This is an author produced version of *Energy-efficient resilient optical networks: Challenges and trade-offs*.

White Rose Research Online URL for this paper: http://eprints.whiterose.ac.uk/87957/

**Article:**

http://dx.doi.org/10.1109/MCOM.2015.7045403
Energy-Efficient Resilient Optical Networks: Challenges and Trade-offs

Yabin Ye, Felipe Jiménez Arribas, Jaafar Elmirghani, Filip Idzikowski, Jorge López Vizcaíno, Paolo Monti, Francesco Musumeci, Achille Pattavina, Ward Van Heddeghem

1) Huawei Technologies Duesseldorf GmbH, Munich, Germany
2) Telefónica I+D, Madrid, Spain
3) University of Leeds, Leeds, UK
4) Sapienza University of Rome, Rome, Italy
5) KTH Royal Institute of Technology, Kista, Sweden
6) Politecnico di Milano, Milan, Italy
7) iMinds, Ghent, Belgium

Abstract

Energy-efficiency and resiliency are two well-established research topics in optical transport networks. On the other hand, their overall objectives (i.e., power minimization and resource-utilization/availability maximization) are in contrast. In fact, provisioning schemes optimized for best resiliency performance are in most cases not energy-efficient in their operations, and vice versa. However, very few works in the literature consider the interesting issues that may arise when energy-efficiency and resiliency are combined in the same networking solution. The objective of this paper is to identify a number of research challenges and trade-offs for the design of energy-efficient and resilient optical transport networks from the perspective of: long-term traffic forecasts, short-term traffic dynamics, and Service Level Agreement (SLA) requirements. We support the challenges with justifying numbers based on lessons learnt from our previous work. The paper also discusses suitable metrics for energy-efficiency and resiliency evaluation, in addition to a number of steps that need to be taken at the standardization level to incorporate energy-efficiency into already existing and well-established protocols.

1 Introduction

Traffic in core networks has been growing at a yearly rate of 45% since 2004 [1]. The ever increasing traffic levels have forced operators to focus on ways to increase the network capacity accordingly. However, such an increase is expected to bring higher energy consumption. The worldwide electricity consumption of communication networks was estimated to be 350 TWh in 2012 (or nearly 2% of the worldwide electricity consumption) showing an average annual growth rate of 10% since 2007 [2]. Therefore, energy-efficiency is becoming one of the key design parameters for planning and operating today’s telecommunication networks. In this regards, optical transport solutions might be beneficial because they reduce the number of opto/electrical/opto (O/E/O) operations, which are very costly from power consumption point of view.

A critical issue when dealing with optical transport networks is the amount of traffic that can be disrupted by a network fault, an aspect that needs to be addressed properly if certain Quality of Service (QoS) levels are to be guaranteed to the end-user. This makes it necessary to reserve some redundant resources to ensure network resiliency, i.e., the ability of the network to provide and maintain an acceptable level of service in the face of different faults. These redundant resources are then used to re-route the traffic, bypassing the failed network element(s), over the so-called backup path(s).

In general, resilient schemes can be divided into two categories: restoration or protection. Their main difference lies in the way they compute the backup paths. In a restoration approach, the backup paths are computed “on the fly”, i.e., only after a failure occurs, while in a protection approach the backup paths are pre-computed, with the backup resources already reserved and ready to be used in the occurrence of a failure. Both schemes have their pros and cons. Restoration allows for a more efficient use of network resources, but has longer recovery time and does not provide a 100% recovery guarantee against failures. Protection schemes, on the other hand, have a faster recovery time and can

Part of this work was performed when Filip Idzikowski was with Technische Universität Berlin.
guarantee 100% recovery from the failure scenario for which they were designed, but require more energy in absolute terms. Protection schemes can be further categorized in the way backup resources are used. 

Dedicated schemes do not allow sharing of backup resources among multiple backup paths. 

Shared schemes, under specific conditions, allow some backup paths to use the same wavelength resources. Dedicated protection (DP) can be implemented in a 1+1 or a 1:1 fashion. In the former, the traffic is duplicated over two disjoint paths, and one of the paths is selected by the destination node, whereas in the latter, the disjoint backup path may carry traffic that can be pre-empted in case of a failure happening in the working path.

Operators tend to implement the resiliency concepts presented above following three high-level mechanisms. The first one is network overprovisioning, i.e., the duplication of a number of line cards in a node, up to the duplication of complete geographically separated routers, nodes, and fiber links. In the latter case, the end result will be a dual plane network at one or several layers. The second mechanism is to use 1+1 protection at the optical layer, which can provide fast switch-over time (< 50 ms) when a failure occurs. On the other hand, when the backup path carries the switched-over traffic (and the failed element along the primary path is under repair), this traffic is in a vulnerable state, i.e., it is unprotected against any additional failures. For this reason, while the traffic is carried by the backup path, restoration is used to replace the old (failed) working path. This is the third option, i.e., dedicated protection plus restoration. After a new backup path is computed, the network is again protected against any single (link or node) failure scenario, with the original backup path serving as the current working path, and the repaired (old) working path now acting as the backup path. Resilient networks rely on a number of duplicated resources, which are unused most of the time. This is obviously not the most energy-efficient solution.

Core networks have always been designed focusing on the minimization of the network equipment and the deployment cost, i.e., Capital Expenditures (CapEx). Given a certain request of data traffic, the definition of a network topology, and the set of equipment to be installed, a design procedure determines the processing capability of the switching/routing nodes on the one hand, and the capacity of the inter-node transmission links on the other hand. However, reducing Operational Expenditures (OpEx) is becoming increasingly important for operators, and energy consumption is one of the key contributors. Despite the numerous studies focused on resiliency versus cost or QoS, energy consumption cannot simply be interchanged with cost and QoS, so that previous techniques cannot be directly applied to the problem of energy-efficient resilient optical network design. Firstly, energy-efficiency has to be traded against QoS. A simple cost minimization approach may save energy by shaping traffic over fewer routes, but on the other hand reduce the network reliability due to the increased impact of a cut in a consolidated route. Secondly, network devices implementing sophisticated energy saving functionalities could be initially more expensive than the conventional ones. Finally, backup (and working path) routes may be designed to use time-varying renewable energy sources efficiently and/or adapting the transmission to the instantaneous traffic demand, which reduce the network carbon footprint, but may not reduce or minimize the costs directly.

Two facts should be considered for an energy-efficient resiliency design: (1) core networks are dimensioned assuming peak traffic levels; (2) the type of protection, i.e., service guarantees, can be differentiated and adapted to the specific traffic type. Both observations provide opportunities to reduce the energy consumption. As for the peak-traffic dimensioning, real traffic demands vary over time, i.e., the traffic demands at night are usually much smaller than those during the day. This means that unused resources (line cards in nodes or even full nodes) can be put into sleep to save energy. As for the service guarantees, not all traffic types need the same level of protection, and therefore resource redundancy for service guarantees can be restricted to a subset of traffic types, thus again leading to further energy savings. In other words, in order to improve the energy-efficiency of resilient optical core networks, there is a clear need for design and provisioning strategies specifically tailored to reduce the network energy consumption.

Certainly, optical network design is complicated and needs to balance among a number of metrics, e.g., cost, energy, resiliency, and scalability, etc. The purpose of this paper is to identify the major research challenges in the relatively new field of energy-efficient resiliency in optical transport networks. In this respect, our objective, rather than presenting univocal solutions, is to highlight and reason around the nature of the potential options and the challenges that they present, e.g., their performance trade-offs.
2 Research topics / Open problems

Core networks are generally designed and provisioned in response to three main inputs, i.e., long-term traffic forecasts, short-term traffic dynamics, and Service Level Agreement (SLA) requirements. These factors influence the design and the choice of protection and restoration mechanisms as well as the energy-efficiency performance. In this section, we provide an overview to address these long-term static network architecture design choices, the protocols and hardware functionalities needed for adaptation to the short-term traffic dynamics, and the strategies needed to meet the SLA requirements in an energy-efficient fashion. Most of the presented challenges are justified by numerical results achieved in our previous work (though in different network and traffic scenarios).

![Diagram of network architectures](image)

Figure 1: (a) 1+1 protection at the IP layer, (b) 1+1 protection at the optical layer, (c) employing optical bypass at the IP layer, (d) grooming at the Optical Transport Network (OTN) layer. (LC = line card, TXP = transponder, OXC = optical cross connect, OLA = optical line amplifier, wp = working path, bp = backup path)

2.1 Long-term static architecture network design choices

Network Architecture — Current optical networks are based on wavelength division multiplexing (WDM) technologies, which operate over fixed channel spacing defined by the International Telecommunications Union –Telecommunication sector (ITU-T). WDM network architectures may consider the Single Line Rate (SLR) transmission in all the wavelength channels, or Mixed Line Rate (MLR) transmissions, where channels transmitted with different rate coexist on the same fiber. Elastic Optical Networks (EONs), which allow for flexible-bandwidth transmissions and adaptive modulation, have emerged as the future technology for the optical network. The choice of a particular network architecture over another, together with the adopted resilient scheme (i.e., restoration versus protection, or dedicated versus shared protection), will have clear effects in energy consumption. For instance, changing the protection scheme from DP 1+1, the most reliable and more energy-consuming scheme, to more energy-efficient ones such as DP 1:1 or SP will result in different energy savings depending on the network architecture. E.g., adopting SP or DP 1:1, the energy consumption of the optical layer can be reduced up to 49% with respect to DP 1+1 in EON and up to 42% in WDM with a SLR of 40 Gbps [3].
Embodied energy — The energy consumed during the whole lifetime of an installed device needs to be considered (including e.g., manufacturing and decommissioning) in the design of optical transport networks. In fact, the embodied energy accounts for approximately 70% of the total energy consumption of the network (i.e., operational plus embodied energy consumption) [4]. The fiber has actually the highest impact in terms of embodied energy [4] due to the large amount of material used to protect the fibers in a cable and the hundreds of thousands of kilometers of fiber cable in a core network. Therefore, from an energy perspective, it is especially important to minimize the number of redundant fibers in resilient networks.

Resiliency at different layers — Core networks are typically organized as multilayer networks where each layer can be viewed as a single network. Higher layers rely on the resources and services provided by the lower layers and resiliency can be provided at any layer. Fig. 1(a) illustrates 1+1 protection in the IP layer, whereas Fig. 1(b) shows the same traffic demand protected at the optical layer. Note the subtle difference of the latter with the employment of optical bypass in combination with 1+1 protection at the IP layer shown in Fig. 1(c). If protection is provided simultaneously at multiple layers and there is no coordination between the layers, parallel recovery actions may take place, which can have a significant impact on the overall network stability, and lead to sub-optimal resource usage. Multilayer recovery schemes have been explored in earlier research, however, it is not yet clear how these strategies perform from an energy-efficiency point of view. Intuitively the best option is to provide protection at the (most energy-efficient) bottom layer, perhaps augmented with a bottom-up escalation strategy where recovery starts in the lowest detecting layer and escalates upwards. However, this strategy comes with the drawback of poorer handling of higher-layer failures, and the impact of escalation timings on the overall recovery time has to be assessed. Furthermore, network operators often duplicate their backbone networks and simultaneously operate two nearly-identical networks. With 1+1 protection often used at the Optical Transport Network (OTN) layer as shown in Fig. 1(d), this means in effect that each requested demand results in four times the capacity at the optical layer. Energy-wise, this is clearly not an optimal situation.

Topologies — Looking at a single layer, the network topology should keep a relatively high number of alternative paths for the traffic demands coming from the higher layers. Although a full-meshed and highly over-provisioned topology is ideal from the perspective of protection, the number of deployed links and their capacities should be traded against energy consumption. This trade-off was considered in [5] for physical topology design. The constraint on minimum nodal degree equal to 2 (securing at least one alternative path) leads to negligible increase of network power consumption (0.5% increase with respect to a network with minimum nodal degree equal to 1 using multi-hop bypass approach and symmetric traffic demand for the NSFNET network). Furthermore, the size of the network will also play an important role when trading energy-efficiency and resiliency. For instance, if some links become highly occupied due to the traffic growth, backup paths may become much longer and require the usage of signal regeneration or more robust modulation format, with the consequent increase in energy consumption.

2.2 Open issues in adaptation to short-term traffic dynamics

Novel equipment features — The nominal power consumption of devices/equipment considered for installation must be assessed during the network design phase. Emerging equipment features—such as sleep modes and dynamic reconfiguration of modulation format and transmission reach—could provide new opportunities to reduce the power consumption of backup links. For example, more energy-efficient protection schemes could be devised to exploit sleep modes at the expense of slower recovery. However, to the best of our knowledge, there is no current information of any commercial device implementing these innovative features, making all related work reporting results based on assumptions. New protection schemes could also allow multi-rate transponders to fallback to lower transmission rates depending on the SLA and reduce energy consumption. This is related to Section 2.3, where we discuss quality of protection (QoP) differentiation. Moreover, there are open issues which must be addressed: 1) Are complex devices more prone to failures? 2) How quickly can the devices be brought up and down to react to traffic variations? 3) How can the energy consumption information of the device and electricity cost be accurately monitored?

Tradeoffs in dynamic adaptation to temporal variation of traffic — Traffic in core networks is usually higher during the day and lower during the night. Putting idle devices into sleep or energy-saving mode (e.g., adaptation of transmission rates) can effectively reduce the energy consumption of the backup resources (e.g., adapting the backup transmission in DP 1+1 can save up to 22% of energy
consumption at the optical layer during the low-traffic hours of the weekend [3]). Another strategy relies on concentrating backup paths into separate fibers, as to be able to put the devices on these links into sleep mode without being constrained by the presence of working paths. This is applied to DP [6] and SP [7] with power savings of up to 35% with respect to the non sleep-mode scenarios. Moreover, the sleep mode functionality can also be exploited considering grooming at the electrical layer with energy savings of up to 60% in both the optical and IP layers [8]. Nevertheless, for these mechanisms, it will be important to determine the best trade-off between resiliency, energy-efficiency, and low reconfiguration costs.

Geographical traffic distribution — Geographical traffic distribution in networks covering large areas may be diverse, especially in networks spanning several time zones. Is it possible to use idle resources from one time zone as protection resources in another time zone? Traffic distribution can also suffer strong variations due to external influences like natural disasters or wide audience events. Extra redundancy can be provided to face the potential traffic changes, but it will be costly. Therefore, how to handle this potential variation from the protection point of view is an open question. There are some works that differentiated traffic in different time zones (e.g., [5]) or used traffic data originating from measurements (e.g., [9]). However, to the best of our knowledge, there is no related work focused on geographical traffic distribution and green networking.

Granularity of traffic demands — The size of a single traffic demand determines the possibility of grooming it into a set of lightpaths and links to utilize the available network resources in an efficient way. However, grooming of multiple traffic demands over a limited number of network resources poses a challenge from the resiliency point of view, since some traffic demands need to traverse multiple hops from their origin to their target. This, in turn, increases the number of devices on their path, each being prone to failures. This leads to important questions relating to the maximum granularity of traffic demands (with respect to network resources) so that traffic grooming for energy saving pays off without violation of protection constraints. While an initial exploration on this issue (see Fig. 7 in [9]) indicates that grooming can be more energy-efficient when the granularity of the traffic demands is below half of the linerate, the cited work does not look at the impact of grooming on the availability (in terms of relative uptime per year).

2.3 Strategies to meet the SLA requirements

Physical impairments — Network design techniques have to consider physical impairments (e.g., fiber loss, dispersion or nonlinear effects) and their impact on the resulting Quality of Transmission (QoT). Some energy saving techniques can exacerbate the effect of some physical impairments (e.g., cross phase modulation and cross talk) [10], as they tend to concentrate most of the traffic on a few links to put unused devices into sleep mode. This problem can be overcome by introducing design techniques that are both energy and impairment aware. The work [10] shows that such techniques are able to achieve the same energy savings of conventional (i.e., non impairment aware) green strategies while providing QoT levels close to those provided by the impairment aware design strategies (i.e., only a small percentage of the lightpaths cannot be established due to insufficient QoT). Even though the work in [10] considers an unprotected scenario, we can expect that this aspect will be even more critical when designing protected networks, where the backup paths are on average longer than their respective working paths. Power transient phenomena will also have to be considered when setting network elements into sleep mode and adequate mechanisms (e.g., disabling the transponder electronics while keeping the injected optical power) must be provided. Moreover, the continuous development of high-speed systems comes at the cost of reduced reach. Consequently, a higher number of O/E/O regenerations (highly energy consuming processes) would be needed for a generalized deployment. Employing more robust modulation formats for long backup paths can be an alternative to regeneration, but at the expense of requiring higher energy per bit [3].

Differentiated Quality of Protection — Commonly, a single policy is applied to all aggregated traffic demands as shown in Fig. 1. An alternative to this policy would be to apply Differentiated QoP by assigning different resiliency levels to demand “subsets” having different SLA needs [11]. These “subsets” could be allocated, for instance, to different residential services or to different performance classes of a corporate service (like optical Virtual Private Networks - VPNs). The application of Differentiated QoP allows for a reduction of protection resources, which may result in improved energy efficiency with respect to the conventional DP 1+1 scheme (Energy Efficiency per GHz (EEPG) improvements of up to 300% are reported in [11]). Another possibility is to adapt the reliability performance of a given protection scheme to the reliability requirements of the provisioned traffic, i.e.,
the Differentiate Reliability (DiR) concept. Thus, if for some traffic demands, the backup path does not need to be always available for any possible failure scenario, it will be possible to selectively assign protection resources only to those demands that need them the most. This approach can lead to significant energy savings (i.e., up to 25% [12]) with respect to conventional protection strategies that are always 100% survivable.

**EON/BVT** — EON technologies provide an extra degree of freedom on assigning traffic demands to QoP levels. E.g., a sliceable Bandwidth Variable Transponder (BVT) based on a multcarrier approach could be used to assign a number of carriers to a specific QoP. This can allow for optimizing the resource utilization and providing a finer granularity on QoP mapping. Besides, a BVT can adapt the transmission rate (modifying the bandwidth and/or modulation format) to the traffic variations, and thus reduce energy consumption. EON is shown to provide much higher EEPG than traditional fixed-grid WDM technologies with any protection scheme (e.g., EEPG can be improved up to 200% with respect to MLR with SP [3]).

**Network Virtualization** — Network resources can be partitioned over different operators and/or services. In that respect, the physical network would comprise of a number of virtual networks, each one receiving a different degree of resiliency, according to the amount of resources allocated to each service/operator. The virtualization of the network components combined with the creation of segregated virtual topologies with different architectures (including different routing and resilient protocols) can help in supporting variable QoP levels and thus reducing the energy consumption by sharing the resources among different users/virtual networks. The differentiation can be achieved in one or more dimensions including time delay, availability, protection scheme / protection level, QoT and the network layer selected (i.e., electrical and/or optical). Despite numerous works that consider realization of a virtual network over a physical one, only a few of them consider network virtualization and protection. E.g., 28% of power can be saved by allowing virtual links to bypass physical nodes in an IP-over-WDM network (multi-hop bypass compared to non-bypass approach for NSFNET network at 6:00 pm, Fig. 5 of [5]) providing protection in the physical layer by keeping minimum nodal degree equal to 2. Protection is not the focus of [5] though.

**Data Pre-emption** — Pre-emption is based on the intuition that low priority services can be provisioned with the option to be discarded, should a connection with a higher criticality level need to use the network resources. This could happen after a failure concatenation or some unexpected or very infrequent event. In this case, lightpath pre-emption levels are used to decide which signal should be allocated first, releasing the resources in use by lightpaths with lower pre-emption levels. We can consider this mechanism as a subset of a QoP scheme. This, of course, has to be done without compromising the required QoS level of each connection. Pre-emption based strategies can improve energy-efficiency, since powering-on extra resources is not always necessary. For instance, in a scenario where pre-emption is used to dynamically provision the service subject to different QoP requirements, redundant resources used for protection could be used to convey extra traffic if the carried traffic is able to tolerate a disruption, should one or more failures occur in the network. This will be accomplished without unnecessarily turning on extra resources in the network, with evident benefits in terms of energy efficiency. How effective is the use of such energy-efficient pre-emption techniques is still an open question which, to the best of our knowledge, has not been studied so far. The answer will depend on a series of factors such as the protection mechanism used, and the traffic composition, just to name a few. Higher power reduction can be expected in the presence of traffic with highly diverse survivability needs, as well as with protection techniques that are less stringent in their requirements. Finally, there is one last aspect to consider. If backup resources are used to route pre-emptable traffic, they need to stay in an active state, potentially nullifying the benefits of some energy-efficient protection techniques. There is obviously a trade-off that needs to be assessed here to understand under which conditions pre-emption will be beneficial from an energy reduction point of view.

### 3 Metrics and standards

This section considers how issues related to energy-efficiency and resiliency can be evaluated and what this would mean in terms of standardisation requirements.

**Metrics** — The energy-efficiency of a network with various resiliency schemes can be assessed with common metrics such as W/bps (or J/b), or the inverse of both. With this approach, less protective schemes (e.g., SP) will result in more energy-efficient designs than more protective schemes (e.g., DP
While this metric is a good indicator for the actual power or energy required to transport a bit of information, it does not take into account the level of protection. To do so, it might make sense to express the energy-efficiency normalized by the protection factor, where the protection factor could be, e.g., 2 for 1+1 DP, 1.5 for SP, etc. This metric provides a fairer indication of the energy required for a given level of protection. Finally, with the potential evolution towards more flexi-grid equipment, common W/bps metric (or its derivatives) might be insufficient to capture the efficiency in utilizing the available spectrum. For example, a BVT can make use of adaptive bandwidths depending on the required transmission rate and distance. Therefore it might make sense to take into account the spectral-efficiency of transmission with an energy-efficiency per GHz metric, as proposed in [3].

**Standards** — Putting unused network devices into sleep mode is one way to save power. However current Internet protocols operate based on the assumption that network elements are always-on. The applications and services re-initializations when these devices wake up again would potentially result in a non-negligible amount of signaling overhead [13]. Modifications to the control plane considering links or nodes removal should provide the ability to choose the level of redundancy available after the network topology has been trimmed. The complete removal of nodes or links from the network topology has several impacts on the control plane which must be considered [14]. For example it is essential to modify the network topology so that the removed links or devices are not used to forward traffic remembering that such links exist, possibly including the neighbors and destinations reachable through those links or devices. One solution to this sleep mode problem could be based on the use of a proxy [13]. Before going into sleep, a node delegates its functionalities to such a proxy, which will then respond to routine network traffic on behalf of the sleeping node and it will wake the node up when needed. The protocols and procedures for proxy operation such as discovery, selection, delegation and wake-up have to be defined. Another example would be to require that nodes can negotiate timeouts (in protocols that make use of timeouts), so that a node might be able to go into sleep mode or to attempt to synchronize periodic messages across a number of protocols. Thus, all these messages fall into a certain timeframe and in between the node can sleep. The issues described above can also be addressed by designing sleep-compatible protocols or by extending existing protocols (where possible) to include the ability to distinguish sleeping elements from failed ones. Some extensions required in existing Generalized Multi-Protocol Label Switching (GMPLS), Open Shortest Path First (OSPF) routing protocol, Resource ReserVation Protocol (RSVP) signaling protocol, and Link Management Protocol (LMP) are proposed in [15] to support energy-efficient traffic engineering technology.

4 Conclusion

Both energy-efficiency and resiliency in telecommunications networks are well established topics in the research community. However, the combination of both for energy-efficient resilient optical network design is still a relatively new research field. In this paper, we identify the corresponding major research challenges and performance trade-offs for core networks from three different aspects: long-term traffic predictions (including network architecture, embodied energy, resiliency at different layers and topologies), short-term traffic dynamics (including novel equipment features, tradeoffs in dynamic adaption to temporal variation of traffic, geographical traffic distribution, and granularity of traffic demands), and SLA requirements (including physical impairments, Differentiated QoP, EON/BVT, network virtualization and data pre-emption). All these factors need to be considered as they influence not only the design and the choice of resiliency mechanisms, but also the energy-efficiency performance.

New metrics need to be used for energy-efficiency resiliency evaluation by either considering the energy-efficiency with a protection factor or the utilized optical spectrum. From standardisation point of view, existing protocols need to be extended for energy-efficient resilient optical networks using sleep mode devices.

5 Acknowledgments

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n. 257740 (Network of Excellence “TREND”).

7
6 References