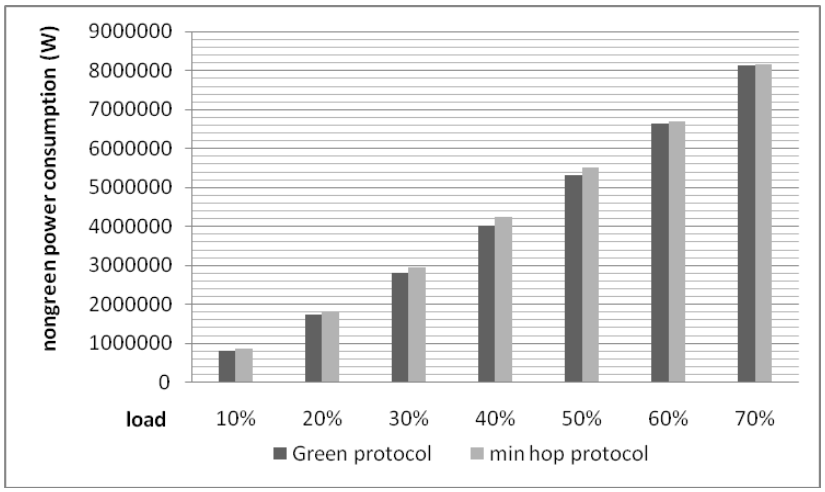

Figure 22: Non green power consumption for Green and minimum hop protocol for k = 1 in scenario 3

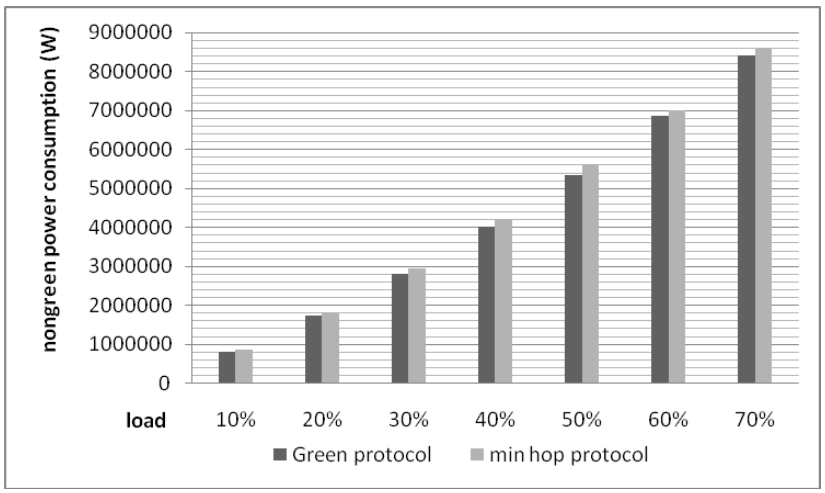

Figure 23: Non green power consumption for Green and minimum hop protocol for k = 2 in scenario 3

Conclusion

The results of this simulation demonstrate that the Green routing protocol is quite distinct from conventional minimum hop routing protocols in terms of non green power consumption. The proposed Green protocol offers up to 62 % improvement in terms of non green power consumptions in the scenarios examined here. In addition, the performance of the Green routing protocol is more beneficial when a network is provided with more renewable resources. The aggregate difference between the amounts of non green power consumption for two mentioned protocols depends on, which nodes and which links in the topology have more access to renewable resources. In other words the geographical locations for Green nodes and links are determining the amount of the difference between the two mentioned routing protocols. In reality, setting of the nodes and links might be in a way that the difference becomes remarkably high. In contrast, the allocation of renewable resources to nodes and links may be in a way that two routing protocols roughly perform similarly in terms of non green power consumption. Needless to say, in large scale scenarios, this difference would be relatively high. Future studies will extend the Green routing protocol to consider further factors and technologies in its path calculation. We will also attempt to merge the Green routing protocol with other conventional protocols.

5. References

[1] Energy Information Administration: "Official Energy Statistics from the U.S. Government", http://www.eia.doe.gov/.

[2] E. Williams, "Energy Intensity of Computer Manufacturing: Hybrid Assessment Combining Process and Economic Input-Output Methods", Environmental Science & Technology, Vol. 38, No. 2, pp. 6166-6174 November 2004.

[3] Kyriakos Vlachos and Apostolos Siokis, "SO-SOON: A Self-Organized, Service-Oriented Optical Network employing Islands of Service Transparency" submitted to IEEE Network Magazine.

[4] Kyriakos Vlachos and Apostolos Siokis, "A Service-Transparent and Self-Organized Optical Network Architecture", submitted to ICC 2009

[5] Chinwe Abosi, Reza Nejabati and Dimitra Simeonidou, "A Service Plane Architecture for Service Oriented Optical Networks" submitted to OFC 2009.

[6] Chinwe Abosi, Reza Nejabati and Dimitra Simeonidou, "A Service Plane Architecture for Future Optical Internet" submitted to ONDM 2009.

[7] Chinwe Abosi, Reza Nejabati and Dimitra Simeonidou, "A Service Plane Architecture for Future Optical Internet" submitted to Journal of Optical Networking special issue on Future Optical Internet.

[8] C. Develder, B. Dhoedt, B. Mukherjee, P. Demeester, "On dimensioning optical grids and the impact of scheduling", to appear in Photonic Network Communications.

[9] T. Stevens, M. De Leenheer, C. Develder, B. Dhoedt, K. Christodoulopoulos, P. Kokkinos, E. Varvarigos, "Multi-Cost Job Routing and Scheduling in Grid Networks", to appear in Future Generation Computer Systems.

[10] 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.

[11] 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.

[12] 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.

[13] M. Ali, L. Tancevski, "Impact of polarization-mode dispersion on the design of wavelength-routed networks", IEEE Photonics Technology Letters, Volume 14, Issue 5, pp. 720 - 722,May 2002.

[14] B. Van Caenegem, W. Van Parys, F. De Turck and P.M. Demeester, "Dimensioning of survivable WDM networks", IEEE Journal on Selected Areas in Communications, vol. 17, no. 7, 1146-1157 (1998).

[15] T. Carpenter, D. Shallcross, J. Gannett, J. Jackel, A. Von Lehmen, "Maximizing the Transparency Advantage in Optical Networks", in Proceedings of Optical Fiber Communication Conference, Vol. 2, Iss. 23-28, pp. 616-617, 2003.

[16] B. Chatelain, S. Mannor, F. Gagnon and D.V. Plant, "Non-Cooperative Design of Translucent Networks", in Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '07), pp. 2348-2352, Washington, DC, November 2007.

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

[18] A. Filho and H. Waldman, "Strategies for Designing Translucent Wide-Area Networks", in Proceedings of International Microwave and Optoelectronics Conference (IMOC), 931-936, 2003.

[19] A. Fumagalli and M. Tacca, "Differentiated reliability (DiR) in wavelength division multiplexing rings," IEEE/ACM Trans. Netw. 14, 159–168 (2006).

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

[21] R. Iraschko, M. MacGregor, and W. Grover, “Optimal Capacity Placement for Path Restoration in STM or ATM Mesh-Survivable Networks,” IEEE/ACM Trans. Netw. 6, 326- 336, (1998).

[22] R. Iraschko, W. Grover, “A Highly Efficient Path-Restoration Protocol for Management of Optical Network Transport Integrity,” IEEE J. Sel. Areas Comm. 18, 779-794 (2000).

[23] M. Kodialam, T. V. Lakshma n, “Dynamic Routing of Bandwidth Guaranteed Tunnels with Restoration,” in Proc. of IEEE conference on Computer Communication (IEEE, 2000), pp. 902-911.

[24] A. Morea, H. Nakajima, L. Chacon, Y. Le Louedec, J.-P. Sebille, “Impact of the reach distance of WDM systems on the cost of translucent optical networks”, in Proceedings of Telecommunications Network Strategy and Planning Symposium, pp. 321-325, June 2004.

[25] Z. Pandi, M. Tacca, A. Fumagalli, and L. Wosinska, “Dynamic provisioning of availability constrained optical circuits in the presence of optical node failures,” J. Lightwave Technol. 24, 3268–3279 (2006).

[26] R. Ramamurthy, Z. Bogdanowicz, S. Samieian, D. Saha, B. Rajagopalan, S. Sengupta, S. Chaudhuri, K.. Bala, “Capacity Performance of Dynamic Provisioning in Optical Networks”, J. Lightw. Tech. 19, 40-48 (2001)

[27] B..Ramamurthy, H. Feng, D. Datta, J.P. Heritage, B. Mukherjee, “Transparent vs. opaque vs. translucent wavelength-routed optical networks”, in Proceedings of Optical Fiber Communication Conference, pp. 59-61, 1999.

[28] S. Ramamurthy and B. Mukherjee, “Survivable WDM Mesh Networks, Part II Restoration,” in Proceeding of IEEE conference on Communications (IEEE, 1999), pp. 2023-2030.

[29] S. Ramamurthy and B. Mukherjee, “Survivable WDM Mesh Networks, Part I - Protection,” in Proc. IEEE Conf. on Computer and Communication Societies (IEEE, 1999), pp. 744-751.

[30] M. Savasini, P. Monti, M. Tacca, A. Fumagalli and H. Waldman, “Regenerator Placement with Guaranteed Connectivity in Optical Networks”, in Proceedings of 11th International IFIP Conference on Optical Network Design and Modeling, pp. 438-447, May 2007.

[31] G. Shen, W. Grover, T. Cheng, and S. Bose, "Sparse placement of electronic switching nodes for low blocking in translucent optical networks", Journal of Optical Networking, Vol. 1, Iss. 12, pp. 424-441, December 2002.

[32] J.M. Simmons, “Network Design in Realistic “All-Optical” Backbone Network”, IEEE Communications Magazine, Vol. 44, Issue 11, pp. 88-94, November 2006

[33] P. Thysebaert., et.al."Scalable dimensioning of Resilient lambda Grids" Future Generation Computer Systems, Volume 24 Issue 6, 2007

[34] W. Van Parys, P. Arijs, O. Antonis,P. Demeester, "Quantifying the benefits of selective wavelength regeneration in ultra long-haul WDM networks", in Proceedings of Optical Fiber Communication Conference and Exhibit (OFC '01), 2001.

[35] X. Yang, B. Ramamurthy, “Sparse Regeneration in Translucent Wavelength-Routed Optical Networks: Architecture, Network Design and Wavelength Routing”, Springer Journal of Photonic Network Communications, 2005.

[36] E. Yetginer, and E. Karasan, “Regenerator Placement and traffic Engineering with Restoration in GMPLS Networks”, Photonic Network Communications, Vol. 62, pp.139-149, 2003.

[37] H. Zhang and A. Durresi, “Differentiated multi-layer survivability in IP/WDM networks,” in Proceedings of IEEE/International Federation of Image Processing (IFIP) Symposium on Network Operations and Management (IEEE, 2002) pp. 681–694.

[38] Carla Raffaelli, Michele Savi, “QoS Aware Optical Packet Switch with Shared Electronic Buffers” Broadnets/GOSP 2008, invited paper, Londra, Settembre 2008.

[39] C. Raffaelli, M. Savi, A. Stavdas, “Multi-Stage Shared-per-Wavelength Optical Packet Switch: Heuristic Scheduling Algorithm and Performance”, accepted for publication on IEEE/OSA JLT

[40] Carla Raffaelli, Michele Savi, Alexandros Stavdas, Hybrid contention resolution in optical packet switching, submitted to OFC 2009.

[41] Michele Savi, Georgios Zervas, Yixuan Qin, Valerio Martini, Carla Raffaelli, Fabio Baroncelli, Barbara Martini, Piero Castoldi, Reza Nejabati, Dimitra Simeonidou, “Data-Plane Architectures for Multi-Granular OBS Network”, submitted to OFC 2009.

[42] C. Raffaelli, M. Savi, T. Politi, A. Stavdas, "High Capacity Hybrid Optical Packet Switch: Data and Control Plane Design", submitted to IEEE Journal on Selected Areas in Communications, Optical Communication and Networking Series.

[43] L. Chiaraviglio, M. Mellia, F. Neri, Energy aware Networks: Reducing Power Consumption by Switching Off Network Portions, GTTI 2008, Italy.

[44] Huang Y, Heritage J.P, Mukherjee B, 'Connection Provisioning With Transmission Impairment Consideration in Optical WDM Networks With High-Speed Channels' , Journal of Lightwave Technology, Vol. 23, No. 3

[45] Opnet tutorial. [Online] http://www.opnet.com