An analysis of Regenerator Placement strategies for a Translucent OBS network architecture

Oscar Pedrola, *Student Member, IEEE,* Davide Careglio, *Member, IEEE,* Miroslaw Klinkowski, *Member, IEEE,* and Josep Solé-Pareta

***Abstract*—Most research works in optical burst switching (OBS) networks do not take into account the impact of physical layer impairments (PLIs) either by considering fully transparent (i.e., using optical 3R regeneration) or opaque (i.e., electrical 3R regeneration) networks. However, both solutions are not feasible for different reasons. In this paper, we propose a novel translucent OBS (T-OBS) network architecture which aims at bridging the gap between the transparent and opaque solutions. In order to evaluate its performance, a formulation of the routing and regenerator placement and dimensioning problem (RRPD) is presented. Since such formulation results in a complex problem, we also propose several alternative heuristic strategies. In particular, we evaluate the trade-off between optimality and execution times provided by these methods. Finally, we conduct a series of simulation experiments that prove that the T-OBS network model proposed effectively deals with burst losses caused by the impact of PLIs and ensures that the overall network performance remains unaffected.**

I. Introduction

WITH the advent of ultra high bandwidth access systems such as the gigabit passive optical network (GPON) and the next generation mobile networks (i.e., long term evolution (LTE) and 4G), we are forced to move into the next phase of broadband backbone technologies. Indeed, multi-industry initiatives have already started the definition of new business models with the aim of accelerating mass adoption of new devices and services such us video streaming/conferencing, HDTV, VoIP and VoD.

After becoming a real networking layer, optical technology and Optical Transport Networks (OTNs) in particular are the preferred candidates to meet the demands of such applications. Recent advances in optical technologies are fostering the deployment of fully transparent OTNs which involves all-optical switching and full end-to-end optical paths. Nevertheless, the Physical Layer Impairments (PLIs) of the optical domain and, concurrently, the lack of effective all-optical regeneration devices prevent it from taking place, at least, in the short-medium term [1]. For that very reason, translucent OTNs are the ideal yet feasible candidates for bridging the gap between opaque (i.e., with Optical-Electrical-Optical (O/E/O) conversion at each node) and transparent networks. Indeed, translucent networks combine features of both opaque and transparent networks allowing signal regeneration only at selected points in the network [2].

However, for translucent OTNs to attract continuing interest, they should be designed in such a way that both the construction cost and power consumption is minimized. Both constraints are clearly related to the number of O/E/O regenerators deployed across the network, and therefore, their number must be reduced as much as possible. For this very reason, the definition of algorithms either for Regenerator Placement (RP)[3] or for Routing and Regenerator Placement (RRP) (see e.g., [4], [5]), if routing constraints are added to the problem, is essential to the problem's success. These techniques are aimed at minimizing the number of regenerators deployed in the network by finding their optimal location.

Due to the maturity of the technology that Wavelength Switched Optical Networks (WSONs) require (e.g., Reconfigurable Optical Add-Drop Multiplexers (ROADMs) and Optical Cross-Connects (OXCs)), translucent WSONs have been the first to receive the attention from the research community. Indeed, protocol extensions and requirements to take into account the presence of PLIs in WSONs are currently under development within IETF [6]. Nonetheless, the inflexibility and coarse granularity of WSONs motivates the development of sub-wavelength switching technology. Nowadays, in fact, technologies like Optical Packet Switching (OPS) and Optical Burst Switching (OBS) [7] which were initially proposed ten years ago, are re-gaining much of the research interest together with more recent proposals such as Optical Data-unit Switching (ODS) and Optical Flow Switching (OFS). Among these sub-wavelength solutions, in this paper, we focus on the OBS switching paradigm. In OBS, edge nodes aggregate the client data into bigger data containers called bursts which, once ready, are launched optically into the network. Together with each burst, a control information called Burst Control Packet (BCP) is transmitted out-of-band and delivered to the core nodes with some offset-time prior to the burst. The offset-time provides the necessary time budget to route the incoming burst properly, that is, the amount of time required for both the electronic processing of the BCP and the reconfiguration of the optical switching matrix of the node. In such a way, a wavelength is booked on-the-fly, only temporarily, and can be reused afterwards by any other burst (i.e., the resources are shared among all nodes and subject to statistical multiplexing).

In this paper, we propose a novel Translucent OBS (T-OBS) network architecture, which we first presented as a preliminary work in [8], and derive a PLI model and some design principles. Afterwards, we deal with the RRP prob-

Oscar Pedrola, Davide Careglio and Josep Solé-Pareta are with the Department of Computer Architecture, Universitat Politècnica de Catalunya, Barcelona, Catalunya 08034, Spain (e-mail: opedrola@ac.upc.edu; careglio@ac.upc.edu; pareta@ac.upc.edu).

M. Klinkowski is with the National Institute of Telecommunications, 04-894 Warsaw, Poland (e-mail: M.Klinkowski@itl.waw.pl).

lem using such PLI model as a constraint. Nonetheless, in contrast to the classical RRP problem found in WSONs, where there exists a one-to-one correspondence between optical path/connection and electrical regenerator, in T-OBS the access to the signal regenerators is, like any other resource, subject to statistical multiplexing and so the introduction of an additional dimensioning phase which eventually extends the problem to the Routing and Regenerator Placement and Dimensioning (RRPD) problem. Since the RRPD problem leads to an extremely complex joint formulation, we simplify it by decoupling RRPD into the routing and the regenerator placement and dimensioning (RPD) subproblems, and thus, we eventually provide a formal model to solve the so-called R+RPD problem by means of Mixed Integer Linear Programming (MILP) formulations. Since the resulting relaxation is still difficult when large problem instances are considered, we also propose several alternative RPD methods and evaluate their performance by considering the trade-off between optimality and complexity they provide. Finally, we study the performance of the proposed T-OBS network under the considered R+RPD strategies by means of network simulation.

The rest of the paper is organized as follows. In Section II, we survey the previous work in this topic and highlight the main contributions of this paper. In Section III, we give a complete description of both the proposed T-OBS network architecture and the network model we use to capture the impact of the main PLIs. In Section IV, first we define the RRPD problem, and then, we present a MILP model to solve it. In Section V, several alternative resolution methods based on either MILP or heuristic algorithms to solve the RRPD problem are proposed. All strategies proposed are compared and evaluated in Section VI. Finally, the conclusions of this study are given in Section VII.

II. Related work and contributions

The evolution of optical networks from traditional opaque towards transparent network architectures has brought to light the serious impact that PLIs have on the optical end-to-end signal quality. In fact, due to these physical constraints and the lack of optical 3R regeneration, the deployment of a fully transparent long-haul network is still not viable. Hence, for the sake of scientific progress, the consideration of PLIs in the design and development of next-generation OTNs has become unavoidable. As a matter of fact, the study and evaluation of translucent WSONs, which rely on already mature technology, has recently received increasing attention from the research community. Such an infrastructure makes use of a limited set of 3R regenerators which are strategically deployed across the network for signal regeneration purposes [9]. Since the research interest on translucent arquitectures lies in the trade-off between network construction cost (i.e., O/E/O devices are expensive) and service provisioning performance (i.e., proper optical end-to-end Quality of Transmission (QoT) must be ensured), both the routing and RP issues must be carefully engineered. However, the RP problem is known to be N-complete [10], and hence, heuristic approaches are generally employed [3]. Indeed, recent studies in WSONs (see e.g., [4], [5]) show that by combining the RP problem with the routing problem in the so-called RRP problem, an improvement in the network performance can be achieved.

Moreover, it goes without saying that owing to the cumulative effect of PLIs, which eventually determine the feasibility of each optical end-to-end connection, there exists the need for an optical control plane (OCP) to efficiently manage such transmission constraints. Therefore, the OCP inevitably requires some modifications and enhancements. For instance, in [11], various signalling-based architectural options for a PLI-aware OCP are proposed and evaluated in a Generalized Multi-Protocol Label Switching (GMPLS) framework. In addition, in [12], translucent-oriented GMPLS protocol extensions similar to those being discussed within IETF WSON [6]-[13] are experimentally validated.

However, in light of the foreseen highly dynamic data traffic scenario, fine-grained and flexible technologies such as the sub-wavelength paradigms (e.g., OPS, OBS, ODS) have emerged as potential candidates to cope with the needs of next-generation OTNs. In this work, we focus on OBS networks, a technology which, in essence, overcomes the technologycal constraints of OPS and the bandwidth inefficiency of WSONs. An OBS network is made up of two types of nodes, namely edge and core nodes. In an OBS network, the transport of client data, which comes from different sources (e.g., IP packet traffic, Ethernet), is based on the following principles. Edge nodes are in charge of both assembling client input packets into outgoing bursts and of disassembling incoming bursts. For each outgoing burst, edge nodes emit a separate BCP in advance, to reserve resources (i.e., bandwidth on a desired output channel) along the way from the ingress node to an egress node. Core nodes and their corresponding control units are responsible for switching individual bursts and for reading, processing, and updating BCPs. In OBS, core nodes are generally assumed to be wavelength conversion capable.

In the case of OBS, however, research has been mainly geared towards evaluating the opaque and transparent network architectures. Consequently, the vast majority of the works consider that either an ideal physical layer or signal regenerators at every channel, port and switching node of the network are available (i.e., OBS is either fully transparent or opaque). Recently, however, owing to the increasing interest on assessing the effect of the PLIs in the optical networks field, we find few interesting works that involve the PLI constraint in the evaluation of the OBS network performance. For example, some impairment-aware scheduling policies with the aim of minimizing the burst loss probability are presented in [14]. Another interesting study that incorporates PLIs in the definition of an algorithm for distributing manycasting services over an OBS network can be found in [15]. An extensive study that evaluates the design and maximum size and throughput for OBS core nodes considering the effects of a range of PLIs such as amplifier noise, crosstalk of WDM channels, gain saturation and dynamics can be found in [16]. However, in [16], all nodes are equipped with O/E/O regenerators, one per each wavelength, also performing wavelength conversion, and thus, an opaque OBS network is being considered.

Our preliminary work in [8] tackled, for the first time to

the best of our knowledge, the issue of designing a complete T-OBS network architecture. To be precise, we first presented a feasible T-OBS network architecture and a model to capture the impact of the main PLIs which uses the Optical Signal to Noise Ratio (OSNR) as the signal quality performance indicator. Finally, we evaluated the performance of the T-OBS network by means of two simple RRPD heuristics. Both algorithms aimed at grouping the regenerators in as few nodes as possible but the one which relied on optimal MILP routing formulations stood out as the best method. In this paper, we present a more detailed description and a more complete analysis of our T-OBS network architecture, proposing novel MILP formulations and heuristics to solve the RRPD problem and assessing their performance and comparing it with that of the transparent and opaque reference scenarios. For this purpose, the contribution of this work is twofold. First, we present the design of a feasible T-OBS network (i.e. with either commercially available or at most lab trial devices like [17]) which has O/E/O regenerators available at selected nodes. In addition, a PLI model based on the calculation of the OSNR figure at the receiving end is derived for the T-OBS architecture. Second, we propose and analyze several design strategies for solving the RRPD problem efficiently. The study here presented follows a static/off-line approach since RRPD decisions are taken during the network planning stage. The consideration of a dynamic traffic matrix, by contrast, would result in an on-line routing and regenerator allocation problem, an issue which is left out of the scope of this paper.

III. Translucent OBS network model

In this section, we present in detail the proposed translucent OBS network model. First, we specify an architecture for an all-optical OBS node which incorporates a limited number of shared electrical regenerators. Second, we present the analytic model that we consider for the calculation of the OSNR level. Finally, a power budget and noise analysis of the characteristic signal path between two adjacent OBS nodes is provided.

A. Node architecture

The node architecture here presented is based on the model proposed in [16], where an opaque OBS network solution is considered. To be precise, the authors present two Semiconductor Optical Amplifier (SOA)-based node architectures for OBS networks, namely Broadcast-And-Select (BAS) and Tune-And-Select (TAS). Both architectures rely on the promising SOA technology and on wavelength converters performing electrical 3R regeneration as their fundamental switch modules. Indeed, SOA as switching elements (SW-SOA) bring some interesting advantages such as high on/off ratios and loss compensation capabilities. Despite this, however, SOA technology also entails some non-desirable effects such as power consumption, noise and nonlinearity that must be taken into account during the design process of the node. Among them, the authors conclude that TAS is more appropriated for OBS networks because BAS displays some major drawbacks (e.g., high power requirements and large inter channel crosstalk) inherent to its architecture.

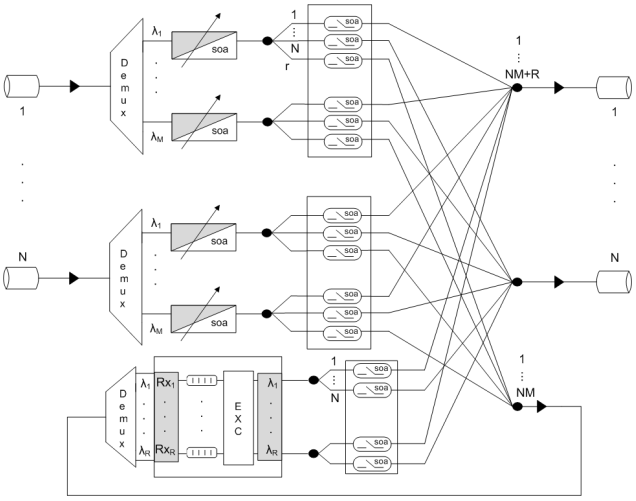

Fig. 1. T-OBS node architecture.

In this paper, we modify the aforementioned opaque TAS OBS core node architecture by replacing each inline electrical wavelength converter with a block consisting of a tunable laser and a wavelength conversion-type SOA (WC-SOA) device. Hence, this modified TAS node architecture (depicted in Fig. 1) is able to perform an all-optical switching operation. The node consists of N input/output fibers with M channels each and a limited number R of regenerators available. After the signal is amplified by the Erbium-Doped Fiber Amplifier (EDFA) pre-amplifier at each node input port, it is demultiplexed and passes through a fixed-input and variable-output WC-SOA. Then, the signal is split into $N+1$ branches, one per each fiber plus an extra branch that allows the access to the regenerator pool, which consists of a set of R fixed receivers, an electrical buffering stage and a set of R lasers emitting in predefined wavelengths (i.e., $\lambda_1, ..., \lambda_R$). The signal is then transported to the output ports of the node following the decisions of the OBS node controller by turning the SW-SOAs either ON or OFF. After the combiner stage, an EDFA booster amplifier provides the signal with enough power to cope with the losses of the first fiber span. Note also that, in this case, the combiners behind the SW-SOAs port merge $NM+R$ signals at each output port as a consequence of the presence of the regenerator pool.

It is worth mentioning that since the output of the WC-SOA is handled by the OBS node controller, all wavelengths from all input ports have the same privileges when requesting a regenerator, and thus, fairness in the access to the regenerator pool is provided by this architecture.

In the next subsections, we evaluate the performance of the proposed node architecture by means of an OSNR model.

B. OSNR model

In this OSNR model, the impact of PLIs is captured by considering the power of both the signal and the noise, which are affected by different gains and losses along the path, at the destination node. Although there exist many other PLIs, either linear or non-linear, here we consider the amplified spontaneous emission (ASE) noise introduced by both the

EDFA and SOA amplifiers as the significant signal impairment factor. In fact, ASE noise is commonly considered as the most severe impairment that limits the reach and capacity of optical systems. In this case, OSNR is defined as the ratio between the signal channel power and the power of the ASE noise in a specified bandwidth (e.g., $0.1nm$ are usually taken by convention) and is generally the fundamental metric which literature studies are based on. For instance, an OSNR model and its evaluation in transparent WSONs is proposed in [18], while in [19] a translucent WSON is experimentally validated.

To quantify the OSNR degradation along the optical path, we define the optical path OSNR (P_{osnr}) by taking advantage of the model described in [20]. Specifically, the OSNR consists of two main components, namely the link and node OSNR that we denote as L_{osnr} and N_{osnr} respectively. Since a link is composed of several amplifier spans, each ending with an in-line EDFA amplifier, the longer the path the higher the impact of the ASE noise in the OSNR received. Similarly, to minimize the ASE effect caused by the internal node amplifiers, gain values should be designed such that each node presents an OSNR level as high as possible. We can compute P_{osnr} for an optical end-to-end path traversing k links by using the following equation,

$$P_{osnr} = 1/(\sum_{i=1}^{k} \frac{1}{L^i_{osnr}} + \sum_{i=1}^{k} \frac{1}{N^i_{osnr}}), \quad (1)$$

where for a link consisting of r amplifier spans, L^i_{osnr} is defined as follows,

$$L^i_{osnr} = 1/(\sum_{j=1}^{r} \frac{1}{AS^j_{osnr}}), \quad (2)$$

where AS^j_{osnr} is the amplifier span OSNR, which can be calculated as,

$$AS^j_{osnr}[dB] = P_j[dBm] - QN[dBm] - F_j[dB] - G_j[dB], \quad (3)$$

where P_j, QN, F_j, G_j, correspond to the output power after the j^{th} amplifier span, the quantum noise, the noise figure and the gain of the j^{th} amplifier (i.e., either EDFA in-line or pre-amplifier) respectively. The expression that we use to compute N_{osnr} is equal to the one that we have defined for AS_{osnr}, however, due to the presence of several components (e.g., amplifiers, splitters and combiners) in our translucent node, both an equivalent noise and gain figure, namely F_{eq} and G_{eq} respectively, have to be derived.

In the next subsection, we provide specific values for all these figures by considering performance parameter values obtained from datasheets of commercially available devices (see e.g., [21]-[22]).

C. Power budget and noise analysis

We consider the power and noise constraints together in order to evaluate the OSNR of a signal that follows the characteristic path between two TAS neighboring nodes depicted in Fig. 2. Component specifications are provided in Table I and

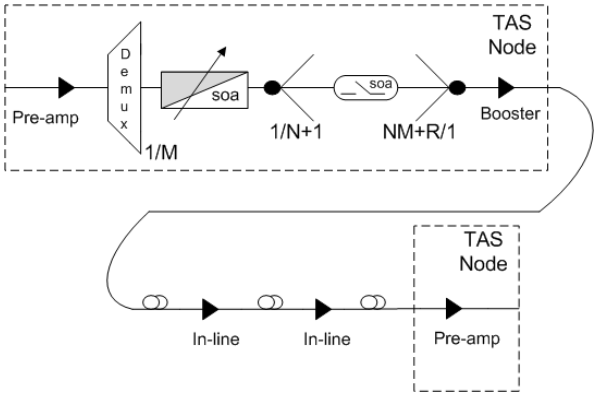

Fig. 2. Signal path between two TAS OBS core nodes.

Channels (M)	32	
Span length	$65km$	
Fiber attenuation	$0.2dB/km + 3dB$ (cable margin)	
Quantum Noise	$-58dBm$	
EDFA (pre-amp)	noise figure	$5.5dB$
	max. gain	$20dB$
	max. output power	$13dBm$
	min. input power	$-30dBm$
EDFA (booster)	noise figure	$5.5dB$
	max. gain	$15dB$
	max. output power	$18dBm$
	min. input power	$-15dBm$
EDFA (in-line)	noise figure	$5.5dB$
	max. gain	$25dB$
	max. output power	$18dBm$
	min. input power	$-25dBm$
WC-SOA	noise figure	$9dB$
	max. gain	$16dB$
	max. output power	$5dBm$
	min. input power	$-25dBm$
SW-SOA	noise figure	$10dB$
	max. gain	$10dB$
	max. output power	$3dBm$
	rise-fall time	$500ps$
WDM Demux	insertion loss ($M = 32$)	(≈ 5.5) dB
Splitter	insertion loss	$(0.5 - 1)$ dB
Combiner	insertion loss	$(1.5 - 2)$ dB

TABLE I
PARAMETER VALUES CONSIDERED

the power constraints for this analysis are: the output power of the node (i.e., output of the EDFA booster amplifier) set to 0dBm/channel, and its input power (i.e., input of the EDFA pre-amplifier) set by link losses to -16dBm/channel.

From (3) and bearing in mind that the objective is to have a N_{osnr} as high as possible, it can be inferred that both F_{eq} and G_{eq} must be designed so that its resultant value is minimized. For this particular case, the equivalent noise and gain figures of the TAS node are obtained as follows,

$$F_{eq} = F_{wc-soa} + \frac{MF_{sw-soa} - 1}{\frac{G_{wc-soa}}{L_{splitter}}} + \frac{F_{edfa-booster} - 1}{\frac{G_{wc-soa}G_{sw-soa}}{L_{splitter}L_{combiner}}}, \quad (4)$$

$$G_{eq} = \frac{G_{wc-soa}G_{sw-soa}G_{edfa-booster}}{L_{splitter}L_{combiner}}. \quad (5)$$

The most critical point is the combiner where, in the worst case, the ASE noise power from M SW-SOAs is merged. Both

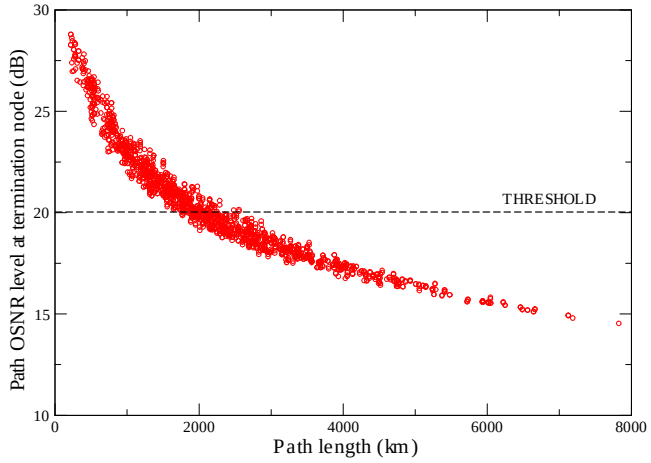

Fig. 3. Pan-European paths OSNR evaluation.

the pre-amplifier and booster EDFAs and the WC-SOA and SW-SOA have to be used to compensate the internal losses. Their gain values must be carefully designed so that both equivalent figures are minimized and the power constraints are respected. In order to minimize F_{eq}, it can be deduced from (4) that, as long as the saturation output power is not reached, it is better to set the gain on the WC-SOA. In this way, the impact of the M ASE powers is reduced. The EDFAs pre-amplifier and booster and SW-SOA gains, by contrast, are kept as low as allowed by the system power requirements. The exact set up for each component depends on the number of input/output ports of each particular node, which eventually define the splitting losses that are to be covered by G_{eq}.

In Fig. 3, we show the result of the application of the OSNR model presented throughout this whole section considering the optical end-to-end paths of the Large and Core topologies (see Appendix A for the simulation details). Figure 3 makes it clear that the length, and thus, the number of amplifier spans, have a strong impact on the received OSNR. Consequently, all bursts arriving at the destination node with an accumulated OSNR value lower than a predefined quality threshold (T_{osnr}) cannot be read correctly, and thus, are discarded. Being the T-OBS network architecture and the network model used to capture the impact of PLI described, the next Section is devoted to present a formal model to solve the RRPD problem.

IV. MILP FORMULATION OF THE RRPD PROBLEM

In this Section, we focus on the modelling of the RRPD problem in a T-OBS network presenting the problem definition and its particular design assumptions. In general, our approach to RRPD concerns, respectively, the design of explicit paths to be used to route bursts through the network, and the placement of regenerators in selected nodes on those paths together with the dimensioning of such regenerators in each node.

The result of this design procedure is a set of routing paths and a subset of regenerative nodes which is specified for each individual path that does not comply the quality of signal requirements. It is essential to our approach that a burst whenever sent on a path will be regenerated only at the nodes that are specified as regenerative nodes for this path. It is worth pointing out that since we are addressing an off-line design problem, we can assume that burst control packets (BCPs) are provided at their respective source node with the information on the set of nodes where their corresponding data burst will be regenerated. We also assume that the signal quality of the BCPs is always satisfactory because they undergo an O/E/O conversion at each node for processing purposes and a successful transmission must be assured at least between two adjacent nodes.

A. RRPD Problem definition

We address the RRPD problem by uncoupling the routing formulation from that of the RPD issue, and therefore, we provide a model to tackle the problem of R+RPD. Two main reasons support this modelling decision. First, treating both problems together and at once would definitely make of the problem an extremely complex undertaking, particularly in terms of computation times or even of solving feasibility. Second, and most compelling, is the fact that in OBS networks, routing must be carefully engineered since the main source of performance degradation is the contention between bursts that arise due to both the lack of optical buffering and the generally considered one-way resource reservation scheme.

Hence, given a set of traffic demands, we first find a proper routing that minimizes burst losses due to congestion in bottleneck network links. Then, this routing solution is used as input information to solve the RPD problem. Since in the T-OBS network the access to the regenerator pools is based on statistical multiplexing, the RPD method must deal with both the selection of regeneration nodes and the dimensioning of regenerator pools so that a given target burst loss rate due to OSNR non-compliant bursts is satisfied. The aim of the RPD formulation here proposed is hence the minimization of the number of O/E/O regenerators deployed in the network.

B. Global notation

We use $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ to denote the graph of an OBS network; the set of nodes is denoted as $\mathcal{V}$, and the set of unidirectional links is denoted as $\mathcal{E}$. Let $\mathcal{P}$ denote the set of predefined candidate paths between source s and termination t nodes, $s, t \in \mathcal{V}$, and $s \neq t$. Each path $p \in \mathcal{P}$ is identified with a subset of network links, that is, $p \subseteq \mathcal{E}$. Adequately, subset $\mathcal{P}_e \subseteq \mathcal{P}$ denotes all paths that go through link e. Let s_p and t_p denote the source and termination nodes of p. Let $\mathcal{D}$ denote the set of demands, where each demand corresponds to a pair of source-termination nodes. For each demand $d \in \mathcal{D}$, $h_d \in \mathbb{R}_+$ denotes the volume of burst traffic; Let $\mathcal{N}_p$ be the set of all nodes constituting path p. Finally, let $\mathcal{V}_p$ denote the set of intermediate nodes on path p such that $\mathcal{V}_p = \mathcal{N}_p \setminus \{s_p, t_p\}$.

C. Routing problem

1) Model assumptions: The routing model that we consider and the routing algorithm that we apply are similar to the Linear Programming (LP) based approach presented in [23]. To be precise, the authors consider a Multi-Path Routing (MPR) approach (i.e., splittable routing) to solve the routing problem.

The objective of this method is to distribute traffic over a set of candidate paths so that to reduce congestion in network bottleneck links. For this purpose, the network is assumed to apply source based routing, and hence, the source node is able to determine the path that a burst entering the network must follow. Although we take the same routing objective, in our study we consider unsplittable (non-bifurcated) routing and, accordingly, all the traffic offered to demand $d \in D$ is carried over a single path in the network.

Let $\mathcal{P}_d \subseteq \mathcal{P}$ denote the set of candidate paths supporting demand d; $\mathcal{P} = \bigcup_{d \in D} \mathcal{P}_d$. Each subset $\mathcal{P}_d$ comprises a (small) number of paths, for example, k shortest paths. The selection of path p from set $\mathcal{P}_d$ is performed according to a decision variable x_p, which later is referred to as the path selection variable or routing variable. In this study, on the contrary to the assumption taken in [23], variables x_p are forced to be binary. To be precise, a burst flow is routed over path p iff $x_p = 1$. Moreover, there is only one path $p \in \mathcal{P}_d$ such that $x_p = 1$. Hence, these routing constraints can be expressed as:

$$\sum\nolimits_{p \in \mathcal{P}_d} x_p = 1, \quad \forall d \in \mathcal{D}, \tag{6a}$$

$$x_p \in \{0, 1\}, \quad \forall p \in \mathcal{P}, \tag{6b}$$

and the traffic ρ_p to path $p \in \mathcal{P}_d$ can be calculated as:

$$\rho_p = x_p h_d = \left\{ \begin{array}{cc} h_d & \text{if } x_p = 1, \\ 0 & \text{otherwise.} \end{array} \right\} \tag{7}$$

As a consequence, the problem formulations in the next subsection are MILP formulations. Notice that vector $\mathbf{x} = (x_1, ..., x_{|\mathcal{P}|})$ determines the distribution of the traffic over the network. This vector has to be optimized in order to reduce link congestion and to improve the overall network performance.

2) Problem formulation: Following the LP algorithm presented in [23], the next two MILP models are sequentially solved to find a solution to the routing problem. First, let variable y represent the average traffic load on the bottleneck link. Then, the first MILP formulation, which aims at minimizing the load on such particular link of the network, can be written as follows:

$$\begin{array}{ll} \textit{minimize} & y \\ \textit{subject to} & \\ & \sum\nolimits_{p \in \mathcal{P}_e} x_p h_d - y \leq 0, \ \forall e \in \mathcal{E} \end{array} \tag{RMILP1}$$

$$\tag{8}$$

and subject to the routing constraints given by (6a) and (6b).

Despite minimizing the average traffic load on the bottleneck link, many solutions to this problem may exist and most of them exploit unnecessary resources in the network (i.e., solutions that select longer paths). Therefore, the next MILP is solved in order to obtain, between the solutions of RMILP1, the one that entails the minimum increase of the average traffic load offered to the remaining network links. For this purpose, let us denote y^* as an optimal solution of RMILP1, then we solve the following problem:

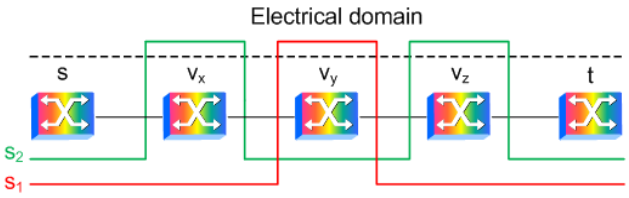

Fig. 4. Two different valid options to perform the regeneration for a particular source-termination pair.

$$\begin{array}{ll} \textit{minimize} & \sum\nolimits_{e \in \mathcal{E}} \sum\nolimits_{p \in \mathcal{P}_e} x_p h_d \\ \textit{subject to} & \\ & \sum\nolimits_{p \in \mathcal{P}_e} x_p h_d \leq y^*, \ \forall e \in \mathcal{E} \end{array} \tag{RMILP2}$$

$$\tag{9}$$

and subject to the routing constraints given by (6a) and (6b). Note that, in constraint (9), we ensure that the maximum average traffic load on the bottleneck link is bounded by the solution of RMILP1.

These MILP models, if sequentially solved, determine the path p that will be in charge of carrying the traffic for each demand d. Hence, only one path $p_d \in \mathcal{P}_d$ is selected as the valid path to be followed by all bursts belonging to demand d. Thus, we can now denote $\mathcal{Q}$ as the set of valid paths, $\mathcal{Q} = \{p_d, d \in \mathcal{D}\}$. In the next Section, we use $\mathcal{Q}$ as input information to solve the RPD problem.

D. RPD problem

1) Model assumptions: Let $\mathcal{P}^o \subseteq \mathcal{Q}$ denote the subset of paths for which the OSNR level at receiver t is non-compliant with the quality of signal requirements, and thus, paths $p \in \mathcal{Q}$ requiring regeneration at some node $v \in \mathcal{V}_p$. For each $p \in \mathcal{P}^o$ there may exist many different options on how to build an end-to-end OSNR compliant path, composed by its transparent segments, since the node or group of nodes where the regeneration has to be performed might not be a unique solution. Thus, let $\mathcal{S}_p = \{s_1, \ldots, s_{|\mathcal{S}_p|}\}$ denote the set of different options to establish an OSNR compliant path for each path $p \in \mathcal{P}^o$, where $s_i \subseteq \mathcal{V}$, $i = 1 \ldots |\mathcal{S}_p|$ and size $|\mathcal{S}_p|$ depends on the length of the transparent segments in path p. Figure 4 illustrates this concept by means of an optical path between a source-termination pair $(s - t)$ with two different options to establish an OSNR compliant path. To be precise, if s_1 is selected, the optical signal only undergoes 3R electrical regeneration at node v_y, whereas if s_2 is the choice, it is converted to the electrical domain two times (i.e., at nodes v_x and v_z). Hence, $s_1 = \{v_y\}$ and $s_2 = \{v_x, v_z\}$. In this particular case, the transparent segments that make it possible to use both regeneration solutions are segments $[s-v_y]$-$[v_y-t]$ and $[s - v_x]$-$[v_x - v_z]$-$[v_z - t]$. Notice that we could also consider other cases like $s_3 = \{v_x, v_y, v_z\}$, however, we have not depicted all of the options for the sake of clarity. Here it is worth pointing out that we obtain $\mathcal{S}_p$, $p \in \mathcal{P}^o$ by means of a precomputation phase where all possible regeneration options are obtained using the OSNR model presented in [8].

We assume that for each path $p \in \mathcal{P}^o$, the selection of the regeneration option s from set $\mathcal{S}_p$ is performed according to a decision variable z_{ps}, which later is referred to as regenerator placement variable, such that the following constraints are fulfilled:

$$\sum\nolimits_{s \in S_p} z_{ps} = 1, \quad \forall p \in \mathcal{P}^o, \tag{10a}$$

$$z_{ps} \in \{0,1\}, \quad \forall s \in \mathcal{S}_p, \forall p \in \mathcal{P}^o. \tag{10b}$$

Let ρ_v^o denote the offered traffic load requiring regeneration at node v. To estimate ρ_v^o (approximately) we add up the traffic load ρ_p offered to each path $p \in \mathcal{P}^o$ that both crosses and undergoes regeneration at node v:

$$\rho_v^o = \sum\nolimits_{p \in \mathcal{P}^o : \mathcal{V}_p \ni v} \sum\nolimits_{s \in \mathcal{S}_p : s \ni v} z_{ps} \rho_p. \tag{11}$$

Similarly,

$$\rho_v = \sum\nolimits_{p \in \mathcal{P}^o : \mathcal{V}_p \ni v} \rho_p, \tag{12}$$

denotes an estimation of the maximal traffic load that is subject to regeneration at node $v \in \mathcal{V}$.

Eventually, we define a regenerator pool dimensioning function $F_v(\cdot)$, which for a given traffic load ρ_v^o, determines the minimum number of regenerators to be allocated in node v. This number must ensure that a given target burst blocking probability (B^{osnr}) for bursts competing for regeneration resources is met. Assuming Poisson arrivals and fairness in the access to regenerator pools among bursts (see subsection V-E) such a function is given by the following discontinuous, step-increasing function,

$$F_v(\rho_v^o) = \left\lceil B^{-1}(\rho_v^o, B^{osnr}) \right\rceil, \tag{13}$$

where B corresponds to the Erlang B-loss formula which for a given number of regenerators $r \in \mathbb{N}$ available at node v can be calculated as,

$$B(\rho_v^o, r) = \frac{(\rho_v^o)^r / r!}{\sum_{k=0}^{r} (\rho_v^o)^k / k!}, \tag{14}$$

and where $B^{-1}(\rho_v^o, B^{osnr})$ is the inverse function of (14) extended to the real domain [24], and $\lceil \cdot \rceil$ is the ceiling function. It is worth noticing that the Poisson arrivals which lead to an Erlang formula for the dimensioning of regenerator pools can be replaced with another distribution for which the blocking probability is attainable. Because B^{osnr} is a predetermined parameter, for simplicity of presentation we skipped it from the list of arguments of function $F_v(\cdot)$. Function $F_v(\rho_{vo})$ is depicted in Fig. 5 for some exemplary B_v^{osnr} values. Note that $B^{-1}(\cdot)$ is a real-valued concave function.

For the purpose of problem formulation, it is convenient to define a_r as the maximal load supported by r regenerators given a B^{osnr}, i.e., $a_r = B^{-1}(r, B^{osnr})$. Note that the inverse function $B^{-1}(r, B^{osnr})$ is expressed with respect to r and B^{osnr}, which is not the same as in function (13).

Although there is no close formula to compute the inverse of (14), we can make use of a line search method (see e.g., [25]) to find the root ρ^* of the function $f(\rho) = B^{osnr} - B(\rho, r)$ so that the value of a_r is approximated by $a_r = \rho^*$ for any index r. Finally, let R denote the number of regenerators required in the most loaded node, that is, $R = \max\{F_v(\rho_v) : v \in \mathcal{V}\}$.

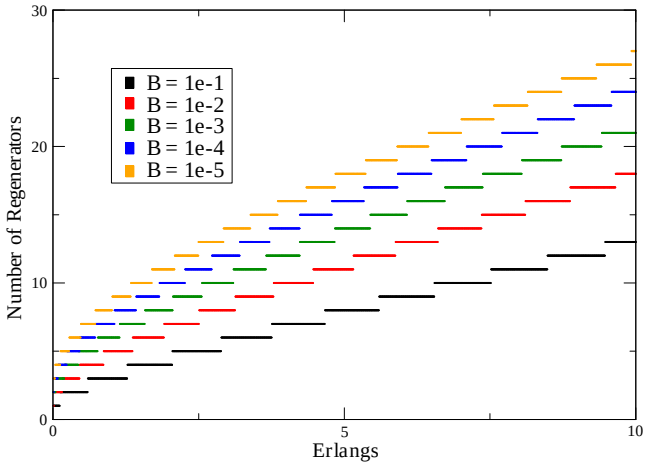

Fig. 5. Discontinuous step-increasing regenerator pool dimensioning function for some exemplary target burst loss probabilities.

Vector $\mathbf{a} = (a_1, ..., a_R)$ will also be used in subsection V-E to determine $F_v(\rho_v^o)$ according to Procedure 1. Note that Procedure 1 is a polynomial time algorithm of complexity $O(R)$.

Procedure 1 Regenerator Pool Dimensioning

1: $r \leftarrow 0$
2: **while** $\rho_v^o > a_r$ **do**
3: $\quad r \leftarrow r + 1$
4: **end while**
5: $F_v \leftarrow r$

2) Problem formulation: Taking into account the network modelling assumptions previously presented, here we present a mathematical formulation for the RPD part of the problem.

The RPD problem can be formulated as a non-convex optimization problem:

$$\underset{\mathbf{z}}{minimize} \quad F = \sum\nolimits_v F_v(\rho_v^o(\mathbf{z})) \tag{NLP1}$$

$$subject\ to \quad \text{(10a) and (10b)} \tag{15a}$$

where $F_v(\cdot)$ is the step-increasing regenerator pool dimensioning function defined by (13) and $\rho_v^o(\mathbf{z})$ is the function representing the traffic load offered to a regenerator node defined by (11). The optimization objective of NLP1 is to minimize the sum of regenerators installed in network nodes. Constraints (15a) represent the selection of an OSNR compliant path from the provided options for each path requiring regeneration. Eventually, the RP decision vector $\mathbf{z}$ is defined as $\mathbf{z} = (z_{11} \ldots z_{1|\mathcal{S}_p|}, \ \ldots, \ z_{|\mathcal{P}^o|1} \ldots z_{|\mathcal{P}^o||\mathcal{S}_{|\mathcal{P}^o|}|})$.

The difficulty of formulation NLP1 lays in the fact that there is no close formula to express $F_v(\cdot)$ since no such formula exists for the inverse of the Erlang function $B^{-1}(\cdot)$. A way to solve the problem is to substitute function $F_v(\cdot)$, $v \in \mathcal{V}$ with its piecewise linear approximation and reformulate NLP1 as a MILP problem.

For a single node $v \in \mathcal{V}$, the piecewise linear approximation of $F_v(\cdot)$ can be expressed as $F_v(\rho_v^o) = \min\{r : a_r > \rho_v^o\}$, or

by means of a 0-1 integer programming (IP) formulation:

$$\begin{aligned} \underset{\mathbf{u}}{minimize} \quad & F_v = \sum\nolimits_r u_v^r r && \text{(IP1)} \\ subject\ to \quad & u_v^r(a_r - \rho_v^o) \geq 0,\ \forall r \in [1,R], && \text{(16a)} \\ & \sum\nolimits_r u_v^r = 1, && \text{(16b)} \\ & u_v^r \in \{0,1\},\ \forall r \in [1,R]. && \text{(16c)} \end{aligned}$$

In IP1, decision variables u_v^r have been introduced in order to represent the number of regenerators required in node v. Due to constraint (16b), in each node only one variable u_v^r is active (i.e., equal to 1), and the one with minimum r satisfying $a_r \geq \rho_v^o$ is found when solving the problem. Notice that formulation IP1, when solved, gives the same solution as Procedure 1. The shortcoming of IP1 is that since ρ_v^o is dependent on vector $\mathbf{z}$ (i.e., ρ_v^o is a function of $\mathbf{z}$), constraints (16a) have quadratic form. To overcome this difficulty, we can consider the following alternative formulation:

$$\begin{aligned} \underset{\mathbf{u}}{minimize} \quad & F_v = \sum\nolimits_r u_v^r r && \text{(IP2)} \\ subject\ to \quad & \sum\nolimits_r u_v^r a_r \geq \rho_v^o, && \text{(17a)} \\ & \sum\nolimits_r u_v^r = 1, && \text{(17b)} \\ & u_v^r \in \{0,1\},\ \forall r. && \text{(17c)} \end{aligned}$$

It is easy to note that formulation of IP2 results directly from IP1; it is enough to add up constraints (16a) and use (16b) for substituting $\rho_v^o \sum_r u_v^r$ by ρ_v^o.

Eventually, taking into account all network nodes and introducing the regenerator placement decision variables, problem NLP1 can be reformulated as a MILP problem:

$$\underset{\mathbf{u},\boldsymbol{\rho}^o,\mathbf{z}}{minimize} \quad F = \sum\nolimits_v \sum\nolimits_r u_v^r r \qquad \text{(MILP1)}$$

$$subject\ to$$

$$\sum\nolimits_r u_v^r a_r - \rho_v^o \geq 0, \quad \forall v \in \mathcal{V}, \qquad \text{(18a)}$$

$$\sum\nolimits_r u_v^r = 1, \quad \forall v \in \mathcal{V}, \qquad \text{(18b)}$$

$$\sum\nolimits_{s \in S_p} z_{ps} = 1, \quad \forall p \in \mathcal{P}^o, \qquad \text{(18c)}$$

$$\sum\nolimits_{p \in \mathcal{P}^o : \mathcal{V}_p \ni v} \sum\nolimits_{s \in \mathcal{S}_p : s \ni v} z_{ps}\rho_p - \rho_v^o = 0,\ \forall v \in \mathcal{V}, \qquad \text{(18d)}$$

$$u_v^r \in \{0,1\}, \quad \forall r \in [1,R], \forall v \in \mathcal{V}, \qquad \text{(18e)}$$

$$z_{ps} \in \{0,1\}, \quad \forall p \in \mathcal{P}^o, \forall s \in \mathcal{S}_p, \qquad \text{(18f)}$$

$$\rho_v^o \in \mathbb{R}^+, \quad \forall v \in \mathcal{V}. \qquad \text{(18g)}$$

where we consider ρ_v^o to be an auxiliary variable representing the traffic load requiring regeneration offered to node $v \in \mathcal{V}$.

The objective of the optimization problem MILP1 is to minimize the total number of regenerators that have to be placed in the network. Constraints (18a) and (18b) result from the 0-1 representation of the dimensioning function and from the reformulation of IP1 as mentioned before. In particular, the number of regenerators in node $v \in \mathcal{V}$ should be such that the maximum traffic load (given a B^{osnr}) is greater or equal to offered traffic load ρ_v^o. Constraints (18c) are the OSNR compliant path selection constraints. Constraints (18d) are the traffic load offered to a regenerator node calculation constraints. Eventually, (18e), (18f), and (18g) are the variable range constraints.

MILP1 is a well-known Discrete Cost Multicommodity Flow (DCMCF) problem [26]. DCMCF was shown to be an extremely difficult combinatorial problem for which only fairly small instances (in our case, situations where $\mathcal{P}^o$ has a rather small size) can be solved exactly with currently available techniques. In the next Section, we propose several less complex heuristic methods to solve the RPD problem.

V. RPD HEURISTIC RESOLUTION METHODS

To overcome the difficulty imposed by the resolution of MILP1, in this Section, we propose several heuristic methods that provide near-optimal solutions to the RPD problem within acceptable computational times. The main idea behind all these strategies is to decouple the RPD problem on the RP problem, which is solved first, and the dimensioning phase performed afterwards. Hence, we derive models to solve the so-called RP+D problem. The performance of these methods is later discussed in Section VI.

A. Load-based MILP formulation

The MILP formulation here proposed is focused on the distribution of the traffic load requiring regeneration (i.e., ρ_v^o, $\forall v \in \mathcal{V}$). Hence, this load must be aggregated in such a way that the number of regenerators to be deployed is minimized. After a ρ_v^o solution is obtained for each node $v \in \mathcal{V}$, we take advantage of the regenerator pool dimensioning function detailed in Section V-E to obtain the number of regenerators required.

Owing to the concave character of the dimensioning function (13), it must be noted that it is of our interest to aggregate the traffic requiring regeneration in as few nodes as possible rather than spreading out such load in little amounts over a large number of nodes. Hence, we propose to solve the problem by making use of two MILP models, namely MILP2 and MILP3. These models can be sequentially solved to obtain a sub-optimal solution of MILP1.

First, MILP2 aims at minimizing the number of nodes where the regenerators must be installed (i.e., nodes such that $\rho_v^o > 0$), and thus, groups as much as possible the load that requires regeneration. Let $\mathbf{y} = (y_1, ..., y_{|\mathcal{V}|})$ denote a vector of binary decision variables. Each value corresponds to one node and determines if this node is used as regeneration point by some path $p \in \mathcal{P}^o$ ($y_v = 1$) or not ($y_v = 0$).
Then, we solve the following problem:

$$\begin{aligned} \underset{\boldsymbol{\rho}^o,\mathbf{z},\mathbf{y}}{minimize} \quad & \sum\nolimits_v y_v && \text{(MILP2)} \\ subject\ to \quad & \rho_v y_v \geq \rho_v^o, \quad \forall v \in \mathcal{V}, && \text{(19a)} \\ & y_v \in \{0,1\},\ \forall v \in \mathcal{V}. && \text{(19b)} \end{aligned}$$

and subject to constraints (10a), (10b), (18d) and (18g).

Although MILP2 minimizes the number of nodes where the regenerations are performed, multiple solutions to this problem may exist and some of them may exploit more regenerations than required, increasing unnecessarily ρ_v^o at some nodes. Therefore, a second MILP model, that is, MILP3, needs to be

formulated with the objective to minimize the total network load requiring regeneration.

Therefore, let k^* denote an optimal solution of MILP2. Second, we solve the following problem:

$$\underset{\boldsymbol{\rho}^o,\mathbf{z},\mathbf{y}}{minimize} \quad \sum\nolimits_v \rho_v^o \tag{MILP3}$$

$$subject\ to \quad \sum\nolimits_v y_v \leq k^*, \tag{20a}$$

and subject to constraints (10a), (10b), (18d), (18g), (19a) and (19b).

Due to the simplicity of both formulations, both models are expected to be promptly solved even for large-sized problem instances.

It is also important to notice that the sequential resolution of both MILP2 and MILP3, which will hereinafter be cited within the text as MILP2/3, provides an optimal solution in terms of the distribution of the traffic and not with respect to the number of regenerators (which is precisely the case of MILP1). This being said, the last step in this method is the dimensioning of regenerator pools as detailed in Section V-E.

B. Reduced MILP1 (MILP1)*

This method aims at reducing the complexity of MILP1 by introducing new constraints to its definition. Specifically, these constraints are the sequentially obtained solutions of both MILP2 and MILP3 as detailed previously in subsection V-A. Although these new constraints are not valid in that they may exclude the optimal solution of MILP1, they bring computation times of good near-optimal solutions (e.g., less than a 2% gap with respect to the optimal solution) within reasonable time limits.

Therefore, let us denote g^*, and again k^*, as the optimal sequentially solved solutions of MILP3 and MILP2 respectively. Then, we reformulate MILP1 as follows,

$$\underset{\mathbf{u},\boldsymbol{\rho}^o,\mathbf{z}}{minimize} \quad F = \sum\nolimits_v \sum\nolimits_r u_v^r r \tag{MILP1*}$$

$$subject\ to \quad \sum\nolimits_v y_v \leq k^*, \tag{21a}$$

$$\sum\nolimits_v \rho_v^o(\mathbf{z}) \leq g^*, \tag{21b}$$

and subject to constraints (18a), (18b), (18c), (18d), (18e), (18f), (18g), (19a) and (19b).

In fact, we sequentially solve all three models in order, that is, first MILP2, second MILP3 and finally MILP1 including all solutions obtained as constraints for the subsequent problem.

It is worth pointing out that, as long as the scenario considered does not involve optical paths that require a large number of regenerations, constraint (21a) is very unlikely to exclude the optimal solution of MILP1. Basically, it is due to the fact that the dimensioning function of our problem is (13), which favours, to some degree, the grouping-like behaviour. Constraint (21b), by contrast, is just an heuristic approach to help solve the problem. Notice that (21b) does not deal with the distribution of the load but with its minimisation, and thus, the optimal solution in terms of the number of regenerators is generally excluded.

C. A Local Search (LS) Algorithm

Here we propose an heuristic solution to the regenerator placement problem which is based on the K-L local search technique [27]. In the proposed algorithm, we assume a neighbouring solution is achieved by means of a flip operation which consists in a permutation of the regeneration points for a specific set of demands.

Let $\mathcal{A}$ be the set of all regeneration vectors that define for each path $p \in \mathcal{P}^o$, the node or set of nodes where the regeneration is performed, that is, $\mathcal{A} = \bigcup_{p\in\mathcal{P}^o} \mathbf{z}_p$, where $\mathbf{z}_p = (z_1, ..., z_{|\mathcal{S}_p|})$. Let $\mathcal{A}_o$ be an initial (randomly selected) solution to the problem where constraints (10a) are met for each $\mathbf{z}_p$, $p \in \mathcal{P}^o$.

Similarly, let $\mathcal{A}_f$, $\mathcal{A}_{tb}$, $\mathcal{A}_i$ and $\mathcal{A}_b$ denote, respectively, the final solution of the algorithm, the global best solution obtained so far, the best solution of a whole iteration and one of the solutions of the iteration in progress. Moreover, let $\Omega_{\mathcal{A}}$ be the set of valid solutions obtained once loop 5-13 is completed.

Procedure 2 LS Heuristic

INPUT: $\mathcal{P}_o, \mathcal{A}, \mathcal{A}_o, \Omega_{\mathcal{A}} \leftarrow \emptyset$
OUTPUT: $\mathcal{A}_f$

1: $\mathcal{A}_{tb} \leftarrow \mathcal{A}_o$
2: $\Omega_{\mathcal{A}} \leftarrow \Omega_{\mathcal{A}} \cup \{\mathcal{A}_o\}$
3: $\mathcal{A}_b \leftarrow \mathcal{A}_o$
4: **repeat**
5: **for all** path $p \in \mathcal{P}_o$ **do**
6: $\mathcal{P}_x \leftarrow \mathcal{P}_o \backslash \{p\}$
7: Take $\mathbf{z}_p$ from $\mathcal{A}$
8: Determine $\mathbf{z}_p^*$ that requires the minimum number of regenerators considering, for all path $p \in \mathcal{P}_x$, the option selected in $\mathcal{A}_b$
9: Let $\mathcal{A}_p$ be a new solution
10: $\mathcal{A}_p \leftarrow \mathcal{A}_b \cup \{\mathbf{z}_p^*\} \backslash \{\mathbf{z}_p\}$
11: $\Omega_{\mathcal{A}} \leftarrow \Omega_{\mathcal{A}} \cup \{\mathcal{A}_p\}$
12: $\mathcal{A}_b \leftarrow \mathcal{A}_p$
13: **end for**
14: Determine the solution of this iteration, $\mathcal{A}_i$, from $\Omega_{\mathcal{A}}$ that requires the minimum number of regenerators
15: $\mathcal{A}_b \leftarrow \mathcal{A}_i$
16: $\Omega_{\mathcal{A}} \leftarrow \mathcal{A}_b$
17: Let r_{tb} and r_i be the number of regenerators required by $\mathcal{A}_{tb}$ and $\mathcal{A}_i$ respectively
18: **if** $r_{tb} > r_i$ **then**
19: $\mathcal{A}_{tb} \leftarrow \mathcal{A}_i$
20: **end if**
21: **until** $r_{tb} \leq r_i$
22: $\mathcal{A}_f \leftarrow \mathcal{A}_{tb}$

Between lines 5 and 13, starting from an initial solution (i.e., $\mathcal{A}_b$), we iteratively take, for each $p \in \mathcal{P}^o$, vector $\mathbf{z}_p \in \mathcal{A}$, and then we set it to $\mathbf{z}_p^*$, which is the solution for vector $\mathbf{z}_p$ that minimizes the number of regenerators to be deployed taking into account the current solutions for all other paths, that is, solutions in the current $\mathcal{A}_b$. Once a choice is made for p, then it remains fixed until the loop is initiated again.

Note that in lines 8, 14 and 17 of Procedure 2, we use Procedure 1 to determine the exact number of regenerators required in each particular case. Despite we call the dimensioning function several times within Procedure 2, vector $\mathbf{a}$ is precomputed only once at the very beginning of the algorithm. It is also worth noticing that in line 12, an update of the current solution is performed even if it entails worsening $\mathcal{A}_b$. Procedure 2 does this in order to increase the probabilities of escaping from the local optima and in the hope that some neighboring solution generated during an iteration will turn out better than the current $\mathcal{A}_{tb}$.

1) Complexity Remarks: To evaluate the complexity of this algorithm let us first define δ as the number of nodes constituting the largest possible path contained in $\mathcal{P}^o$, that is,

$$\delta = \max\{|\mathcal{N}_p| : p \in \mathcal{P}^o\}. \tag{22}$$

Then, the complexity is given by $O(M\,|\mathcal{E}|\,|\mathcal{P}^o|\,(2^\delta - 2))$, where $(2^\delta - 2)$ is the upper bound on the maximum number of regeneration options for path p. Such operation is performed once per path, hence $|\mathcal{P}^o|$, and $M\,|\mathcal{E}|$ (i.e., the number of regenerators required in an opaque OBS network) defines an upper bound on the number of iterations at the worst-case improvement (one per iteration) of the cost function. Although the complexity of this routine is polynomial in time, LS can also perform quickly as shown later in Section VI.

D. A Regenerator Grouping (RG) Algorithm

In this method, the search for appropriate location of regenerators in intermediate nodes for all paths $p \in \mathcal{P}^o$ is performed. For this particular algorithm, let $\mathcal{R}_p$ denote the node or set of nodes where the regeneration is performed for path p. Let $\mathcal{R} = \bigcup_{p\in\mathcal{P}^o} \mathcal{R}_p$ be the set of all nodes where the regenerators have to be installed for all path $p \in \mathcal{P}^o$. Let Ω_p be the set of subpaths of p to be processed. Then, Procedure 3 is executed.

Procedure 3 iteratively processes each path $p \in \mathcal{P}^o$ with the aim of assuring that the OSNR signal level meets a predefined signal quality threshold at each node $v \in \mathcal{N}_p$. To provide a regenerator grouping-like behavior, in lines 4-9, the algorithm searches among all the previously processed paths if there are nodes such that $\{v \in \mathcal{V}_p : \rho_v^o \geq 0\}$, and if so, it takes the node that is nearest to the middle of the path (with respect to the number of hops) and selects it as the first regeneration point for path p. Hence, two new subpaths are added to Ω_p. Between line 10 and 23 the algorithm performs a loop that adds regeneration points to path p until Ω_p becomes an empty set.

Once Procedure 3 finishes, the load ρ_v^o for each node $v \in \mathcal{V}$ is obtained. Although the order of the iteratively processed paths in Procedure 3 may result in different solutions, still we observed that the algorithm performance does not vary significantly, and thus, we consider an arbitrary order. After Procedure 3 is executed, we can proceed to dimension the regenerator pools in all nodes having $\rho_v^o > 0$ (i.e., $\forall v \in \mathcal{R}$) as detailed in Procedure 1.

Procedure 3 RG Heuristic

INPUT: $\mathcal{P}_o, \mathcal{R} \leftarrow \emptyset, \Omega_p \leftarrow \emptyset$
OUTPUT: $\mathcal{R}$

1: **for all** path $p \in \mathcal{P}_o$ **do**
2: $\quad\Omega_p \leftarrow \Omega_p \cup \{p_d\}$
3: $\quad\mathcal{R}_p \leftarrow \emptyset$
4: $\quad\mathcal{T}_p \leftarrow \mathcal{R} \cap \{\mathcal{V}_p\}$
5: $\quad$**if** $\mathcal{T}_p \neq \emptyset$ **then**
6: $\quad\quad$Let $g \in \mathcal{T}_p$ be the nearest node to the middle of the path (with respect to the number of hops)
7: $\quad\quad\mathcal{R}_p \leftarrow \mathcal{R}_p \cup \{g\}$
8: $\quad\quad\Omega_p \leftarrow \Omega_p \cup \{p_{s-g}, p_{g-t}\}\backslash\{p\}$
9: $\quad$**end if**
10: $\quad$**while** $\Omega_p \neq \emptyset$ **do**
11: $\quad\quad$Take the first subpath q from Ω_p
12: $\quad\quad$**if** q meets OSNR **then**
13: $\quad\quad\quad\Omega_p \leftarrow \Omega_p\backslash\{q\}$
14: $\quad\quad$**else**
15: $\quad\quad\quad$**repeat**
16: $\quad\quad\quad\quad$Let q^* be a clone of q
17: $\quad\quad\quad\quad$Remove the last link (and node) from q^*
18: $\quad\quad\quad$**until** q^* meets OSNR
19: $\quad\quad\quad$Consider t_{q^*} as the regenerative node,
20: $\quad\quad\quad\mathcal{R}_p \leftarrow \mathcal{R}_p \cup \{t_{q^*}\}$
21: $\quad\quad\quad\Omega_p \leftarrow \Omega_p \cup \{q\backslash q^*\}$
22: $\quad\quad$**end if**
23: $\quad$**end while**
24: $\quad\mathcal{R} \leftarrow \mathcal{R} \cup \{\mathcal{R}_p\}$
25: **end for**

1) Complexity Remarks: The complexity of this algorithm is given by $O(|\mathcal{P}^o|\,(\frac{(\delta-1)(\delta-2)}{2}))$, where the second term is the upper bound on the maximum possible number of iterations required to create a feasible path, that is, when a regenerator is required at every node $v \in \mathcal{V}_p$, $p \in \mathcal{P}^o$. Such operation is performed once per path $p \in \mathcal{P}^o$, and hence, $|\mathcal{P}^o|$. Note that for all path $p \in \mathcal{P}^o$, $\delta \geq 3$ since T_{osnr} is dimensioned so that the feasibility of all network links (i.e., two-node paths) is always guaranteed.

E. Regenerator dimensioning phase

The load of burst traffic requiring regeneration at any node $v \in \mathcal{V}$ is (approximately) given by (11). In order to determine the number of regenerators required in node v, we define a dimensioning function $f(\rho_v^o, B_v^{osnr}) : (R^+, R^+) \mapsto \mathbb{Z}^+$. Under the assumption that any burst may access any regenerator in a node (as shown in Section III, the architecture proposed allows a fair access to the regenerator pool), we make use of the inverse of the Erlang B-loss function as the dimensioning function f. An straightforward way to implement this dimensioning function is to make use of vector $\mathbf{a}$ and Procedure 1, which have been both detailed in Section IV-D.

VI. Results and discussion

In this Section, we first present and compare the performance results of all the resolution methods to solve the RRPD

Parameter	NSFNET	Core	Base	Large
$\|\mathcal{P}_o\|$	35	18	109	282
$T_{osnr}[dB]$	18	20	20	20

TABLE II
NUMBER OF PATHS THAT REQUIRE REGENERATION AND OSNR THRESHOLD VALUES

problem presented in Section IV and Section V. Then, we study the performance of the T-OBS network architecture under some of the methods evaluated in order to prove that they are effective at keeping OSNR losses under control.

A. Resolution methods comparison

The evaluation has been performed by considering four different network topologies that are detailed in Appendix A. For this experiment and hereinafter in this paper, we consider the T_{osnr} values provided in Table II. Note that for the NSFNET network topology, due to larger link distance values, we had to consider a lower T_{osnr}. In particular, the highest T_{osnr} that guarantees that any link of the network is feasible (in terms of the OSNR signal quality) was selected. For the Pan-European networks we consider a value that is in accordance with recent studies (see e.g., [11]).

This parameter also determines the number of paths that require regeneration (i.e., $|\mathcal{P}_o|$), and hence, the level of complexity that is given to the problem. $|\mathcal{P}_o|$ values are also given in Table II for each considered network.

The results obtained are presented in Table III (number of regenerators) and Table IV (computation times). III also provides the number of regenerators required when an opaque network architecture is considered. In this study, each node injects into the network 11.2 erlangs into the network. One can note that MILP1 is solved very effectively when small instances are considered (i.e., NSFNET and Core). This is not, however, the case with both the Basic and the Large network, where MILP1 struggles several hours to reach poor solutions. In fact, if executions are not interrupted, they last until a memory error is dispatched, and worst, without achieving good enough solutions.

Among the heuristic MILP algorithms proposed, both the MILP1* and the MILP2/3 methods provide the most satisfactory near-optimal solutions. However, the trade-off between computation time and optimality is much more favourable to the latter due to the very large computation times of the former. Comparing the two heuristic algorithms proposed, it is easy to note that whilst LS outperforms RG in terms of the number of regenerators, RG has an extremely fast execution compared to all other methods considered. In fact, LS is even outperformed by MILP2/3 in terms of computation time.

From the results obtained in this Section, it can be deduced that MILP2/3 is the best method since it provides the best trade-off between optimality and execution times. In Figure 6, the number of regenerators required by the MILP2/3 method for some exemplary B_v^{osnr} and load values are shown. However, if computation resources are the top priority, the RG heuristic clearly outperforms all other methods considered.

Method	NSFNET	Core	Basic	Large
MILP1	112	55	499 ($>6\%$ gap)	971 ($>17\%$ gap)
MILP2/3	113	56	500	866
MILP1*	112	55	496 ($<2\%$ gap)	860 ($<2\%$ gap)
LS	112	55	556	932
RG	112	55	607	1021
OPAQUE	1344	1472	2624	2648

TABLE III
RP RESULTS COMPARISON

Method	NSFNET	Core	Basic	Large
MILP1	0.5	0.61	>7 hours	>11 hours
MILP2/3	0.1836	0.254	3.97	15.71
MILP1*	0.75	0.658	717	1864
LS	0.52	0.28	10.09	39.33
RG	0.086	0.116	0.37	0.55

TABLE IV
RP EXECUTION TIMES (SECONDS) COMPARISON

It is for these reasons that we use both the MILP2/3 and the RG heuristic in the next subsection in order to evaluate the performance of both methods when applied in the translucent OBS network architecture.

B. Impact on the OBS network performance

In order to evaluate the effectiveness of both the MILP2/3 and the RG methods, in this section, we conduct extensive simulations on the T-OBS network. In this study we consider the overall Burst Loss Probabilily (BLP) as the metric of interest. In Fig. 7, we show the results obtained under both the MILP2/3 and RG methods in the Large topology when the number of erlangs offered per node is equal to 6.4. In this experiment, two different B^{osnr} targets are considered, namely 10^{-3} and 10^{-5}. In addition, the opaque and transparent scenarios are plot and used as benchmarking indicators. It is easy to observe that the progressive and even placement of regenerators (i.e., the amount of regenerators to be placed is fairly distributed among all selected nodes) reduces the overall BLP until both B^{osnr} targets are reached (i.e., the required

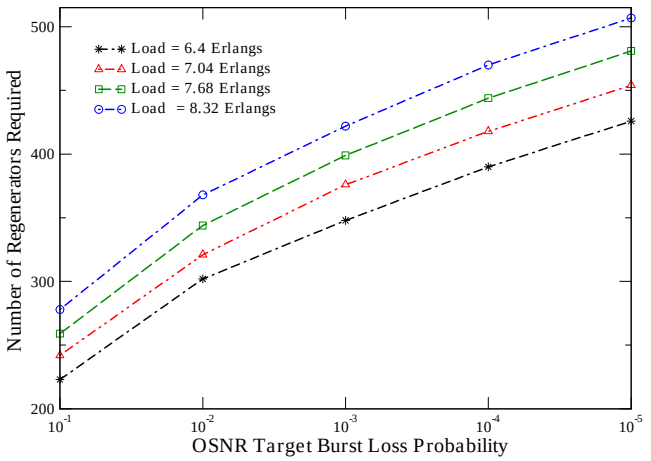

Fig. 6. MILP2/3 placement method results for some exemplary OSNR target burst loss probabilities and network load values.

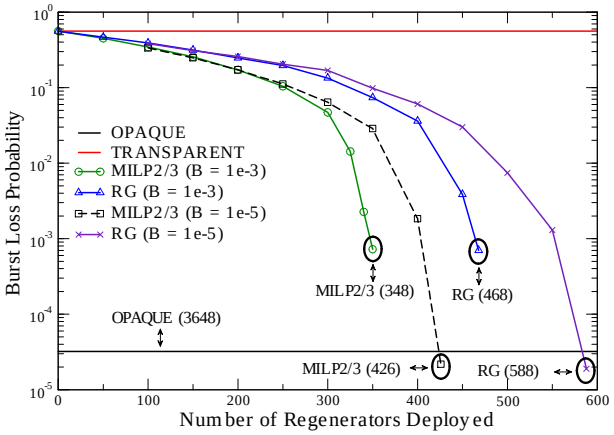

Fig. 7. MILP2/3 vs. RG performance comparison in the Large topology.

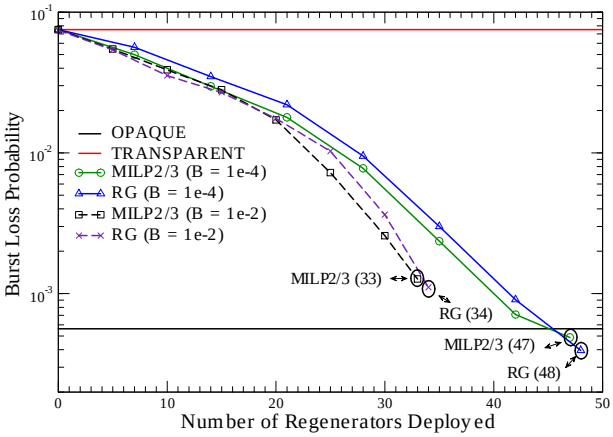

Fig. 8. MILP2/3 vs. RG performance comparison in the Core topology.

number of regenerators has been deployed). As it was to be expected, MILP2/3 reaches both B^{osnr} targets well before than RG. In the $B^{osnr} = 10^{-3}$ case the BLP is dominated by OSNR losses, and consequently, when all the regenerators have been deployed $BLP \approx B^{osnr}$. On the other hand, if B^{osnr} is set to 10^{-5}, contention losses become predominant, and therefore, $BLP \approx BLP_{OPAQUE}$.

Similarly, Fig. 8, shows the same experiment performed in the Core topology. However, this time each edge node offers 12.8 erlangs and B^{osnr} targets are set to 10^{-2} and 10^{-4}. It is worth pointing out that both the load and B^{osnr} values were selected in order to illustrate two different and representative situations in both figures.

In Fig. 8 both methods require nearly the same amount of regenerators, and thus, their performance is quite similar in both B^{osnr} cases. Notice that, in the $B^{osnr} = 10^{-2}$ case, although OSNR lossess have a noticeable impact on the network performance, the BLP decreases up to nearly 10^{-3}. This is due to the fact that the percentage of the traffic requiring regeneration in the network is quite low, or in other words, $|\mathcal{P}^o|$ has a small size. If B^{osnr} is set to 10^{-4}, in contrast, we observe the same behaviour as in Fig. 7, that is, contention losses are predominant, and hence, $BLP \approx BLP_{OPAQUE}$. Note that in both figures provided, the BLP found in the case where contention losses are predominant slightly improves that of the opaque case. This is due to the differences in node architectures between the opaque and translucent networks: whilst the opaque network relies on in-line regenerators as in [16], our translucent architecture operates in the feed-back mode as proposed in [8].

Eventually, we assess how effective at keeping OSNR losses under control the MILP2/3 method is. For this purpose, we study how both contention and OSNR losses contribute to the total BLP. In Fig. 9 the impact that the load injected into the network has on both types of burst loss is made clear. Note that whilst in the bottom x-axis the load is considered, in the top one the number of regenerators placed is shown. It is easy to observe that with the load increase, contention losses, which are the main source of performance degradation in OBS networks, become dominant. On the contrary, OSNR losses are kept satisfactorily under control regardless of the load.

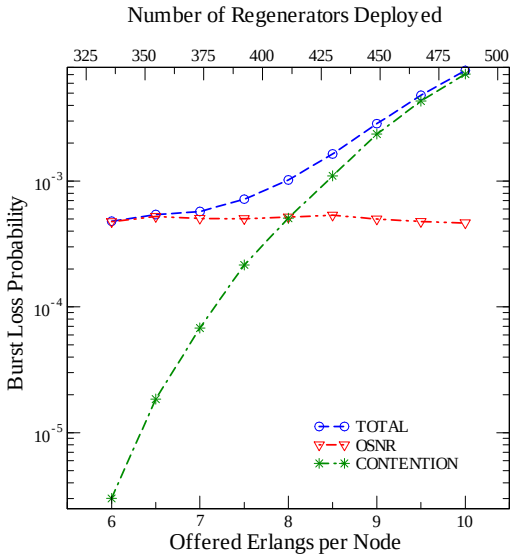

Fig. 9. MILP2/3 performance evaluation in the Large topology.

VII. Conclusions and Future Work

In this paper, we propose several methods for the sparse placement of regenerators in a translucent OBS network. Such methods are based either on MILP or heuristic techniques. For this purpose, we have focused on the problem of PLIs in OBS networks. In particular, we have proposed a novel T-OBS network architecture consisting of all-optical TAS nodes equipped with a limited number of O/E/O regenerators. Then, we have provided an OSNR model to evaluate the impact of the main PLIs (i.e., ASE noise and splitting losses) and illustrated a method to compute a power budget and noise analysis between two TAS OBS core nodes.

Then, this model has been used to address the RRPD problem. To be precise, we have uncoupled the routing issue from the RPD problem, and eventually solved the so-called R+RPD problem. We have presented a link congestion-reduction unsplittable routing strategy which is based on a MILP formulation aimed at reducing congestion in bottleneck network links. The routing solution obtained has then been

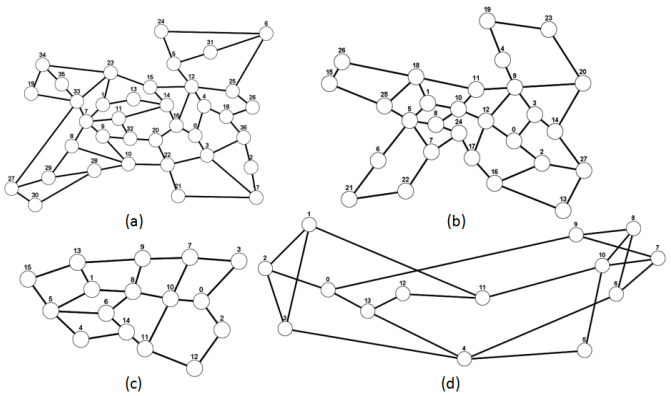

Fig. 10. a) Large (37 nodes), b) Base (28 nodes), c) Core (16 nodes), d) NSFNET (14 nodes).

used as input for the RPD problem. The RPD scheme presented relies on the piecewise linear approximations of the inverse of the Erlang-B loss formula. Since such formulation corresponds to the complex DCMCF problem, we have also developed several heuristic methods to help solve the RPD problem (i.e., RP+D heuristics). We have evaluated and compared these methods by considering the trade-off between optimality and complexity they provide. Among them, the load-based formulation (MILP2/3) stood out from the rest as the best trade-off, and the regenerator-grouping (RG) heuristic as the fastest method.

Finally, we have conducted a series of exhaustive simulations in the T-OBS network proposed considering both the MILP2/3 and RG methods. From the results obtained, we have concluded that both the architecture and model proposed in this paper ensure that, according to a pre-specified target performance, losses caused by OSNR signal degradation are kept satisfactorily under control and do not impact negatively the overall network performance.

In our future work, we plan to extend our model to consider the case of an on-line/dynamic scenario.

Appendix A
Simulation Scenario

In our simulation scenario, we consider several topologies (see Fig. 10), all of which being real network topologies: a set of Pan-European [28] networks known as: Large (a), Basic (b) and Core (c) with 37, 28 and 16 nodes and 57, 41 and 23 links respectively; an American backbone network called NSFNET [29] (d) with 14 nodes and 21 links.

Network links are bidirectional and dimensioned with the same number of wavelengths $M = 32$. The transmission bitrate is set to $10Gbps$.

We assume that each node is both an edge and a core bufferless node capable of generating bursts destined to any other nodes. We consider the offset time emulated OBS network architecture (E-OBS) [30] and the Just-In-Time (JIT) [31] resources reservation protocol. For the sake of simplicity, the switching and processing times are neglected.

The traffic is uniformly distributed between nodes. We assume that each edge node offers the same amount of traffic to the network; this offered traffic is normalized to the transmission bitrate and expressed in Erlangs. In our context, an Erlang corresponds to the amount of traffic that occupies an entire wavelength (e.g., 20 Erlangs mean that each edge nodes generates $200Gbps$).

Bursts are generated according to a Poisson arrival process and have exponentially distributed lengths. The mean duration of a burst is $100\mu s$ ($1Mb$).

All simulations have been conducted on the JAVOBS [32] network simulator on an Intel Core 2 Quad 2.67 GHz with $4GB$ RAM.

The RMILP1, RMILP2, MILP1, MILP2, MILP3 and MILP1* problems have all been solved using the IBM ILOG CPLEX v.12.1 solver [33].

Acknowledgment

The work described in this paper was carried out with the support of the STRONGEST-project, an Integrated Project funded by the European Commission through the 7th ICT-Framework Programme, the Spanish Ministry of Science and Innovation under the "DOMINO" project (Ref. TEC2010-18522), the Catalan Government under the contract SGR-1140, and the Polish Ministry of Science and Higher Education under the contract 643/N-COST/2010/0.

References

[1] S. Azodolmolky et al., “A survey on physical layer impairments aware routing and wavelength assignment algorithms in optical networks,” *Com. Netw.*, vol.53, no.7, May 2009.

[2] B. Ramamurthy et al., “Transparent vs. opaque vs. translucent wavelength-routed optical networks,” *IEEE/OSA OFC’1999*, Feb. 1999.

[3] X. Yang and B. Ramamurthy, “Sparse regeneration in translucent wavelength-routed optical networks: Architecture, network design and wavelength routing,” *Phot. Netw. Commun.*, vol.10, no.1, Jan. 2005.

[4] S. Pachnique, T. Paschenda, and P.M. Krummrich, “Physical Impairment Based Regenerator Placement and Routing in Translucent Optical Networks,” in *Proc. OFC/NFOEC 2008*, San Diego, USA, Feb. 2008.

[5] W. Zhang et al., “REPARE: Regenerator Placement and Routing Establishment in Translucent Networks,” in *Proc. IEEE GLOBECOM2009*, Nov. 2009.

[6] Y. Lee, G. Bernstein, D. Li, G. Martinelli, “A Framework for the Control of Wavelength Switched Optical Networks (WSON) with Impairments”, *IETF draft*, draft-ietf-ccamp-wson-impairments-02.txt, May 2010.

[7] C. Qiao, M. Yoo, Optical burst switching (OBS)-A new paradigm for an optical Internet, *J. High Speed Netw.*, vol. 8, no. 1, Jan. 1999.

[8] O. Pedrola, D. Careglio, M. Klinkowski, and J. Solé-Pareta, “Modelling and performance evaluation of a translucent OBS network architecture,” in *Proc. IEEE GLOBECOM 2010*, Miami, USA, Dec. 2010.

[9] G. Shen and R. S. Tucker, “Translucent optical networks: the way forward,”, *IEEE Commun. Mag.*, vol.45, 2007.

[10] A. Sen, S. Murthy, S. Bandyopadhyay, “On Sparse Placement of Regenerator Nodes in Translucent Optical Network,” in *Proc. IEEE GLOBECOM 2008*, Nov. 2008.

[11] E. Salvadori et al., “Distributed Optical Control Plane Architectures for Handling Transmission Impairments in Transparent Optical Networks,” *J. Lightwave Technol*, vol.27, no.13, 2009.

[12] R. Martínez, R.Casellas, R. Muñoz and T. Tsuritani, "Experimental translucent-oriented routing for dynamic lightpath provisioning in GMPLS-enabled wavelength switched optical networks", *J. Lightwave Technol.*, vol.28, no.8, Jul. 2009.

[13] G. Bernstein, Y. Lee, and B. Mack-Crane, “WSON Signal Characteristics and Network Element Compatibility Constraints for GMPLS,” *IETF draft*, draft-ietf-ccamp-wson-impairments-02.txt, Oct 2019, work in progress.

[14] Y. Fan, B. Wang, “Physical impairment aware scheduling in optical burst switched networks,” *Phot. Netw. Commun.*, vol.18, no.2, Oct. 2009.

[15] B.G. Bathula, V.M. Vokkarane, R.R.C. Bikram, “Impairment-Aware Manycasting over Optical Burst-Switched Networks,” in *Proc. IEEE ICC2008*, May 2008.

[16] H. Buchta and E. Patzak, “Analysis of the Physical Impairments on Maximum Size and Throughput of SOA-Based Optical Burst Switching Nodes, ”, *J. Lightwave Technol.*, vol.26, no.16, Aug. 2008.

[17] M. Mestre et al., “Tuning Characteristics and Switching Speed of a Modulated Grating Y Structure Laser for Wavelength Routed PONs, ”in *Proc. ANIC 2010*, Karlsrhue, Germany, June 2010.

[18] H. Pereira et al., “OSNR model to consider physical layer impairments in transparent optical networks,” *Phot. Netw. Commun.*, vol.18, no.2, Oct. 2009.

[19] R. Martínez et al., “Experimental GMPLS routing for dynamic provisioning in translucent wavelength switched optical Networks,” in *Proc. OFC/NFOEC 2009*, San Diego, USA, Mar. 2009.

[20] T. Tsuritani et al., “Optical Path Computation Element interworking with Network Management System for Transparent Mesh Networks,” in *Proc. OFC/NFOEC 2008*, San Diego, USA, Feb. 2008.

[21] INPHENIX: *http://www.inphenix.com/soa_devices.html*, accessed in Mar. 2010.

[22] MRV: *http://www.mrv.com/product/MRV-LD-OAB*, accessed in Mar. 2010.

[23] M. Klinkowski et al., “An overview of routing methods in optical burst switching networks,” *Opt. Swit. Netw.*, in Press, available online Jan. 2010.

[24] R. Syski, “Introduction to Congestion Theory in Telephone Systems,” *North-Holland*, 1960.

[25] M. Minoux, “Mathematical Programming: Theory and Algorithms,” *John Wiley and Sons*, 1986.

[26] M. Minoux, “Discrete Cost Multicommodity Network Optimization Problems and Exact Solution Methods, ” *Annals of Operations Research*, vol.106, no.1, 2001.

[27] J. Kleinberg, E. Tardos, “Algorithm Design,” *Addison-Wesley*, 2005.

[28] S. De Maesschalck et al., “Pan-European Optical Transport Networks: An Availability-based Comparison,” *Phot. Netw. Commun.*, vol.5, no.3, May 2003.

[29] K.C. Claffy, G.C. Polyzos, H.W. Braun, “Traffic characteristics of the T1 NSFNET backbone,” in *Proc. IEEE INFOCOM93*, San Francisco, USA, Apr. 1993.

[30] M. Klinkowski et al., “Performance overview of the offset time emulated OBS network architecture,” *J. Lightwave Technol.*, vol.27, no.14, pp. Jul. 2009.

[31] J. Y. Wei, R. I. McFarland, “Just-in-time signaling for WDM optical burst switching networks,” *J. Lightwave Technol.*, vol. 18, no. 12, Dec. 2000.

[32] O. Pedrola et al., “JAVOBS: a flexible simulator for OBS network architectures,” *J. Netw.*, vol.5, no.2, Feb. 2010.

[33] IBM ILOG CPLEX: *http://www-01.ibm.com/software/integration/optimization/cplex/*, accessed in May 2010.