ADONIS project
Algorithms for Dynamic Optical Networks based on Internet Solutions

Technical Report

Version: December 2003

Telecommunications Networks Group at Politecnico di Milano,
Telecommunications Networks Group at Politecnico di Torino,
Computer Networks Group at Università di Trento

Contents

Chapter 1

Introduction

Objectives of the research project

ADONIS (Algorithms for Dynamic Optical Networks based on Internet Solutions) is a FIRB research project, sponsored by the Italian Ministry for University and Scientific Research, aiming at proposing novel methodologies for the design and management of optical networks in order to support the migration from the current static assignment and routing techniques to adaptive, dynamic techniques based on IP-centric control.

According to the original project proposal, the research activities are organized into two workpackages:

WP1 *Intelligent static network design*, whose main objective is to suggest a unified evolutionary design approach, which supports the current static network structure, but is aware of the dynamic future requirements. By a proper design of the static network, these techniques allow a seamless migration to dynamic management as soon as equipment and protocols become available.

WP2 *Evolutionary design of dynamic networks*, which aims to develop efficient algorithms and protocols for establishing lightpaths in wavelength-routed networks with dynamic traffic demands. The proposed protocols will consider traffic engineering capabilities and guarantee effective protection/restoration mechanisms. Furthermore, new switching architectures are proposed and evaluated to support the dynamic traffic patterns.

In the following section, an overview of the activities developed during the first year of the project is given. The activities pursued by each Research Unit participating in the project are described in greater detail in the internal chapters of the Technical Report.

Research activities developed during the first year

The activities developed during the first year of the project are divided in the following according to the specific workpackage to which they refer.

Activities in WP1 - Intelligent static network design

A novel approach to WDM static network planning which introduces “availability” as optimization objective function has been proposed in the framework of the first workpackage. The introduction of availability as objective function during this phase is a new research target (see references [1, 2] in Chapter 2). The planning tool is based on a heuristic approach which lead to designing large networks by taking into account both limitations of state-of-the-art technology and strict reliability requirements of current networks. As stated in Sec. 2.1, reliability is measured by the connection availability, i.e. the fraction of mean failure-repair cycle in which the connection is active and working properly. The availability of an optical circuit depends on the availability of the network elements (WDM links and nodes) it crosses. The operator can choose to spend more to equip his network with high-quality hardware, while a more feasible strategy is the adoption of a protection mechanism for the connections. This requires extra transmission capacity to support the backup circuits. The availability of each connection strictly depends on the Routing, Fiber and Wavelength Assignment (RFWA) of lightpaths that have been setup to support it. The RFWA for the link-disjoint w/p pair that minimizes the non-linear function can be found by non-liner programming, with a high computational complexity. Here instead a heuristic method is proposed in Sec. 2.2.1 which is based on a graph in which all the arcs representing idle wavelength channels are labeled with a weight equal to their unavailability. The new tool exploiting such a procedure allows to investigate the trade-off between cost and availability, allowing to fine-tune the availability target. A first planning solution has been found with a realistic case-study network which minimizes the cost while guaranteeing the minimum possible unavailability for all the connections; a second less expensive solution has been given by relaxing this constraint (see Sec. 2.3.1 and authors' work [8] in Chapter 2 for further details).

Activities in WP2 - Evolutionary design of dynamic networks

A new optical node architectures of the switching core for an IP over WDM switching fabric has been proposed. Here the long-term view of a full packet switching network performing IP packet transport is addressed, in which optical operations are performed as much as possible exploiting the currently available optical device technology. This activity deals with the architecture of an optical packet switching node first proposed in [7] (see Chapter 2), which is equipped with a fiber delay line stage used as an input buffer for optical packets. Some new AWG-based alternative structures of the switching core of the node, which exploit the routing in the wavelength domain inherently available in the switching components being used, are described in Sec. 2.2.3. Two different architectural solutions have been compared in terms of complexity and traffic performance both with or without recirculation delay lines to implement to obtain *shared buffering* (open vs. closed switching architectures). The promising result described in Sec. 2.3.3 is that, under reasonable assumptions on the offered IP traffic, the simplest of the new proposed structures outperforms the original one (see ref. [15, 16] in Chapter 2 for further details).

Another activity in the context of WP2 led to the proposal of two new Traffic Engineering (TE) schemes. The first is based on a load balancing mechanism for MPLS networks which optimizes the allocation of LSPs arriving dynamically in order to minimize the blocking probability. The proposed algorithms are based on the idea to efficiently re-route LSPs from the most congested links in the network, in order to balance the overall links load and to allow a better use of the network resources. Results reported in 4.3.1 (and [5] in Chapter 4) shows that the proposed algorithms performed better than well-known algorithms for load-balancing in MPLS networks such as Minimum Interference Routing Algorithm (MIRA), for both symetric and asymmetric dynamic traffic. The second is a novel on-line scheme to route sub-wavelength requests with QoS requirements in a G-MPLS based optical network. Compared to previously proposed TE schemes, its objective is to minimize the rejection probability for high-priority traffic by respecting specific constraints on the maximum tolerable end-to-end delay and packet-loss ratio at the same time. The proposed scheme, described in Sec. 4.2.1, consists of an on-line dynamic grooming scheme which routes an incoming request by respecting specific QoS requirements, while a preemption algorithm guarantees that high-priority requests experience a reduced blocking probability when compared to low-priority ones. Simulations performed on different topologies show that the proposed "local" preemption mechanism minimizes the network disruption both in term of number of preempted connections and new set-up lightpaths. The reader is referred to Sec. 4.3.1 and [10] in Chapter 4 for further details on these results.

Joint activities in both WP1 and WP2

The activities described in this section are related to both workpackages WP1 and WP2.

Among them, one is related to the analysis of the behavior of a wavelength routed network optimized for static traffic but employed to provide on-demand lambda-connection service. The idea here was to investigate whether the set of resources left idle after an optimization procedure of the network for a particular static traffic can be used to accommodate an expansion with dynamic traffic, without disrupting the static lightpaths. In practical terms, this happens when each original customer of the network operator wants to add more optical connections to his already established set in order to increase the bandwidth of his traffic relations. The operator has to satisfy the new requests exploiting the current network idle capacity, without adding extra physical resources to his existing infrastructure. The study on traffic scaling comprises two different steps, as described in Sec. 2.2.2. First, a WDM network is optimized for a given static traffic (with dedicated 1+1 path protection), setting up all the static optical connections initially requested and minimizing the total number of fibers of the network. Secondly, traffic growth is simulated by setting up new lightpaths until the idle capacity is exhausted; furthermore, during this step the performance of several well-known dynamic RFWA algorithms is evaluated. Simulations carried out on some case-study networks in various conditions prove that a substantial number of extra-connections can be setup exploiting this idle capacity without adding new capacity to the optimized physical topology. For details on the traffic scaling gains obtained both on USA National-Science-Foundation Network (NSFNET) and the European Optical Network (EON) topologies, the reader is referred to Sec. 2.3.2 and to authors' work [11] in Chapter 2.

The proposal of efficient algorithms for the dynamic logical topologies design over optical networks has been developed based on the adoption of an Optical Time Division Multiplexing (OTDM) scheme together with classic WDM multiplexing scheme. The joint use of WDM and OTDM technologies (see [1, 2] in Chapter 3) offers the interesting opportunity of splitting the bandwidth of a lightpath into a fixed number of sub-channels, using a Time Division Multiplexing (TDM) scheme directly in the optical domain. According to this proposal, a *fixed bit level framing* is determined, such that each bit in a given position in the frame, called bit slot, identifies a particular sub-channel. Using a bit interleaver, the (single) transmitter multiplexes sub-channels into the frame, and transmits the resulting stream using the same wavelength, obtaining point-to-multi-point channels which is called "*Super-Lightpaths*". An intermediate node receives in the electronic domain a de-multiplexed stream of bits, while it forwards in the optical domain the entire wavelength toward the next destination node. The objective of this activity is to support the on-demand establishment of IP tunnels in dynamically reconfigurable wavelength routed networks. By grooming at the source node several IP-tunnels on the same wavelength and by extending classic RWA algorithms, results in Sec. 3.3.1 (and in the authors' works [4, 5, 6] in Chapter 3) show that the adoption of this joint WDM-OTDM technique can lead either to a reduction of network costs, or to a significant improvement of the network performance. This is obtained with an increase of the aggregate network capacity, which however comes at a limited cost in the optical domain, and additional multiplexing/demultiplexing at network nodes, where complexity is minimal due to the choice of having one source in each Super-Lightpath.

Another activity is related to the development of a new methodology to design the logical topology in single-hop, ring-based physical WDM topologies capable of offering datagram service under dynamic traffic patterns. This activity refers to the design of innovative Metropolitan Area Networks (MANs) architectures to carry datagram traffic. In this context, single-hop optical networks offer a reasonable and cost-effective balance for future MANs. The single-hop architecture employs WDM to provide connectivity among network nodes, exploiting the large bandwidth available on fibers without requiring node transceivers to operate at the full network bandwidth. The authors took as the reference the RingO architecture experimentally studied in the PhotonLab of Politecnico di Torino. In that architecture, the case of tunable transmitters was considered, so that each fixed receiver is permanently assigned to one WDM channel. An important limitation of this kind of architectures is that the number of nodes is often assumed to be equal to the number of wavelengths available in the network. This limitation is overcome by statistically time multiplexing packets to several destinations on the same wavelength channel (that is, the same wavelength can be used to transmit to different nodes). From a network dimensioning perspective, since more than one node can receive on the same wavelength, a decision problem arises concerning the allocation of the different receivers to WDM channels. Good solutions to this problem should aim at equalizing the load on the different channels, i.e., the maximum load among all channels must be minimized (see [14] in Chapter 3 for further details). The study of different allocation policies of transceivers under dynamic datagram traffic reported in Sec. 3.3.2 shows that greedy allocations may have great impact on network performance, while efficient allocation algorithms help to find suitable solutions based on actual traffic.

Finally, a new control protocol for lightpath establishment called Source and Destination Cooperative Reservation with Conditional One-way (SDCRCOW) has been proposed, which can be used in both static and fairly dynamic wavelength routed networks. The proposed distributed control protocol is based on the combination of (i) SDCR, a hybrid scheme of Source Initiation Reservation (SIR) and Destination Initiation Reservation (DIR) which combines the advantages of both these well-known reservation mechanisms and (ii) COW, an efficient data transmission scheme. As shown in Sec. 4.3.2, the proposed protocol performs very well compared to classic Link-state and Distributed-routing approaches in term of blocking probability, lightpath setup time and network utilization (see ref. [13] in Chapter 4 for further details).

Chapter 2

Politecnico di Milano Research Unit

A. Pattavina, M. Tornatore, S. Bregni
Dipartimento di Elettronica – Politecnico di Milano
Via Ponzio, 34/5- 20133 Milano, Italy
{pattavin,tornator,bregni}@elet.polimi.it

Abstract

The tasks of the unit of Politecnico di Milano in this project are focused mainly on the following topics:

- development of methodologies for designing optical networks under static traffic demand and with reliability requirements;
- development of methodologies for the evolution of optical networks able to provide transmission capacity under dynamic traffic demand;
- development of methodologies for designing highly dynamic optical networks and study of switching architectures for IP traffic.

According to the previous objectives our works have been carried on in order to obtain significant results in these three areas.

First of all we have developed a heuristic procedure to optimize WDM networks with reliability as objective function. Optimization experiments have been carried out on a case study network, exploring the trade-off between network cost and reliability.

On the other hand, in these last years both the corporate and research community agree in stating the key-role that dynamic traffic could play in the future IP over WDM scenario. So the second part of our work investigates the behavior of a WDM network optimized in case of an unpredictable increase of the number of connection requests. New lightpaths are setup as dynamic traffic exploiting idle capacity left in the network after static optimization. Traffic is grown until a given blocking probability value is reached so that the maximum allowable traffic scaling factor can be evaluated.

Actually, in order to support the continuous growth of transmission capacity demand, the optical burst switching (OBS) technology has been proposed, that is able to allow fast dynamic allocation of WDM channels, combined with a high degree of statistical resource sharing. So we propose some innovative optical switch architectures for OBS applications, that use an arrayed waveguide grating (AWG) device to route packets, which better exploit the switching capability of this component in the wavelength domain. This analysis is also extended considering an architecture exploiting input/shared-buffered optical switching node based on recirculation delay lines. Different node configurations are compared in terms of complexity, and traffic performance of these new structures in a full optical packet switching scenario is also examined.

2.1 Introduction

In recent years the importance of WDM networks has rapidly grown, leading them to become the core of the telecommunication infrastructure, able to face the increase of data and video traffic. If WDM technology can offer a solution to the large bandwidth demand, WDM protocols (management and control systems) have been developed in order to guarantee that the WDM layer will act in the future as a common transport platform able to operate in an integrated multi-protocol environment, providing a high quality of service.

ADONIS project aims at proposing an evolutionary methodology for the design and management of optical WDM networks in order to support the migration from the current static assignment and routing techniques to adaptive, dynamic techniques based on IP-centric control as soon as new generation equipment becomes available.

The activities described in this report have been carried on by the Research Unit at Politecnico di Milano within the FIRB ADONIS project.

The first step of our activity consists in introducing a quality of service parameter during the static optimization phase of WDM layer. Contracts between operators and their customers are almost always made on the basis of service-level agreements which are very strict on optical circuits reliability, an important quality-of-service (QoS) aspect. So, as far as the first topic is concerned (workpackage 1), we aim at developing methodologies for dimensioning and optimizing optical networks able to provide lightpaths under some assigned traffic demand matrix. taking into account the degree of quality of service provided by each optical connection, expressed in terms of availability. The introduction of availability as objective function during this phase is a rather new research target [1, 2], important for the future. In our work, to our knowledge one of the first on this topic, we develop a network planning tool able to optimally plan a WDM network when availability is a key parameter. A heuristic approach will be adopted, in order to face the issue of dealing with large networks. The project will be carried out by taking into account both limitations of state-of-the-art technology and strict reliability requirements of nowadays networks. Reliability is measured by the connection availability, i.e. the fraction of mean failure-repair cycle in which the connection is active and working properly. The ability to guarantee high availability is fundamental for the operator in order to attract customers: it comes however at extra network costs. The availability of an optical circuit depends on the availability of the network elements (WDM links and nodes) it crosses. The operator can choose to spend more to equip his network with high-quality hardware. However, improving the quality of these elements above the levels of present commercial systems is hard and usually beyond the control of the operator. A more feasible strategy is the adoption of a protection mechanism for the connections. This requires extra transmission capacity to support the backup circuits. The availability of each connection strictly depends on the Routing, Fiber and Wavelength Assignment (RFWA) of lightpaths that have been setup to support it. Thus, careful network planning is in any case fundamental to fully exploit the benefits of good-quality hardware and protection strategy.

Up to the present time WDM networks have mainly been regarded as static systems, both in planning and management. One of the main issues of research on optical networks, developed under the drive of the operators' needs, has been network optimization: a vast number of procedures and algorithms have been proposed in literature to evaluate the lowest-cost capacity dimensioning given a certain static set of connection requests. The "static paradigm" was justified by the fact that optical connections has been mostly used as long-distance backbones to carry large aggregates of telephone traffic between distant locations, mainly involving end-customers of a single operator, thus with highly predictable traffic volume.

We are witnessing a new evolution of WDM networks today, as a consequence of a big change in the application scenario. Data traffic is going to overcome traditional telephone traffic in volume: statistical modeling of network load has to be modified to describe a new reality with less regular flows, more and more independent from geographical distances. Moreover, WDM has been successfully adopted in regional- and metro-area networks, in which flows are less aggregated and more sensitive to traffic relations due to single, large bandwidth applications. Finally, many operators are beginning to offer the "lambda service" (i.e. optical connections for lease) and the carrier-of-carries service to support the so-called "bandwidth-trading" business.

The static conception of the network is showing its limitations since it is no longer able to cope with these new applications. The focus of research is thus moving from static planning to lightpath on-demand provisioning. The change is also reflected by the evolution of WDM protocol standardization. The simple static Optical Transport Network (OTN) is already well-defined by the main standard bodies [3, 4], while the new model known as Automatic Switched Optical Network (ASON) is currently under development. Its main feature is the ability to accommodate on-line connection requests issued to the network operating system, which is responsible of the activation of new lightpaths in real time. Despite current economy difficulties may impose a slower evolution pace, this change of paradigm from static to dynamic system does not seem to be reversible.

It is however likely that the evolution from OTN to ASON is going to happen as a gradual process in order to preserve the investments of the network operators. While upgrading network control and management systems to ASON will be a relatively easy and quick task, the installed transmission systems and the existing capacity of an operator will probably remain unchanged for a certain period. In this transition phase static and dynamic traffic will co-exist and share the same WDM network infrastructure. Mostly likely, this infrastructure has been designed and optimized in order to support a given original static traffic, according to the OTN design approach. So an interesting issue is to evaluate how many on-demand lightpaths can be provided in an optimized WDM network without disrupting or reconfiguring the original static connections. So, the second topic concerns both workpackages and addresses new methodologies for designing optical networks under dynamic traffic demand. This implies that not only dimensioning the network, but also its management must be based on the statistical description of incoming bandwidth requests. Therefore, routing and wavelength assignment (RWA) algorithms will aim at minimizing both the resources allocated to accepted traffic and the probability of rejecting new connection requests.

In our activity we have more precisely identified the particular situation in which the new on-line connection requests are generated as an expansion of the original static traffic. In practical terms, this happens when each original customer of the network operator wants to add more optical connections to his already established set in order to increase the bandwidth of his traffic relations. The operator has to satisfy the new requests exploiting the current network idle capacity, without adding extra physical resources to his existing infrastructure. It should be noted that we chose this particular scenario to perform a sort of worst-case analysis: obviously, a wise operator would probably have over-dimensioned the network during the design phase pre-adding some extra capacity to the amount prescribed by an optimization procedure. However our purpose here is to ascertain whether the set of resources left idle after an optimization procedure is suitable to be used to accommodate a traffic expansion, so no over-dimensioning is considered.

Moreover, recently considerable research is currently devoted to the design of IP fully-optical backbone networks, which will provide the possibility of overcoming the capacity bottleneck of classical electronic-switched networks.

Unfortunately, although WDM networks have evolved to support some network functions as circuit routing and wavelength conversion and assignment, optical devices used in market equipment are not mature enough to meet packet-by-packet operation requirements yet. An interesting solution which tries to represent a balance between circuit switching low hardware complexity and packet switching efficient bandwidth utilization is the optical burst switching ([5] and [6]). In an optical burst switching system, the basic units of data transmitted are bursts, made up of multiple packets, which are sent after control packets, carrying routing information, whose task is to reserve electronically the necessary resources on the intermediate nodes of the transport network.

Such operation results in a lower average processing and synchronization overhead than optical packet switching, since packet-by-packet operation is not required. However packet switching has a higher degree of statistical resource sharing, which leads to a more efficient bandwidth utilization in a bursty, IP-like, traffic environment.

We address in the third point of these activities the long-term view of a full packet switching network performing IP packet transport, in which optical operations are performed as much as possible exploiting the currently available optical device technology. Apparently most of the operations related to the packet header processing needs to be done in the electronic domain. This last activity deals with the architecture of an optical packet switching node first proposed in [7], which is equipped with a fiber delay line stage used as an input buffer for optical packets. Some new alternative structures of the switching core of the node, which exploit the routing in the wavelength domain inherently available in the switching components being used, are described. In this third field of optical networks research we could identify a twofold approach related to the presence or absence of shared input buffers based on recirculation delay lines (open vs. closed switching architectures).

Concluding, the third topic concerns workpackage 2 and considers traffic that varies in a highly dynamic fashion and with fine granularity. For instance, let us consider requests for lightpaths lasting very short intervals, such those of typical packet burst durations. Since it is very likely that future optical networks will transport directly IP traffic, it is evident how important it is to characterize thoroughly in statistical terms the IP traffic, both aggregated and from single sources, to dimension the optical network. To this aim, grooming techniques play a key role. Such techniques aggregate IP flows in order to made up packet bursts that have sufficient length to minimize switching time compared to burst duration. This study cannot be disjoint from the evaluation of the availability of new switching systems able to switch IP bursts with an acceptable degree of efficiency, by possibly supporting different per-flow quality-of service.

Architectural and performance characteristics of these switches have a significant impact on the project methodologies of the overall optical network.

2.2 Research unit activities

Let's have a deeper look on the activities that have been introduced in the previous section. We can clearly distinguish three main research tracks, respectively (1) static planning with availability constraints, (2) dynamic traffic scaling and (3) architecture and performance of optical switching nodes for IP-networks.

2.2.1 Static planning with availability constraints

We have developed a network planning tool able to optimally plan a WDM network when availability is a key parameter [8]. The availability parameters of the network elements are given as an input, as well as the static traffic matrix (the set of all the requested optical point-to-point connections) and the physical topology. The number of wavelengths per fiber W is preassigned, while the number of fibers per each link is a variable of the problem. All the connections are protected by Dedicated Path Protection (DPP), i.e. two link-disjoint lightpaths (working + protection, or a w/p pair) are setup per connection. A heuristic algorithm performs RFWA of all the w/p pairs, simultaneously dimensioning the necessary transmission capacity of the links (in terms of number of fibers) according to a specific objective function (e.g. connection availability or network cost).

A heuristic has been preferred to an exact approach due to its lower computational complexity, which allows to plan also large networks in a reasonable time. The tool is based on a deterministic heuristic procedure described in details in [9]. Here we would like to present only the procedure added specifically to support availability optimization, which finds the RFWA for the w/p pair of a connection that has the maximum possible availability under link-disjoint constraint, given source and destination nodes and graph of the idle wavelength channels.

An availability model of a connection adopting DPP is a parallel of two series, representing the working and the protection lightpath, respectively (figure 2.1). Each series models the sequence of WDM channels assigned to the lightpath from the source to the destination node. Each element of the series accounts for the total availability of the sequence of devices crossed by the lightpath form one node to the next: a booster Optical Amplifier (OA), a pre-OA, a O/E/O transponder (a lightpath crosses one transponder in every node with the possibility of wavelength conversion - VWP network) and line OAs (equally spaced along the link), their number depending on the link length. In this first study switching nodes has been assumed ideal, i.e. always available. According to well known equations [10], the

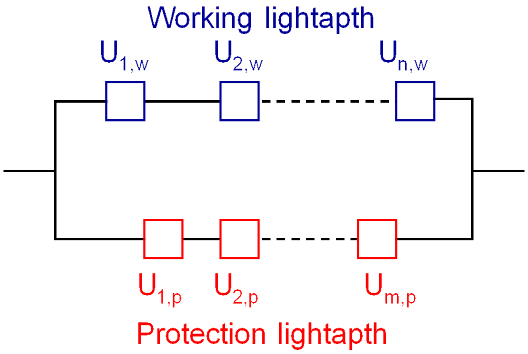

Figure 2.1: DPP connection availability model

availability A of a connection is given by: $A = 1 - U_w \cdot U_p$, where $U_w = 1 - \prod_{i=1}^{n} A_{iw} \simeq \sum_{i=1}^{n} U_{iw}$, and $U_p = 1 - \prod_{i=1}^{n} A_{ip} \simeq \sum_{i=1}^{n} U_{ip}$ is the unavailability of the i-th element of the working (protection) lightpath. The RFWA for the link-disjoint w/p pair that minimizes the non-linear function could be found by non-liner programming, with a high computational complexity. We propose instead a heuristic method based on a graph in which all the arcs representing idle wavelength channels are labeled with a weight equal to their unavailability. Two known algorithms can be applied: the ”One-step” (or Bandhari's), finds the link-disjoint pair of paths having the minimum total weight $U_w + U_p$; the ”Two-step”, finds out the least-unavailability path (working) and the second link-disjoint least-unavailability path (protection). None of the two algorithms actually minimizes $U_w \cdot U_p$, but the sub-optimal solutions they find are expected to be very close to the absolute optimum. Our method applies both the algorithms for each connection and keeps the one solution found which gives the lowest unavailability. It can be proved that when the two solutions are identical, they also coincide with the actual optimum: we verified that this happened for all the connections of the case-study we solved. We conjecture that the heuristic approach gives results sensibly far from the optimum only in extremely connected networks.

2.2.2 Dynamic traffic scaling

Our study [11] comprises two different steps. First, a WDM network is optimized for a given static traffic, setting up all the static optical connections initially requested and minimizing the total number of fibers of the network. Secondly, traffic growth is simulated by setting up new lightpaths until the idle capacity is exhausted, as we are going to describe later on in section.

Static optimization

The optimization is carried out by exploiting a tool that we developed and which is reported in details in a previously published work (Ref. [9]). Let us briefly describe the tool operation.

The set of requests for static connections (virtual topology) is fed to the design tool, together with a description of the physical characteristics of the network (topology, wavelength conversion, etc.). The tool gives the possibility to select a WDM protection strategy for the optical connections. In this work we have chosen the Dedicated Path-Protection (DPP).

Once the protection technique is selected, the design tool proceeds in evaluating Routing, Fiber and Wavelength Assignment (RFWA) for all the w/p pairs or for all the lightpaths, in order to satisfy all the static connections of the virtual topology. In doing this, it also determines the number of fibers that must be installed in each WDM link of the network, thus solving the dimensioning problem of the physical topology. This is done by exploiting a heuristic optimization cycle which assumes the total number of fibers installed in the network as cost function. The heuristic technique we have defined leads to very good suboptimal results, i.e. returning a network design (lightpath configuration and link capacity) very close to the one necessary and sufficient to support the given set of connection requests (the heuristic results are compared to integer linear programming results in Ref. [9]).

Traffic scaling

In the second phase the optimized network with all the static w/p pairs set up is fed to a discrete-event simulator, whose basic event is the provisioning of a connection in the network keeping the current resources occupancy state into account at the arrival of a new request.

It should be noted that for the purpose of this work the time instant of arrival of a new requests does not matter, since all the connections are supposed to be permanent (there are no death events). The only relevant aspect is the sequence of the requests. To simulate an homogeneous growth of static traffic we have adopted the following request-generation procedure. A couple of source and destination nodes is randomly and uniformly chosen among all the couple of nodes having a static traffic relation. One new connection is requested for that couple. If available resources are not sufficient to satisfy the request, then it is blocked and lost forever [12, 13]. If instead resources are sufficient, the connection is setup by allocating the suitable sequence of WDM channels for an indefinite time. Then, independently of the result of the previous allocation, a new couple is chosen and another request is issued, and so on.

We considered two different and alternative cases in which the connections requested in the traffic-growth phase are either all unprotected or all protected by 1+1 DPP. In the first case a single lightpath is sufficient to satisfy the request, while in the second a w/p pair has to be setup.

All the requests for new optical connections are managed as dynamic traffic. At each arrival the control system applies a heuristic RFWA algorithm trying to setup the corresponding lightpath (or w/p pair) using the available WDM channels. The network resources available to support a new connection are the WDM channels still unassigned at the request arrival, i.e the residual capacity unassigned after the optimization phase.

The chance of being able to accept a new connection is measured by the blocking probability parameter P. At a given simulation event, P is defined as the ratio between the number of unsuccessful events (requests which could not have been satisfied) and the total number of events occurred so far. At the beginning of the traffic-growth phase a threshold value of P is given to the simulator. The simulation is stopped when P reaches the pre-fixed threshold value of blocking probability. For example, a threshold value $P = 0$ implies that the simulation is stopped at the first connection refusal.

At the end of the simulation the scaling factor parameter is measured and returned as output. The traffic scaling factor is defined as the ratio between the number of extra connection requests accepted during the traffic-growth phase and the total number of static connections in the network.

Heuristic Routing algorithms

One particular aim of this work is to compare the effectiveness of different RFWA algorithms in terms of maximum scaling factor obtainable in the traffic-growth phase, given a threshold on the average blocking probability.

In our work we adopted a heuristic method which jointly solves routing, fiber and wavelength assignment without imposing constraints on the viable routes (unconstrained routing [14]). This method is basically the same employed in our static-network design tool [9] and adapted to the dynamic network environment. It is based on the *multifiber layered graph* (MLG) representation of the network. Each WDM channel of the network is represented by an arc of the MLG, as well as each traffic-generating node has an image node in the MLG. In order to jointly solve RFWA, all the arcs of the MLG are assigned proper weights prior to setup the new connection.

The key point of our heuristic approach is weight assignment to the MLG arcs. First of all an infinite (actually, very large) weight is given to arcs corresponding to unavailable WDM channels. Secondly the MLG weight system is used to support all the optimality heuristic criteria adopted to solve RFWA. In our approach specific criteria for routing, for fiber assignment and for wavelength assignment can be combined together with a given priority order. To do this each MLG arc is actually assigned an array of weights instead of a single scalar weight. Each weight of the array is determined by a specific criterion. Each time several alternative MLG routes have an equal total weight according to the primary criterion, they are compared according to the secondary criterion, and so on.

In all the simulations we have always assigned the highest RFWA priority to routing; fiber and wavelength assignment follow in order of decreasing priority. For fiber and wavelength assignment we used two traditional criteria: Most Used (MU) (this algorithm attempts to route the connection requests on the most utilized wavelength first i.e., wavelength are searched in descending order of utilization, in order to maximize the utilization of available wavelengths), and First Fit (FF) (the search order is fixed *a priori*, e.g., λ_0, λ_1, . . ., λ_{W-1}).

As far as the routing strategies are concerned, the two best-known and simplest routing algorithms for dynamic traffic in a WDM network are the Shortest Path Routing (SPR) and the Least Loaded Routing (LLR).

The first one routes the lightpath on the minimum distance available path between source and destination: distance is evaluated as the number of hops (i.e. WDM links) crossed by the lightpath. It is very easily implemented by setting to 1 the weights of all the available WDM channels. This routing algorithm is *static* since the corresponding weights do not depend on the state of the network.

LLR tries instead to route the new lightpath on a path which carries the lowest possible amount of traffic generated by already active connections at the time of connection set up. The algorithm allocates the new lightpath on the route having the least possible *route congestion parameter*. This latter variable is equal to the maximum link congestion parameter among all the links crossed by the route itself.

It should be noted that each of the above algorithms is effective in reducing blocking probability of dynamic connections on a single different front. SPR tends to minimize the amount of resources that a new connection is going to subtract from the pool of available WDM channels of the network. LLR tends to uniformly distribute the load over the links of network. A very interesting option offered by our network model is that more criteria concerning the same aspect of RFWA can be applied in sequence, taking advantage of the best heuristic quality of each one. We applied this to routing, creating a new algorithm from the combination of LLR and SPR in a prioritized sequence, named LLR/SPR. The highest priority is given to LLR; when two routes are equal according to the least-loaded criterion, the shortest one is selected according to SPR. We expect that cascading LLR and SPR can improve blocking probability compared to each single algorithm.

2.2.3 Architecture and performance of optical switching nodes for IP-networks

The architecture of the optical transport network we propose consists of $M = 2^m$ *optical packet-switching nodes*, each denoted by an optical address made of $m = \log_2 M$ bits, which are linked together in a mesh-like topology. A number of *edge systems* (ES) interfaces the optical transport network with IP legacy (electronic) networks (see figure 2.2).

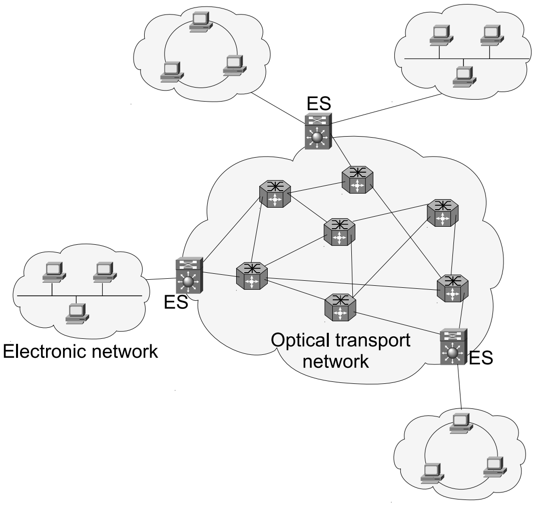

Figure 2.2: The optical transport network architecture

The transport network operation is *asynchronous*; that is, packets can be received by nodes at any instant, with no time alignment. The internal operation of the optical nodes, on the other hand, is *synchronous* or *slotted*, since the behavior of packets in an unslotted node is less regulated and more unpredictable, resulting in a larger contention probability.

An ES receives packets from different electronic networks and performs *optical packets* generation. The optical packet is composed of a simple optical header, which comprises the m-bit destination address, and of an optical payload made of a single IP packet or, alternatively, of an aggregate of IP packets. The optical packets are then buffered and routed through the optical transport network to reach their destination ES, which delivers the traffic it receives to its destination electronic networks. At each intermediate node in the transport network, packet headers are received and electronically processed, in order to provide routing information to the control electronics, which will properly configure the node resources to switch packet payloads directly in the optical domain. In the following we will describe three innovative switching element architectures [15, 16].

Node architecture

The general architecture of a network node consists of three stages: a first stage of channel demultiplexing, a second stage of switching and a third stage of channel multiplexing. The node is fed by N incoming fibers each having W wavelengths. In the first stage the incoming fiber signals are demultiplexed and G wavelengths from each input fiber are fed into each one of the W/G second-stage switching planes, which constitute the switching fabric core. Once signals have been switched in one of the parallel planes, packets can reach every output port through multiplexing carried out in the third stage using any of the G wavelengths that are directed to each output fiber. We note that the number of inlets of each third-stage multiplexer varies, depending on the specific structure of the switching planes. Wavelength conversion must be used for contention resolution, since at most G packets can be concurrently transmitted by each second-stage plane on the same output link.

The detailed structure of one of the W/G parallel switching planes is presented in figure 2.3. It consists of three main blocks: an input *synchronization unit*, as the node is slotted and incoming packets need to be slot-aligned, a *fiber delay lines unit*, used to store packets for contention resolution, and a *switching matrix unit*, adopted to achieve the switching of signals.

These three blocks are all managed by an *electronic control unit* which carries out the following tasks:

- optical packet header recovery and processing;

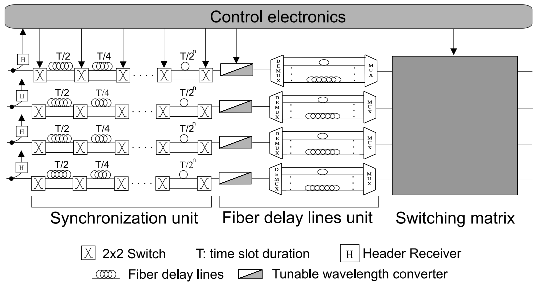

Figure 2.3: Detailed structure of one of the W/G parallel switching planes

- managing the synchronization unit in order to properly set the correct path through the synchronizer for each incoming packet;
- managing the tunable wavelength converters inside the delay unit and in the switching matrix, in order to properly delay and route incoming packets.

One electronic control unit is implemented in each switching plane and, since at each plane output packets are transmitted using one of the G input wavelengths, the controllers' job is carried out in a completely parallel and independent way.

Once packets have crossed the fiber delay lines unit, they enter the switching matrix stage in order to be routed to the desired output port. This is achieved using a set of tunable wavelength converters combined with an arrayed waveguide grating (AWG) wavelength router [17].

The AWG is used as it gives better performance than a normal space switch interconnection network, as far as insertion losses are concerned. This is due to the high insertion losses of all the high-speed all-optical switching fabrics available at the moment, that could be used to build a space switch interconnection network. Commercially available 40 channel devices have a channel spacing of 100 Ghz and show an insertion loss of less than 7.5 dB [18].

Three different structures are proposed for the realization of this stage, referred to as structure (a), (b) and (c). In the following sections we will consider single plane structures, that is $W = G$, in which the switching matrix has NW inlets and NW outlets. The extension to multi-plane nodes is easily achieved for the first two structures by selecting $W' = W/G$. The third structure will require further considerations. Actually, in the switching structures (a) and (b) we will consider all the case in which R AWG ports are reserved to manage recirculation delay lines, which act as a shared buffer. Each of these lines delays a packet by the same amount of time, $D_{rec} = kT$. These recirculation lines can make easier than pure input buffers the provision of different switching service classes by dynamically adjusting the priorities of packets being recirculated.

Structure (a)

The simplest switching matrix structure, first proposed in [7], is shown in figure 2.4. It consists of $2NW$ tunable wavelength converters and an AWG with size $NW \times NW$. Only one packet is routed to each AWG outlet and this packet must finally be converted to one of the wavelengths used in the WDM channel, paying attention to avoid contention with other packets of the same channel. Let's observe that in the structure comprising recirculator each switching plane requires an $(NG + R) \times (NG + R)$ AWG.

Structure (b)

In order to reduce the number of planes of the node and thus to better exploit the *channel* grouping effect (i.e. the sharing of different channels for transmitting a large number of packets, the load per channel being constant) more than one packet can be routed in each AWG inlet. In the resulting switching matrix structure illustrated in figure 2.5, up to k different packets are sent to the same AWG inlet using different wavelengths.

Structure (c)

Node structure (b) can be simplified by selecting $k = N$, so that each AWG input can receive up to N packets using different wavelengths. Therefore, the number of AWG inlets is now exactly W. In this last structure the last TWC stage isn't needed anymore, provided the employed AWG works on the same wavelengths used in the outgoing fibers. In fact, if the electronic controller takes care of avoiding wavelength contention between AWG outlets connected to the same output channel, packets are ready to be transmitted as soon as they exit the AWG. Therefore, a packet entering the AWG inlet i and destined to the output WDM channel j can not be transmitted using every color in the

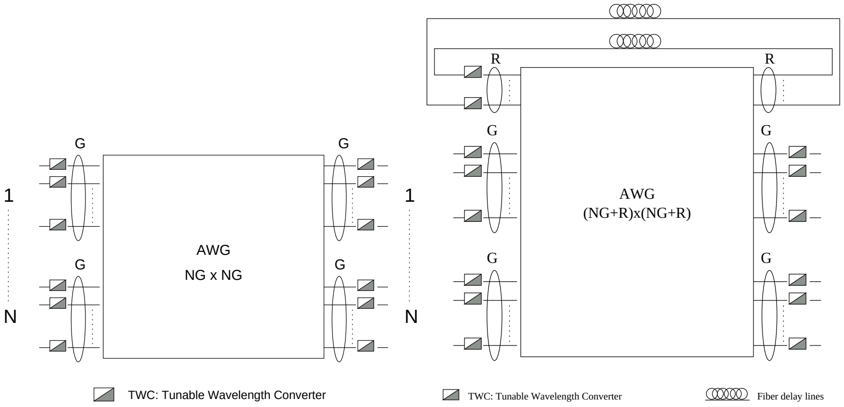
G
G
1
N
AWG
NG x NG
G
G
1
N
TWC: Tunable Wavelength Converter
R
R
G
G
AWG
(NG+R)x(NG+R)
G
G
TWC: Tunable Wavelength Converter
Fiber delay lines

Figure 2.4: Switching matrix with structure (a), with or without recirculation

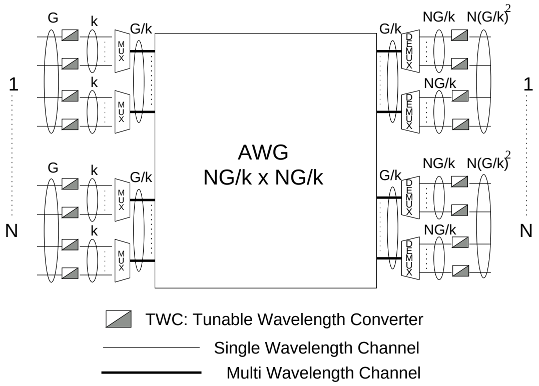

Figure 2.5: Switching matrix with structure (b)

WDM channel, but only using a subset which consist of the W/N wavelengths through which the packet can reach the desired output channel, thus reducing the benefits of *channel grouping* (when N and W are kept constant).

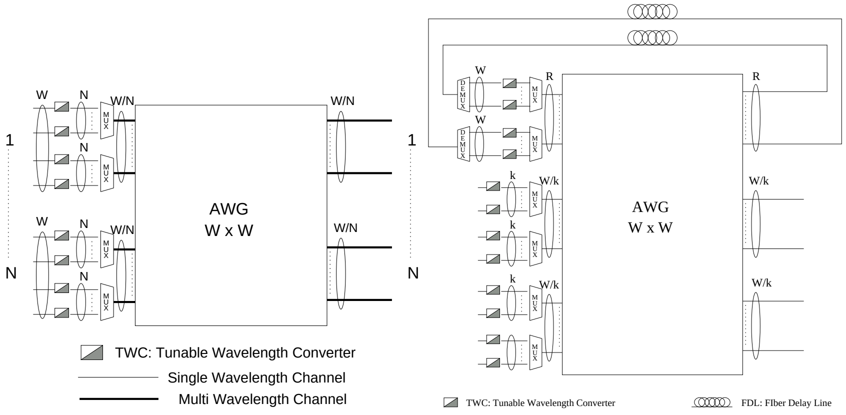
W
N
W/N
MUX
1
N
AWG
W x W
W/N
1
N
TWC: Tunable Wavelength Converter
Single Wavelength Channel
Multi Wavelength Channel
W
DEMUX
R
R
k
W/k
AWG
W x W
W/k
TWC: Tunable Wavelength Converter
FDL: FIber Delay Line

Figure 2.6: Switching matrix with structure (c), with or without recirculation

We would like to point out that the size $W \times W$ of the AWG is not a limiting factor for this node architecture, since the current optical technology enables using optical fibers supporting a number of wavelengths much larger than the maximum size of an AWG. On the other hand if we are willing to fully exploit the external number of wavelengths W in the internal node structure in case of a maximum size $W' \times W'$ of the AWG, with $W' \leq W$, a multi-plane structure must be adopted.

2.3 Discussion of results

The previously introduced studies allows us to obtain a significant amount of results: we will discuss them following the same topic subdivision structure.

2.3.1 Static planning with availability constraints

Let us discuss now the application of our tool to the planning of a case-study network. The physical topology and the matrix of the connection requests (250 requests in total) are based on a realistic model of a hypothetical Italian long distance network. It comprises terrestrial and submarine WDM links, which differ in OA spacing (100 km and 60 km, respectively) and Mean Time Between Failure (MTBF) value. For each lightpath (working and spare) we have also taken into account unavailability of one WDM mux/demux pair and of one electronic receiver. The value of MTBF for all the components have been chosen by estimating real typical values of commercial systems. The Mean Time to Repair (MTTR) for the terrestrial and submarine links is 2 hours and 14 days, respectively.

We developed this study not only as an application experiment of our new approach, but also with the purpose of exploring the trade-off between network availability and cost, obtaining some interesting results we are going to show. We chose a simple but relevant network cost parameter, which is the total number of fibers M that must be installed in the network to support all the demanded (DPP) connections in a given RFWA solution. We repeated planning for three values of W (2, 4 and 8).

The planning tool is used a first time (phase U) setting up a w/p pair for each connection so to obtain the minimum possible unavailability. RFWA is solved by the heuristic procedure described above, while the cost of the network is not taken into account at all. Then the solution found is reprocessed to minimize the cost (by the optimization procedure described in Ref [9]). This is done under the constraints that the unavailability of each connection is kept equal (case mU) or is allowed to rise at most a factor 10 (case rU) compared to the value obtained by phase U. By comparison, the network has been also optimized by minimizing the cost, regardless of the availability (case mM).

Figure 2.7 shows the cumulative distribution of the connection unavailability for the cases mU, rU and mM with $W = 8$. This distribution does not change sensibly with W, so the curve is well representative also for $W = 2$ and $W = 4$. The availability optimization is effective in improving the most unreliable connections, while it leaves the values of the mean and average availability almost unchanged. In fact, while the minimum unavailability is always $5.2 \cdot 10^{-9}$, the maximum unavailability is decreased from $3.85 \cdot 10^{-7}$ in the case mM to $2.24 \cdot 10^{-7}$in the case mU. rU leads to maximum unavailability very close to mU ($2.45 \cdot 10^{-7}$), showing that the relaxed-constraint optimization can be an effective approach to reduce the cost without losing in QoS.

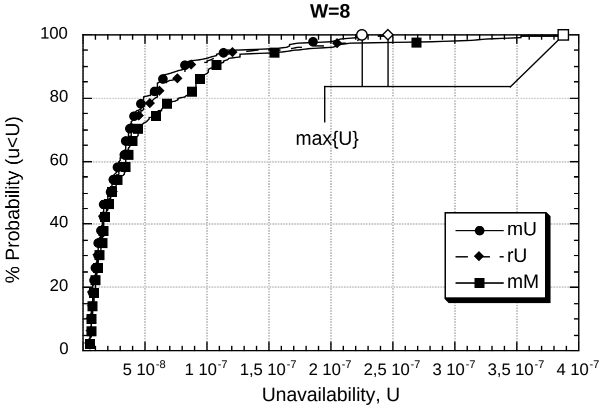

Figure 2.7: Unavailablility cumulative distribution functions

We should note that all the three cases of optimization led to unavailability levels which are well below the figure typically required to an operator by his customers (around 10^{-6}: the five-nines availability). This shows how the DPP technique is effective per-se in improving the QoS.

The resulting cost values are reported in figure 2.8 in term of total number of fibers.

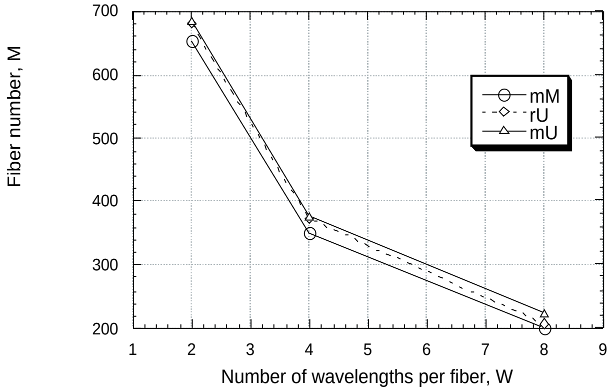

Figure 2.8: Network cost comparison

To better appreciate the differences between the optimization cases Figure 2.9 shows the extra cost due to availability improvement. This cost is measured by the difference in the number of fibers between the unavailability-constrained optimizations (mU and rU) and mM (in absolute and percent terms on the two vertical axes). By comparison, the extra-cost at the end of phase U is also plotted. The cost of the maximum availability can be quite high, even when a cost minimization reprocessing is performed (e.g. 9 fibers = 9 eight-wavelengths transmission systems, comprising all the line devices). The optimization approach with relaxed constraint rU can sensibly reduce the extra cost (e.g. to only 2 fibers) while preserving a good QoS level.

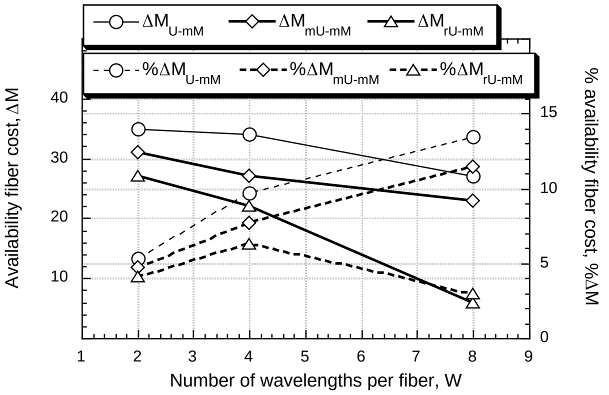

Figure 2.9: Availability extra-cost under different optimization conditions

2.3.2 Dynamic traffic scaling

The final goal of these experiments is the evaluation of the maximum traffic scaling in different conditions.

A first set of simulations was carried out considering two realistic case-study networks, namely the USA National-Science-Foundation Network (NSFNET) and the European Optical Network (EON). Their physical topologies have been derived from data reported in Ref. [19] and Ref. [20] for NSFNET and EON, respectively. A number $W = 32$ of wavelengths per fiber has been chosen to carry out all the experiments of this first set. Moreover, both the two alternative cases with (VWP) and without (WP) wavelength converters have been considered for each network. For sake of brevity in the following we will report only the NSFNET results (for EON case see [11]).

The optimization phase of each of the two networks has been solved with two different static traffic matrices. A first matrix has been defined starting from data based on real traffic measurements and reported in the two papers cited above (Refs. [19, 20]). We name this virtual topology *non-uniform static traffic*, since each node couple has a different number of connection requests.

The second static traffic matrix has been obtained for each of the two networks by evenly distributing the same number of connection requests of the non-uniform virtual topology among all the node-couples. In this way a *uniform static traffic* has been obtained in which all the node couples ask for roughly the same amount of optical connections.

Figures 2.10 represent the result of the static optimization phase carried out on the NSFNET. The labels associated to the links correspond to the number of unidirectional fibers averaged on the two propagation directions, while the link thickness represents the percentage of WDM channels of the link that are left idle after the allocation of all the static connections. In the case of non-uniform traffic (see figure 2.10 (a)) there are parts of the network which are almost disconnected since many links are completely saturated. On the opposite, uniform traffic originates a more uniform distribution of the load on the links. As a consequence, the amount of links that are either extremely underutilized and overloaded decreases (see figure 2.10 (b)).

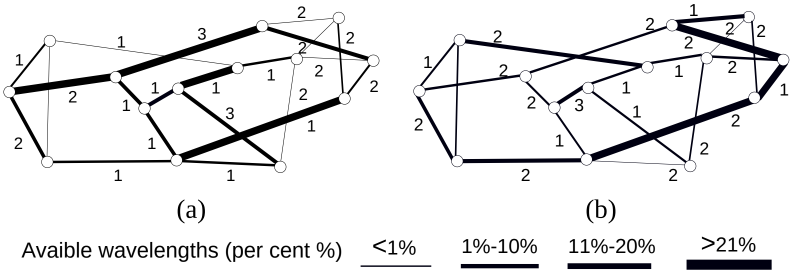

Figure 2.10: NSFNET at the end of the static optimization phase, loaded with (a) non-uniform and (b) uniform static traffic.

A quantitative assessment of the difference between the idle capacity distribution in the two cases of uniform and non-uniform traffic is obtained by evaluating the effective connectivity factor. Given a network with n nodes, $L_f = [n \cdot (n-1)]/2$ is the number of (unidirectional) links of a theoretical fully-connected network having the same number of nodes. Let L be the number of the actual links of the physical topology of the given network. We define the *connectivity factor* as the ratio $\alpha = L/L_f$. At the end of the optimization process, only a subset of the L physical links of the network is not completely saturated by static traffic. Let L_e be the number of links belonging to this subset. We then define the *effective connectivity factor* as the ratio $\alpha_e = L_e/L_f$. In the NSFNET case ($\alpha = 0.242$) the effective connectivity factor with uniform traffic is $\alpha_e = 0.219$, while it decreases to $\alpha_e = 0.154$ in the non-uniform case.

As we are going to show next, the exhaustion of available resources on some particular cut-sets and the saturation of all input and/or output links of a node have a great impact on the possibility to scale the traffic over the optimized network.

The following set of graphs represent the detailed report of all the experiments performed on NSFNET. The maximum scaling factor at the end of the growth phase is always plotted as a function of the preset blocking-probability threshold P. The curves obviously tend to saturate as P approaches 1 ($P = 1$ means that any new connection request is blocked).

A first and a second group of graphs concern the situation in which the traffic scales by setting up new path-protected connections in the NSFNET (figure 2.11). For each network four graphs are reported corresponding to the combination of VWP and WP and uniform and non-uniform traffic. As expected, wavelength conversion and traffic uniformity improve the possibility of scaling. It is however noticeable that traffic distribution is more important than wavelength conversion.

One of the objectives of this study is the evaluation of the impact of the particular RFWA criteria used in the growth phase on the final maximum scaling factor. We can comment on this point by observing the curves of figures 2.11.

As far as routing is concerned, the combined use of LLR and SPR giving to LLR the highest priority (LLR/SPR), is always the best choice, much better than when the two algorithms applied separately. Then LLR has always better performance than SPR. This latter comparison is typical of a dynamic traffic environment, as actually is our growth phase: it is another prove of the importance of avoiding cut-set saturations to lower blocking probability, even at the cost of routing lightpaths on longer routes. The differences between the three routing algorithms are more evident in the uniform traffic case in which the number of routing alternatives for the new connection is higher. In the non-

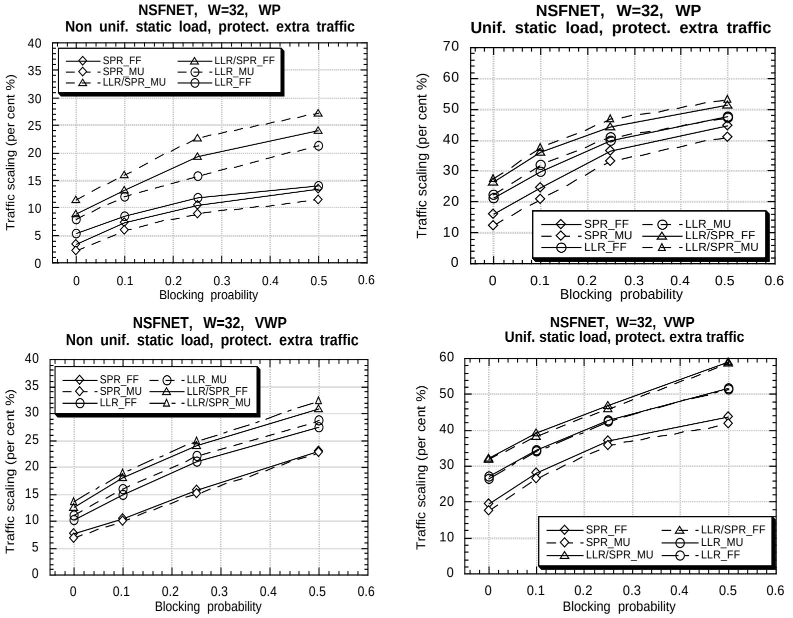

Figure 2.11: Traffic scaling in the NSFNET (WP and VWP case) loaded with non-uniform and uniform static traffic, as a function of the blocking probability threshold P. Dedicated path-protection is adopted for the new connections.

uniform case there are almost no routing choices for many lightpaths since many links are unavailable (see figures 2.11.

Let us concentrate now on the impact of the wavelength and fiber assignment criteria on the scaling factor. The two compared algorithms are the First Fit (FF) and the Most Used (MU), both presented in section 2.2.2.

We first consider the WP network. When the routing algorithm is SPR, FF behaves better than MU. This is due to the fact that FF tends to pack-up short lightpaths on the same wavelengths, leaving wavelengths with high index free to accommodate long monochromatic lightpaths. The MU algorithm behaves "locally" as the FF at each new request, but the priority order of the wavelengths can change from one event to the other, so we can not have the same beneficial effect for long connections. The difference between FF and MU appears more evident when the network is loaded with uniform static traffic compared to the non-uniform case.

When the routing algorithm is LLR or LLR/SPR, then MU is the best performing fiber and wavelength assignment algorithm. This behavior is probably due to the load-balancing effect of LLR. It conveys the idea that an adaptive routing algorithm is better matched by an adaptive wavelength assignment algorithm than by a fixed-order algorithm as FF.

Wavelength conversion does not change the behaviors described above. In the VWP network the differences between FF and MU are less evident than in the WP network (see figures 2.11).

The graphs reported next in figures 2.12 have been obtained by scaling traffic with unprotected connections (only data regarding the WP case have been plotted). The scaling factor is obviously larger than in the protected case (almost the double for low values of P). This gain is partially due to the additional capacity that is required by protection and partially by the link-disjoint constrain. All the other observations on the protected case apply also to the unprotected one. The advantages of LLR-SPR (and LLR) compared to SPR are more evident.

2.3.3 Architecture and performance of optical switching nodes for IP-networks

In order to evaluate the different switching structures, we will compare their traffic performance under a realistic traffic assumption and a complex configuration will be analyzed. For further and theoretically complete information concerning the packet loss probability and the complexity of the structures in terms of number of components let us refer to [15, 16].

So, we show now some traffic performance results given by the different node architecture configurations obtained

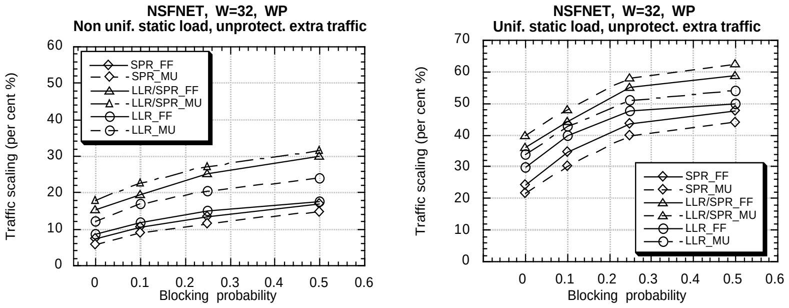

Figure 2.12: Traffic scaling in the NSFNET (WP case) loaded with non-uniform and uniform static traffic, as a function of the blocking probability threshold P. No protection is adopted for the new connections.

through computer simulation. Packet interarrival has been modeled as a Poisson process with negative exponential interarrival times. Based on measurement of real IP traffic [21], the following packet length distribution has been assumed:

$$\begin{cases} p_0 = P(L = 40 \text{ bytes}) = 0.6 \\ p_1 = P(L = 576 \text{ bytes}) = 0.25 \\ p_2 = P(L = 1500 \text{ bytes}) = 0.15 \end{cases}$$

In this traffic model, the resulting average packet length is 393 bytes. Only structures (a) and (c) have been examined now, considering the fact that the implementation of structure (b) brings to an increase of complexity of the electronic controller. On the other hand, in order to implement structure (c), it is only necessary to add a $W \times N$ table to the electronic controller, where each entry contains the W/N wavelengths that can be used to route a packet from inlet i to output channel j. Moreover, this structure reduces the complexity of the controller operations in each time slot, because W/N output wavelengths per packet must be considered, rather than W.

Let us note that the two switching matrices are compared keeping constant the AWG size. Figures 2.13 and 2.14 show a performance comparison for several values of the maximum buffer depth D_{max}. Structure (c) outperforms the original structure (a) and the improvement is as greater as the maximum buffer depth is increased. This improvement is also much bigger as traffic load is decreased.

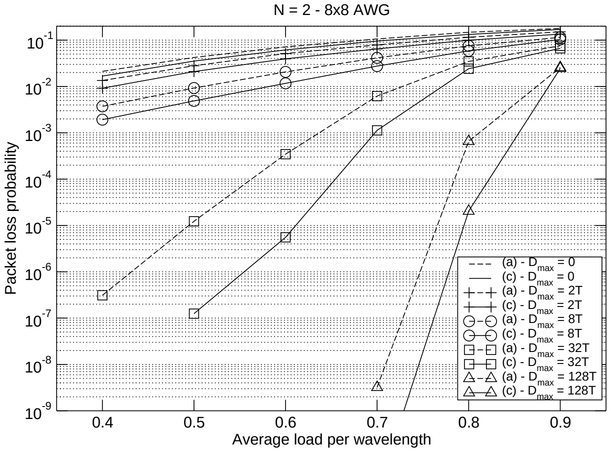

Figure 2.13: Packet loss probability of structures (a) and (c) with $N = 2$ and 8×8 AWGs, for different values of D_{max}.

The number of wavelengths per channel connected to every switching plane is a key parameter to improve node performance, due to the *channel grouping* phenomenon, as was pointed out in [7]. In figure 2.15, different configurations with 32×32 AWGs are compared. It is shown that structure (c), where the value of the G parameter is equal to 32, always gives better results.

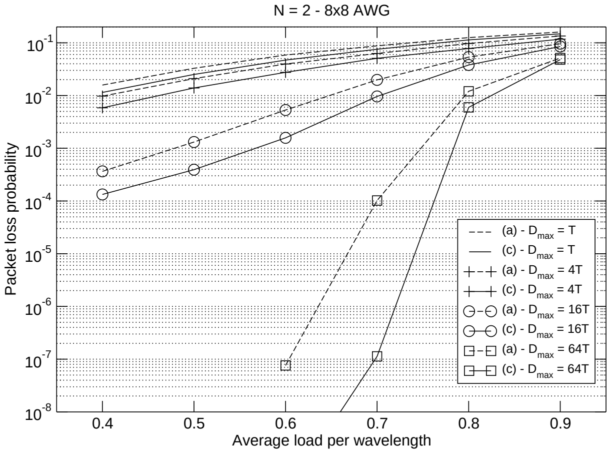

Figure 2.14: Packet loss probability of structures (a) and (c) with $N = 2$ and 8×8 AWGs, for different values of D_{max}.

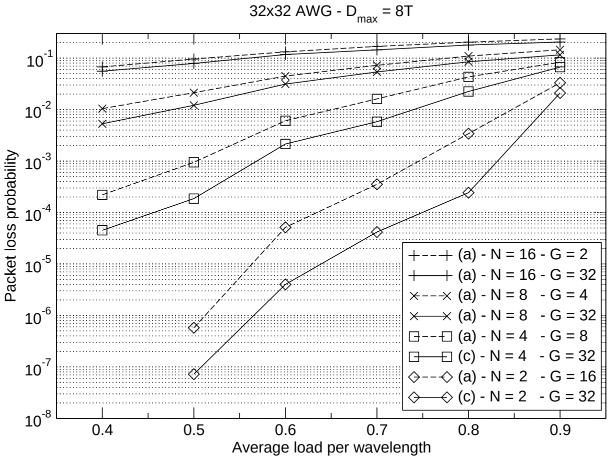

Figure 2.15: Packet loss probability of structures (a) and (c) with 32×32 AWGs, for different values of G.

Finally, in [16] we evaluate how the number of recirculation ports R affects the overall packet loss performance of the node. We first compare for (a) and (c) structures the two configurations with and without recirculation lines. It can be pointed out that the reduction of the *grouping factor* yields a higher loss probability. This performance worsen for lower traffic values, since, as pointed out in [7], the effect of the grouping factor is more apparent for low levels of the offered traffic load. Actually in structure (c), the implementation of recirculation lines always yields to lower loss probability. This is due to the fact that in structures (c), the implementation of recirculation lines brings higher buffering capacity, being able to buffer up to W packets per line, while reducing lightly the node capability to resolve contention in the wavelength domain.

As far as packet delay is concerned, the two structures give similar performance and structure (c) delays are always smaller for low offered traffic and larger for high offered traffic. Moreover, it is clear that, while improving node performance, the implementation of recirculation lines yields higher packet delays.

2.4 Conclusions and future work

We have presented a novel approach to WDM network planning which introduces availability as optimization objective function. The tool that exploits such a procedure is useful to investigate the trade-off between cost and availability, allowing to fine-tune the availability target. A first planning solution have been found with a realistic case-study network which minimizes the cost while guaranteeing the minimum possible unavailability for all the connections; a second less expensive solution has been given by relaxing this constraint.

A deeper analysis on the availability degree provided by the different protection techniques could represent a useful tool to develop a planning tool able to identify the suitable SLA dependable protection strategy. In particular the shared mesh protection, that is an up-to-date proposal and likely subject to real network implementation, deserves an extension of this study.

As far as a dynamic traffic scenario is concerned, we have considered the future scenario of WDM networks designed and optimized for static traffic (with dedicated path protection) and then employed to provide lambda-connection service on demand, without disrupting the static lightpaths. The issue we intended to investigate was whether the set of resources left idle after an optimization procedure of the network for a particular static traffic can be used to accommodate a traffic expansion. Simulations carried out on some case-study networks in various conditions proved that a substantial number of extra-connections can be setup exploiting the idle capacity of the optimized network, even without adding new capacity to the optimized physical topology. We have shown that the maximum scaling factor is very sensitive to the distribution of the initial static traffic. In particular, traffic scalability improves when the static traffic is uniformly distributed.

In this study we have considered the dedicated 1+1 path-protection: this implies that resources allocated to static protection lightpaths can not be used to provision lightpaths in the traffic-growth phase. A different scenario assuming 1:1 DPP for static connections will be considered in future works. In fact, in 1:1 path-protection the signal normally travels on the working path and in absence of failures the spare path is used to carry “low priority” traffic that is lost when a failure occurred. Secondarily the purely dynamic case, with connections characterized by birth and death events, represents a different approach that we are going to analyze in the field of WDM network performance study.

Finally, we have proposed and compared different architectures of the switching core for an IP over WDM switching fabric. Starting from previous proposals of AWG-based optical switching node, it has been shown how to arrange the switch core so as to perform the switching also in the wavelength domain, by thus fully exploiting the AWG properties. Two different architectural solutions have been examined and compared in terms of complexity and traffic performance both with or without recirculation delay lines to implement to obtain *shared buffering*. The results are quite promising in that under reasonable assumptions on the offered IP traffic, the simplest of the new proposed structures outperforms the original one. Other issues will have to be addressed in the future: optical core networks with conflicts for the same outlets solved not only by buffering (input andor shared), but also by deflection routing, in that packets are routed to a downstream node that is not optimal from routing standpoint; switching nodes loaded by different traffic patterns, preferably more realistic than the traffic analyzed so far.

Bibliography

[1] M. Clouqueur and W. D. Grover, “Availability analysis of span-restorable mesh networks,” *IEEE Journal on Selected Areas in Communications*, vol. 20, pp. 810–821, May 2002.

[2] K. C.-H. Chu, M. Mezhoudi, and Y. Hu, “Comprehensive end-to-end reliability assessment of optical network transports,” in *Proceedings, OFC 2002*, Mar. 2002.

[3] *Architecture of Optical Transport Networks*. ITU-T Intern. Telecom. Union - Telecom. Standard. Sector, 1999, no. G.872.

[4] *Network Node Interface for the Optical Transport Network (OTN)*. ITU-T Intern. Telecom. Union - Telecom. Standard. Sector, 2001, no. G.709.

[5] C. Qiao, “Labeled optical burst switching for IP-over-WDM integration,” *IEEE Communications Magazine*, pp. 104–114, Sep 2000.

[6] M. Yoo and C. Qiao, “Optical burst switching for service differentiation in the next generation optical Internet,” *IEEE Communications Magazine*, vol. 32, no. 2, p. 98104, Feb 2001.

[7] S. Bregni, G. Guerra, and A. Pattavina, “Optical packet switching of IP traffic,” in *Proceedings, ONDM 2002*, 2002.

[8] M. Tornatore, G. Maier, A. Pattavina, M. Villa, A. Righetti, R. Clemente, and M. Martinelli, “Availability optimization of static path-protected WDM networks,” in *Proceedings, OFC 2003*, Mar. 2003.

[9] A. Dacomo, S. D. Patre, G. Maier, A. Pattavina, and M. Martinelli, “Design of static resilient WDM mesh-networks with multiple heuristic criteria,” in *Proceedings, INFOCOM 2002 IEEE Conf.*, vol. 3, June 2002, pp. 1793–1802.

[10] E. E. Lewis, *Introduction to Reliability Engeneering*. John Wiley & Sons, 1987.

[11] L. Barbato, G. Maier, and A. Pattavina, “Maximum traffic scaling in WDM networks optimized for an initial static load,” in *Proceedings, ONDM 2003*, Feb 2003.

[12] A. Mokhtar and M. Azizoğlu, “Adaptive Wavelength Routing in All-Optical Networks,” *IEEE/ACM Transaction on Networking*, vol. 6, pp. 197–206, apr 1998.

[13] E. Karasan and E. Ayanoglu, “Effects of Wavelength Routing and Selection Algorithms on Wavelength Conversion Gain in WDM Optical Networks,” *IEEE/ACM Transactions on Networking*, vol. 6, no. 2, pp. 186–196, apr 1998.

[14] L. Li and A. Somani, “Dynamic wavelength routing using congestion and neighborhood information ,” in *Networking, IEEE/ACM Transactions on*, vol. 7(5), Oct. 1999, pp. 779 –786.

[15] S. Bregni, A. Pattavina, and G. Vegetti, “Architectures and performance of AWG-based optical switching nodes for IP networks,” *IEEE Journal on Selected Areas in Communications*, vol. 2, pp. 726–733, 2003.

[16] ——, “Awg-based WDM switching architectures for all-optical ip networks,” in *Proceedings, ONDM 2003*, Feb 2003, pp. 1037–1051.

[17] C. Parker and S. Walker, “Design of arrayed-waveguide gratings using hybrid fourier-fresnel transform techniques,” *IEE Journal on Selected Topics in Quantum Electronics*, vol. 5, p. 13791384, 1999.

[18] “Fiberoptics products: Arrayed waveguide grating,” http://www.samsungelectronics.com/fiberoptics/.

[19] Y. Miyao and H. Saito, “Optimal design and evaluation of survivable WDM transport networks,” *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 1190–1198, sept 1999.

[20] A. Fumagalli, I. Cerutti, M. Tacca, F. Masetti, R. Jagannathan, and S. Alagar, “Survivable networks based on optimal routing and WDM self-heling rings,” *Proceedings, IEEE INFOCOM ’99*, vol. 2, pp. 726–733, 1999.

[21] K. Thompson, G. J. Miller, and R. Wilder, “Wide Area Internet traffic patterns and characteristics,” *IEEE Network Magazine*, pp. 10–23, Nov 1997.

Chapter 3

Politecnico di Torino Research Unit

A. Bianco, V. Curri, J. Finochietto, E. Leonardi, M. Mellia, F. Neri, C. Piglione,
Dipartimento di Elettronica – Politecnico di Torino
Corso Duca degli Abruzzi, 24 – I-10129 Torino, Italy
{lastname}@mail.tlc.polito.it

Abstract

During the first year of the project, the research unit of Politecnico di Torino focused its research efforts in two directions:

- **Dynamic Super-Lightpath in WANs**: we proposed novel algorithms for the design of optical networks jointly exploiting wavelength and time division multiplexing under dynamic connection requests
- **Resource Allocation in MANs**: we developed novel algorithms suitable for the dynamic resource allocation in ring topologies under dynamic packet-based traffic patterns

As regards the first item, belonging to the second workpackage of this project, the effort has been primarily devoted to the study of efficient algorithms for the dynamic logical topologies design over dynamically reconfigurable wavelength routed networks, in which lightpaths carrying IP traffic are on-demand established. In wavelength routed networks, each lightpath is usually established by reserving a wavelength on each fiber belonging to one of the physical paths that connect the end-point routers. We instead exploited the joint use of Wavelength Division Multiplexing (WDM) and Optical Time Division Multiplexing (OTDM) technologies; the latter offers an interesting opportunity of partitioning the bandwidth of each lightpath into a fixed number of sub-channels, using a bit-level Time Division Multiplexing scheme directly in the Optical domain (OTDM). A single transmitter feeds each wavelength, so that no synchronization is required among network nodes. We proposed simple algorithms to allocate users connection requests in a dynamic traffic scenario, and showed that the adoption of a joint WDM-OTDM multiplexing technique leads either to a reduction of the network cost, or to a significantly improvement of the network performance. For what regards instead the second item, belonging to the first and second project workpackages, the research unit developed methodologies to design the logical topology in single-hop, ring-based physical WDM topologies, capable of offering datagram service under dynamic traffic patterns. We considered the case of node interfaces equipped with tunable transmitters and wavelength-fixed receivers, and faced the problem of node allocation to the available WDM channels, which can be viewed as a particular optical network design problem.

3.1 Introduction and motivation

WDM is today a well-established technique to exploit the fiber bandwidth in both core and metro networks, and all major vendors in this field offer a wide range of products and commercial solutions.

When backbone networks are considered, remote high-capacity (electronic) routers are connected through IP-tunnels. IP tunnels are implemented by optical pipes called *lightpaths* that may extend over several physical links. Lightpaths are routed in the optical layer through the physical topology using a wavelength (we do not assume to exploit costly wavelength conversions); at intermediate nodes, incoming wavelengths belonging to in-transit lightpaths are switched to outgoing fibers through an optical cross-connect that does not process electronically in-transit information. At the IP layer, lightpaths are seen as data-link channels through which packets are moved from a router to another router toward their destinations, following the classic IP forwarding procedure. Therefore, in a Wavelength Routed (WR) network, an *IP layer topology* (also called logical topology), whose vertexes are IP routers and whose edges are lightpaths, is overlayed to the *physical topology*, made of optical fibers and optical cross-connects (OXC). If switching of lightpaths is fully performed in the optical domain, the term "transparent" is used.

The switching cost in transparent WR networks can be considered to be essentially insensitive to the transmission data-rate. On the other hand, the complexity and cost of optical switches exhibit a strong dependency on the number of lightpaths to be switched in the nodes, i.e., on the number of input/output ports (or connectivity degree). As a consequence, at the optical layer it is more convenient to increase the lightpaths data-rate, thereby reducing the connectivity degree of the logical IP topology. A lower connectivity degree of the IP topology, however, would entail longer paths on the logical topology, thus resulting in an increase of the bandwidth to be switched at the IP layer. In addition, increasing the lightpath data-rate increases the granularity of the lightpath capacity, thus resulting in possible bandwidth wastes.

Therefore, it is strongly desirable to decouple the data-rate of lightpaths at the optical layer from the data-rate of tunnels at the IP layer, by grooming several lower speed IP-tunnels onto the same higher-capacity optical pipe.

In [1, 2] we considered static WR networks (with semi-permanent lightpaths), and we proposed the joint use of WDM and Optical Time Division Multiplexing (OTDM) technologies, which offers the interesting opportunity of splitting the bandwidth of a lightpath into a fixed number of sub-channels, using a Time Division Multiplexing (TDM) scheme directly in the optical domain. According to our proposal, a *fixed bit level framing* is determined, such that each bit in a given position in the frame, called bit slot, identifies a particular sub-channel. Using a bit interleaver, the (single) transmitter multiplexes sub-channels into the frame, and transmits the resulting stream using the same wavelength, obtaining point-to-multi-point channels (or tree [3]) that we call "*Super-Lightpaths*". An intermediate node receives in the electronic domain a de-multiplexed stream of bits, while it forwards in the optical domain the entire wavelength toward the next destination node. While this approach might not result optimal with respect to the bandwidth utilization when compared with traditional OTDM schemes, it avoids most technological issues related to bit level synchronization.

In the first year of the project we considered dynamically reconfigurable WR networks in which lightpaths are on demand established [4, 5, 6]. We relayed again on the joint use if WDM and OTDM technologies to allow the partition of the lightpath bandwidth into several lower-speed IP tunnels.

While this first activity of the research unit at Politecnico di Torino refers to connection-oriented traffic in backbones, the second activity refers to datagram traffic in Metropolitan Area Networks (MANs). The feature common to the two activities is that they both tackle topology design problems in dynamic traffic scenarios, capitalizing onto the flexibility in the wavelength domain offered by optical networking technologies.

For what regards the second activity, designing innovative MAN architectures means finding cost-effective combinations of optical and electronic technologies, and new networking paradigms that better suit the constraints dictated by available photonic components and subsystems. In this context, single-hop optical networks offer a reasonable and cost-effective balance for future MANs. The single-hop architecture employs WDM to provide connectivity among network nodes, exploiting the large bandwidth available on fibers without requiring node transceivers to operate at the full network bandwidth. Paths between nodes are created by dynamically sharing on a packet-by-packet basis WDM channels, which operate at a data rate compatible with current electronic technology. To find a good balance between optical and electronic complexity, we consider the case in which node transceivers can operate on a single WDM channel at a given time. Tunability at transceivers is required in order to temporally establish an all-optical single-hop connectivity between nodes. However, due to costs of tunability at transceivers, usually media access protocols that require tunability only at one end are more cost effective.

When tunability is present only at one connection end, the other end is permanently assigned to one WDM channel. We took as the reference the RingO architecture experimentally studied in the PhotonLab of Politecnico di Torino [14]. In that architecture, the case of tunable transmitters was considered, so that each fixed receiver is permanently assigned to one WDM channel. Simple slotted medium access control protocols can be used to obtain good global performance while guaranteeing fair access between nodes [13]. However, an important limitation of these architectures is the fact that they often assume that number of nodes must be equal to the number of wavelengths available in the network; in this case, each node is assigned one full wavelength for reception. This largely impairs the scalability and flexibility of the architecture: adding one node would mean re-designing all node interfaces and re-engineering the transmission over the fiber. This limitation is overcome by statistically time multiplexing packets to several destinations on the same wavelength channel (that is, the same wavelength can be used to transmit to different nodes). This is not far from the basic idea exploited in the first activity described above. From a network dimensioning perspective, since more than one node can receive on the same wavelength, a decision problem arises concerning the allocation of the different receivers to WDM channels. Good solutions to this problem should aim at equalizing the load on the different channels [14], i.e., the maximum load among all channels must be minimized.

In the remaining of this chapter, we will briefly summarize the methodologies used to face the two problems above, and present the most important results that were obtained in the first year of the Adonis project.

3.2 Description of the activity

3.2.1 Dynamic Super-Lightpath in WANs

For what regards the study of dynamic resource allocation in WR backbone networks, as mentioned in the previous section, we introduce a new dimension to the problem, overlaying a TDM multiplexing scheme to the WDM channels already offered by the optical layer [1, 2]. In particular, we propose the use of OTDM technologies [7, 8], which have been originally proposed to enable high-speed channels directly in the optical domain. A bit level TDM scheme is created, and D receivers/transmitters can be de/multiplexed to form a D times faster point-to-point optical channel. The multiplexing of D tributaries can be obtained by optical, electronic or mixed technologies [9]. As in [1, 2], we propose to extend the OTDM bit-stream carried by a single wavelength, and generated by a single transmitter, to several destinations nodes, obtaining a Super-Lightpath. Tree shaped Super-Lightpaths become possible if OXCs have wavelength splitting capabilities [3], which is usually the case. Note that a single transmitter feeds each Super-Lightpath, so that technological complexities due to adding information to in-transit synchronous frames are avoided. An intermediate node receives in the electronic domain a de-multiplexed stream of bits, while forwarding in the optical domain the entire wavelength toward the next destination. If the time de-multiplexing is obtained directly in the optical domain, the electronic part of the node receiver (to which most of the bit-rate-dependent costs are normally associated) is only marginally increased.

The introduction of the time dimension modifies the RWA problem, transforming it in a Routing, Time and Wavelength Assignment (RTWA) problem. In the Adonis project, we studied in particular the case of dynamic traffic scenarios, in which end-to-end connections are established and released at different times [4, 5, 6]. Given the similarities between the classic RWA and the novel RTWA problems, most of the algorithms that have been already proposed in the literature can be extended to include the time allocation problem, or novel ad-hoc algorithms may be devised.

The scheme we followed to devise heuristics that solve the RTWA problem was derived from the observation that the accommodation of connections using only the time dimension does not consume precious resources as wavelengths are. Therefore, we will try to allocate the new connection:

1. allocating a free time slot on a Super-Lightpath that already connects the source and the destination node; Super-Lightpaths will be tested following a predefined order or looking for a optimal candidate, and the algorithm will use the first available time slot;

2. otherwise, extending an already present Super-Lightpath rooted in the source node, but not reaching the destination, and performing a time allocation; also in this case the algorithm will select a Super-Lightpath (eventually according to some optimality criterion) that can be successfully extended to accommodate the request;

3. otherwise, creating a new Super-Lightpath rooted at the source node and terminating in the destination node, as done by the classic RWA problem, and then performing a time slot allocation.

Considering a dynamic scenario in which connections have finite durations, we have to define how resources are released when a connection is terminated. In particular, when a connection is torn down, the corresponding bit slot becomes free. When all bit slot on one or more links of a Super-Lightpath are free, the corresponding wavelength will be freed up on those links, resulting in a "shorter" Super-Lightpath, eventually completely releasing the wavelength if no time slots are used on any link of the Super-Lightpath.

Extension of classic heuristics

To solve the RWA problem, we selected two algorithms that were shown to give good performance: the *First-Fit Alternate* (FF-alt) and *First-Fit Least-Congested* (FF-LC) [10]. In more details, when searching for available wavelengths on a given path, a lower numbered wavelength is considered before higher numbered wavelengths, and the first available wavelength is then selected (First Fit Strategy) by both algorithms. As regards the path selection, the FF-alt algorithm considers a pre-ordered list of available paths from source to destination nodes. In case no wavelength is available on the first path, than the second alternate path is tested, followed by the third one, etc. (alternate routing scheme). Paths are sorted according to their hop count, in a static way.

The FF-LC algorithm, instead, dynamically sorts the path list, so that always the least congested is tested first. The "congestion" metric counts the number of wavelengths already used on a fiber, so that the path with the largest number of unused wavelengths is chosen.

Novel heuristics

We propose a novel algorithm to solve the RTWA problem, which tries to exploit the larger number of possibilities available to accommodate a connection when using Super-Lightpaths. In particular, among all possible allocation solutions, we select the one that steals the fewest resources to future connection requests.

We define a metric to estimate the impact of a possible accommodation solution, and then evaluate all of them, selecting the one which exhibits the minimum cost. Indeed, in general the cost of allocating a new connection request comprises two terms: the first one takes into account the cost of consuming a time slot on an already present Super-Lightpath; the second one takes into account the allocation of wavelengths along a new Super-Lightpaths, or along a new portion of Super-Lightpath.

Let $P = \{p\}$ be the set of pre-calculated[1], loop-free paths p and let $P(l) \subseteq P$ be the subset of paths that include fiber l; let $T(s, \lambda)$ be the tree used to route a Super-Lightpath from node s using wavelength λ. Let $\Psi(p, \lambda)$ be equal to 1 if the wavelength λ is free on all the fibers belonging to the path p; $\Psi(p, \lambda)$ equals 0 otherwise.

Denoting with p' the path that should be added to $T(s, \lambda)$ to reach destination node d (both p' and $T(s, \lambda)$ can be null), then the cost $C(s, d)$ to allocate a request from s to d is defined as:

$$C(s, d) = |T(s, \lambda)| + \sum_{l \in p'} \sum_{p \in (P(l) \setminus p')} D\Psi(p, \lambda) \tag{3.1}$$

where the term $|T(s, \lambda)|$ (number of edges in $T(s, \lambda)$) takes into account the fact that a timeslot must be allocated on all links on the selected tree, while the second term considers all the possible (future) connection requests that might be routed using wavelength λ on all fibers of the new path p'. Indeed, allocating λ to the current connection request "steals" D time slots to other possible connections on each link of p'.

For example, if a connection is accommodated by simply allocating a time slot on an already present Super-Lightpath, then $C(s, d) = |T(s, \lambda)|$. On the contrary, $C(s, d) = \sum_{l \in p'} \sum_{p \in (P(l) \setminus p')} D\Psi(p, \lambda)$ when a completely new Super-Lightpath is created.

The algorithm, which is referred as *Maximize Capacity* (MaxCap) in the remaining of the paper, evaluates the cost of *all* possible solutions to the connection allocation problem, and selects the one that has the minimum cost, if it exists. Otherwise, the connection request will be blocked.

Given the complexity of the algorithm, a *Simplified MaxCap* (MaxCap(S)) version is also proposed, which tries first to allocate the connection using only a free time slot on a minimum cost Super-Lightpath which has been already set-up. Only if such a time slot is not available, then all possible solutions to the allocation problem will be considered before blocking the request.

[1] To limit the complexity of the algorithm, P can include only a limited number paths between all source-destination pairs. In the simulation, we limited this number to at most 30 paths for each source-destination pair, selecting the shortest paths in terms of hop number.

3.2.2 Resource Allocation in MANs

For what regards our study of resource allocation in optical packet MANs, we consider the case of an unidirectional WDM ring conveying a certain number of wavelengths. Tunable transceivers are needed to temporally establish an all-optical single-hop connectivity between nodes. We consider the case of tunable transmitters, so that each fixed receiver is permanently assigned to one WDM channel.

We consider the case in which several receivers can be allocated to the same WDM channel. To achieve this, we assume a separation between "downstream" and "upstream" resources: transmissions occur on a set of ("upstream") dedicated WDM channels; after one tour of the ring, the transmitted information is switched to another set of ("downstream") WDM channels, from which receptions occur. The separation can be obtained in space, i.e., by using two separate unidirectional fibers for transmission and reception, or in the wavelength domain, i.e., by using two different wavelength bands for transmission and reception. We typically refer to time separation of upstream and downstream resources (see [14]). A possible alternative to upstream/downstream separation for the accommodation of several receivers in each WDM channel would be to have active switching in the data path, so that each node would selectively drop packets destined to him. Although no space reuse can be achieved under the upstream/downstream separation configuration, the node architecture becomes simpler and cheaper with respect to the case where nodes are capable of switching in the data path and thereby permit wavelength reuse (obtaining better network performance).

When the number of nodes N to be supported is larger than the number of wavelengths W available on each path (i.e., downstream and upstream), if the bandwidth needed on reception by nodes is smaller than the capacity of a wavelength, nodes can be assigned to the same wavelength without exceeded the channel capacity. The allocation of nodes to WDM channels is to be done in a proper way, in order to avoid overloading channels, as well as balancing their loads. The problem of allocation node receivers to WDM channels can be viewed as a particular logical topology design problem.

Our activity has focused on modeling the problem and finding algorithms that aim at allocating receivers to WDM channels. These algorithms are run in a centralized fashion, assuming a perfect knowledge of the traffic scenario.

It is straightforward to notice that the solution of the node allocation problem depends on the traffic on the network. Although this traffic matrix could be dynamically estimated, we suppose for simplicity that the traffic matrix is known.

The problem can be formalized in terms of ILP *(Integer Linear Programming)*, and it can be shown to be equivalent to the problem (well-known in operations research) of scheduling jobs on identical parallel machines, which falls in the class of NP-hard problems. The problem states that, given W wavelengths and N nodes, the receiver bandwidth load can be expressed as:

$$l_i = \sum_{j=1}^{N} p_{ji} r_j \qquad \forall i,\ 1 \leq i \leq N$$

where r_j represents the transmission rate of node j and, p_{ji} its transmission probability to node i. A set of control variables x_{ik} can be defined, where:

$$x_{ik} = \begin{cases} 1 & \text{iff node } i \text{ receives on wavelength } k \\ 0 & \text{otherwise} \end{cases}$$

Receivers allocation is to be done trying to minimize $L_{\max}$, i.e. the load on the most loaded wavelength $L_{\max} = \max_k \sum_{i=1}^{N} l_i x_{ik}$. Thus our problem formulation becomes:

$$\text{Minimize } L_{\max}$$

subject to the following constraints:

$$L_{\max} \geq \sum_{i=1}^{N} l_i x_{ik} \qquad \forall k,\ 1 \leq k \leq W \tag{3.2}$$

$$\sum_{k=1}^{W} x_{ik} = 1 \qquad \forall i,\ 1 \leq i \leq N \tag{3.3}$$

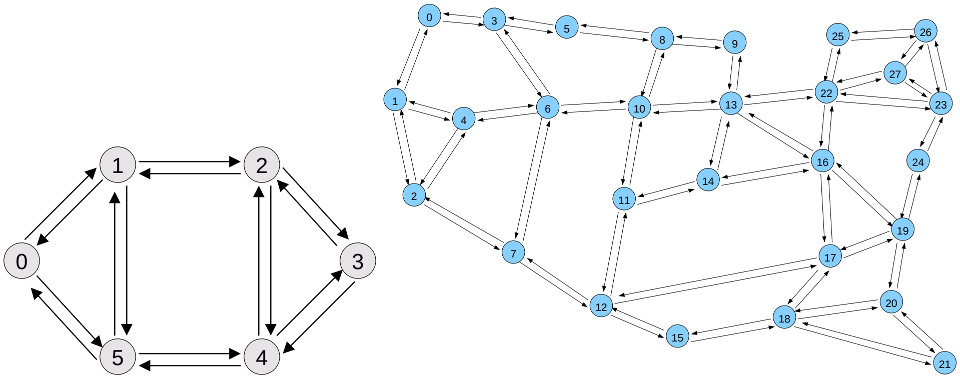

Figure 3.1: Simple Six-node topology (on the left) and US Network topology (on the right).

$$x_{ik} \in \{0,1\} \qquad \forall i,\ 1 \le i \le N \quad \forall k,\ 1 \le k \le W \tag{3.4}$$

Eq. (3.2) ensures that no wavelength has a load larger than $L_{\max}$. Eq. (3.3) ensures that each receiver must be allocated to only one wavelength.

The complexity of the optimal solution may be too large. We sketch a simple but effective heuristic that solves the problem of receiver allocation with a low complexity [it can be shown to be $O(NW\log(W))$]. The algorithm iterates through the following three steps:

1. Order all receiver loads as a non-increasing sequence.
2. Allocate the first receiver of the sequence on the least loaded wavelength, and delete it from the sequence.
3. If the sequence is empty then EXIT; else GOTO step 2.

This algorithm is known, in operations research, as LPT (*Longest Processing Time*) [15].

3.3 Discussion of results

3.3.1 Dynamic Super-Lightpath in WANs

In this Section we compare the performance of our approach with classical wavelength-routed networks. We considered two different physical topologies, depicted in Fig. 3.1. The first one (left part of Fig. 3.1) is a very simple topology, which was proposed in [10] as benchmark topology. It comprises 6 nodes and 8 bidirectional links. The second topology (right part of Fig. 3.1) is derived from a possible U.S. Long-Distance Network [11] comprising 28 nodes and 45 links. 16 wavelengths are supported on each physical link. We remind that that no wavelength conversion is available at OXCs.

New connection requests for destination s arrive at node s according to a Poisson process with parameter $\rho_{s,d}$. The connection duration is supposed to be exponentially distributed with parameter 1. Two traffic scenarios have been used to asses the performance of the algorithm:

- *Uniform traffic scenario*, in which traffic is supposed to be uniformly distributed among the nodes, i.e., $\rho_{s,d} = U[0.5, 1.5]\,\forall s, d$; $U[min, max]$ is random variable distributed according to a uniform distribution between min and max.
- *High-Low load traffic scenario*, in which 1/5 of nodes (selected at random) are considered “High” traffic nodes, and the remaining 4/5 of nodes are considered “Low” traffic. Then $\rho_{s,d} = U[9.5, 10.5]$ if both s and d are labeled as High traffic; $\rho_{s,d} = U[4.5, 5.5]$ if either s or d, but not both, are labeled High traffic; $\rho_{s,d} = U[0.5, 1.5]$ if both s and d are labeled as Low traffic nodes.

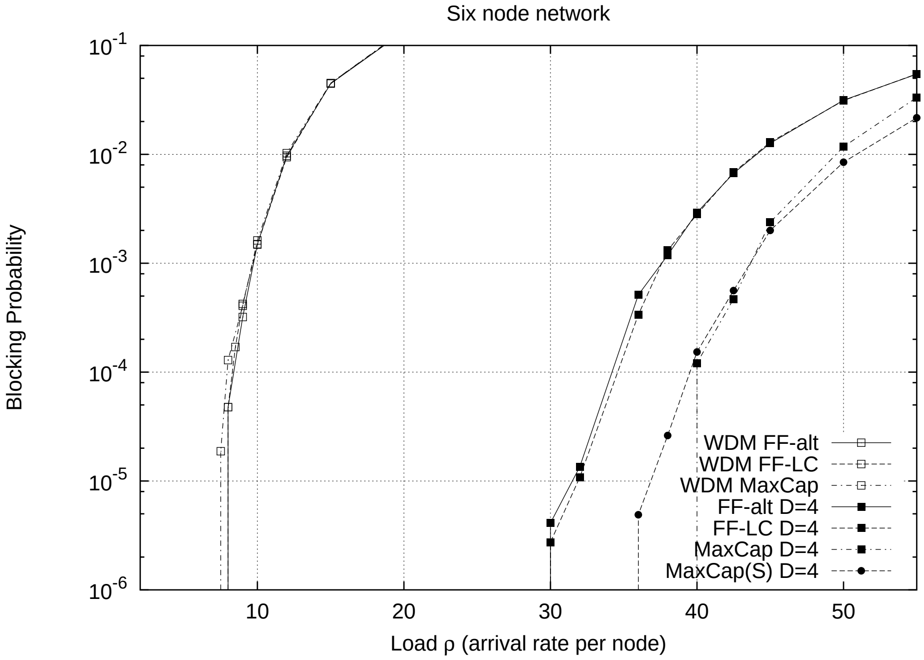

Figure 3.2: Blocking probability: six node network.

For all simulations, the accuracy of the results was assessed using the "batch means" technique with 95% confidence intervals.

Fig. 3.2 reports the simulation results obtained by considering the six-nodes network under uniform traffic pattern. It reports the blocking probability versus the load offered to the network, which is defined as the number of connection requests per time unit per node.

Both classical RWA algorithms and algorithms that jointly exploit WDM and OTDM are considered. As expected, the use of an OTDM multiplexing scheme, superposed to the WDM scheme, leads to a significant reduction of the blocking probability, since it exploits the grooming of several connections onto the same Super-Lightpath. If we identify the maximum amount of traffic that can be transported in the network under the constraint of a maximum blocking probability, we can observe that the capacity of the network is more than four times larger by using a OTDM multiplexing factor $D = 4$; e.g., fixing the blocking probability to 10^{-3}, the offered load carried by a traditional solution is about 10, while adding the OTDM multiplexing scheme the offered load the network can support becomes about 44 using the MaxCap algorithms. This is obtained by multiplying by 4 the capacity of each wavelength, but without increasing the transmission and reception rate at nodes.

Looking at how the different algorithms perform, we observe no differences when the classic RWA problem is faced. On the contrary, the OTDM case makes the problem more difficult, and the increased complexity of the MaxCap and MaxCap(S) algorithms allows to better exploit network resources.

In the left plot of Fig 3.3, the connection blocking probability versus traffic load is reported considering the US network topology under Uniform traffic, while the right plot reports results when the High-Low load traffic pattern is simulated. Both classical RWA algorithms and their extension to the OTDM case with a multiplexing factor of $D = 4$ are considered. As expected, also in these scenarios there is a large reduction of the blocking probability by exploiting the grooming of several connections onto the same Super-Lightpath.

Note that the FF-LC algorithm performs slightly better than FF-alt algorithm. This is mainly due to the intrinsic capability of FF-LC to adapt the routes of new lightpaths to the instantaneous congestion state in the network, resulting in a better usage of network wavelengths. The difference is more evident in the RTWA problem, where the point-to-multipoint paths are longer and therefore the impact of a smarter routing choice is more effective. But also in these cases, the MaxCap and MaxCap(S) algorithms outperform the FF-LC algorithm, also when considering the classic RWA problem. Notice that the blocking probability results are very similar when considering the Uniform or the

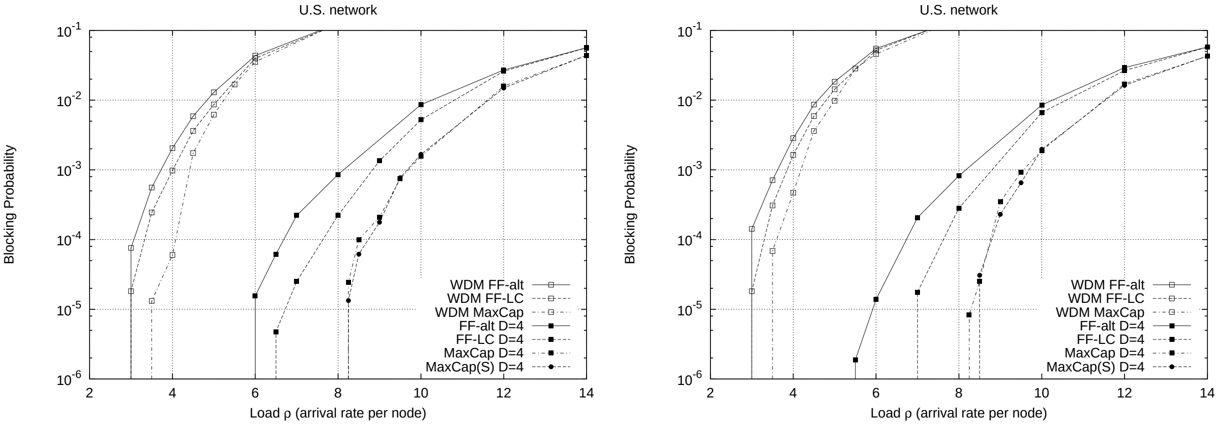

Figure 3.3: Blocking probability: US Network under Uniform Traffic scenario (on the left) and High-Low traffic scenario (on the right).

High-Low load scenario. This is due to the fact that the increased variability in the traffic pattern when considering the High-Low load scenario does not influence the blocking probability averaged among all users, but rather the blocking probability of single traffic relations.

We now give more detailed results considering the US network topology and the Uniform traffic pattern only. The MaxCap(S) algorithm is used to solve the RTWA problem, because it is the algorithm that exhibits the best performance, while maintaining the computation complexity limited. Fig 3.4 plots the blocking probability versus the traffic load for different values of multiplexing factor D. For small values of D, a significant performance gain is observed by increasing the OTDM multiplexing factor D. For larger values of D (when the factor D becomes comparable with the number of nodes in the topology) the performance gain, instead, becomes less significant. This effect is mainly due to the fact that, for large values of D, the beneficial reduction of the number of Super-Lightpaths required to sustain the traffic that would result from a further increase of D, is compensated by the increased Super-Lightpaths span.

3.3.2 Resource allocation in MANs

We studied the performance of different allocation algorithms in simple ring topologies. In particular, we considered an unidirectional ring with 16 nodes, where transmission and reception bands are decoupled, and each band has 4 wavelengths. In this scenario, two nodes named servers transmit at high load, equal to the capacity of one wavelength per server, with equal probability to the remaining 14 nodes, called clients. Client nodes transmit only to servers at a lower rate, equal to $\frac{1}{14}$ of the channel capacity. Hence the input and output load for all servers and for all clients are the same.

The access protocol is based on time slotting, and transmitters give priority to in-transit traffic (i.e., to transmission of upstream nodes), while packets from the longest packet queues are transmitted in free slots. One packet queue is maintained at each node for every destination.

Performance results are plotted in Fig. 3.5, which shows the throughput versus input load (both normalized to the available network capacity) for three different modes of allocating nodes to wavelengths. In particular we compare the optimal allocation obtained with the algorithm described in the previous section with two other allocations. In the first one, called *random allocation*, each node is randomly allocated to one of the W wavelengths, this allocation is fixed ”a priori” at system startup; the plotted curve show an average among several random node allocations. The second one, called *simple allocation*, is similar to the random allocation, but we also force that the number of allocated nodes on each wavelength is the same. We can observe that a non optimal solution to the allocation problem may lead to significant reductions of the total network throughput.

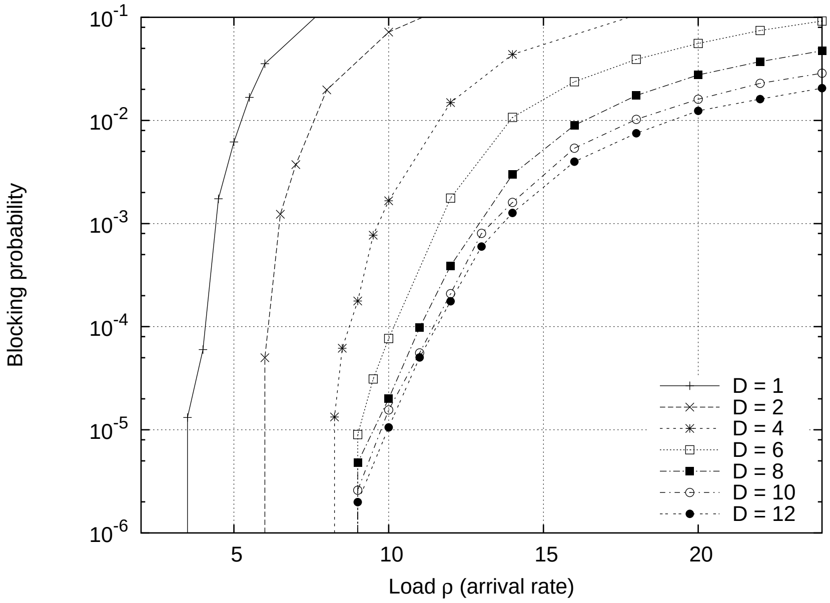

Figure 3.4: Blocking probability: US network under Uniform Traffic for different values of D.

3.4 Conclusions and future work

We have shown how the use of the WDM technology can be improved to both reduce network costs, and to increase resource utilization. In particular, considering the WAN case, we have proposed the adoption of an Optical Time Division Multiplexing scheme together with the classic WDM multiplexing scheme to support IP tunnels in dynamically reconfigurable wavelength routed networks. By grooming at the source node several IP-tunnels on the same wavelength, and by extending classic RWA algorithms, we have shown by simulation that the adoption of a joint WDM-OTDM multiplexing technique can lead either to a reduction of network costs, or to a significant improvement of the network performance. This is obtained with an increase of the aggregate network capacity, which however comes at a limited cost in the optical domain, and additional multiplexing/demultiplexing at network nodes, where complexity is minimal due to our choice of having one source in each Super-Lightpath.

Considering the MAN case, we studied the impact on network performance of different allocation policies of transceivers on WDM ring topologies under dynamic datagram traffic. We showed that greedy allocations may have great impact on network performance, while efficient allocation algorithms help to find suitable solutions based on actual traffic. Despite the fact that the traffic matrix upon which the receiver allocation is chosen must be known a priori, it can show variations over time, i.e., it can behave as a dynamic matrix. In this case it may be worthwhile to re-allocate receivers dynamically in order to keep the network in an optimal operation point.

As future work, we intend to take into account that traffic estimates are known with a certain degree of uncertainty, so that novel “robust” algorithms must be devised in the design of logical topologies. For example, in the case of resource allocation in MANs, instead of a full knowledge of the traffic matrix, a partial knowledge of this matrix can be considered if the traffic is estimated by measurements. While, traffic may be measured quite accurately on unloaded channels, traffic uncertainty arises in measurements done on overloaded channels.

Another topic of future research is the study on the impact of the logical topology reconfiguration on the physical layer. Indeed, when a lightpath is set up, the quality of the signal to noise ratio at the physical layer may change due to physical phenomenon. For example, the working point of optical amplifiers may change, as more power is present on a fiber, and therefore transient phases may cause problems not only on the novel allocated lightpath, but also to already established connections. This is of particular interest when considering the dynamic traffic case, in which reconfigurations can be frequent.

Besides, future investigations for the MAN WDM ring networks will concern the study of dynamic traffic and reallocation of receivers. A feasible solution to this problem is to introduce tunability in node receivers. This tunability

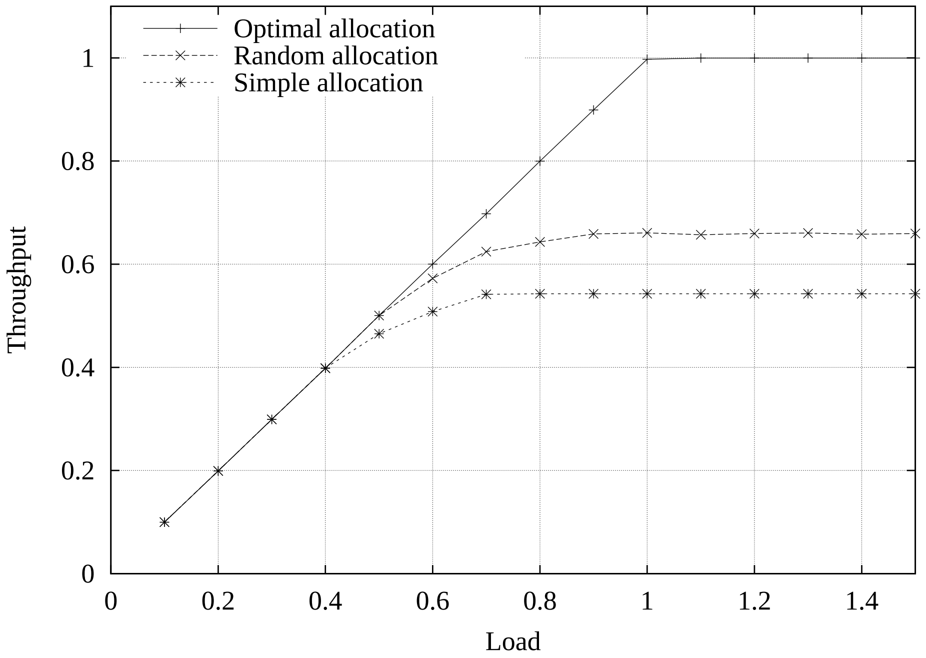

Figure 3.5: Normalized network throughput versus input load for three different allocations of node receivers to WDM channels.

does not need to be fast, and does not need to track packet-by-packet variations. Low-cost devices available today (e.g., mechanical or thermo-optic filters) can be suitable to implement this slow receiver tunability feature. It should be clear that a trade-off arises between keeping the receiver allocation well matched to the dynamic traffic matrix to optimize performance, and throughput losses due to black-outs when receivers are tuning.

In the next references section, items in **boldface** are publications generated at Politecnico di Torino in the first year of the Adonis project.

Bibliography

[1] M. Mellia, E. Leonardi, M. Feletig, R. Gaudino, F. Neri, “Exploiting OTDM technology in WDM networks”, *IEEE Infocom 2002*, New York, NY, USA, pp. 1822-1831, June 2002

[2] M. Mellia, E. Leonardi, M. Feletig, R. Gaudino, F. Neri, “Exploiting OTDM in Wavelength Routed Networks”, *OFC 2002*, Anaheim, CA, March 17-22, 2002

[3] **P. Petracca, M. Mellia, E. Leonardi, F. Neri, “Design of WDM Networks Exploiting OTDM and Light Splitters”, Proceedings *2*nd *International Workshop on QoS in Multiservice IP Networks (QOS-IP 2003)*, Milan, February 2003, available in M. Ajmone Marsan, G. Corazza, M. Listanti, A. Roveri (eds.), *Quality of Service in Multiservice IP Networks*, Lecture Notes in Computer Science, vol. 2601, Springer, pp. 433-446, 2003**

[4] **L. Calafato, *Algoritmi di Instradamento di Connessioni Dinamiche in Reti Tutto Ottiche con Supporto OTDM*, Tesi di Laurea, Politecnico di Torino, January 2003, *in Italian***

[5] **L. Calafato, M. Mellia, E. Leonardi, F. Neri, “Traffic Grooming Using OTDM in Dynamic WDM Networks”, submitted to *IEEE International Conference on Communications (ICC 2004), Optical Networking Symposium*, Paris, France, June 2004**

[6] **L. Calafato, M. Mellia, E. Leonardi, F. Neri, “Exploiting OTDM Grooming in Dynamic Wavelength Routed Networks,” submitted to *IFIP TC-6 (Task Group on Photonic Communication Networks) Working Conference on Optical Network Design and Modelling (ONDM’04)*, Ghent, Belgium, Febr. 2004**

[7] K. Uchiyama, T. Morioka, “All-optical signal processing for 160 Gbit/s/channel OTDM/WDM systems”, *Optical Fiber Communication Conference OFC 2001*, paper ThH2, Los Angeles, CA, Mar. 2001

[8] D. Marcenac, A. Ellis, D. Moodie, “ 80 Gbit/s OTDM using electroabsorption modulators”, *Electronics Letters*, Vol. 34, n. 1, pp. 101-103, Jan. 1998

[9] U. Feiste et al., “160 Gbit/s transmission over 116 km field-installed fibre using 160 Gbit/s OTDM and 40 Gbit/s ETDM”, *Electronics Letters*, Vol. 37, n. 7, pp. 443-445, Mar. 2001

[10] H. Zang, J.P. Jue, B. Mukherjee, “A review of routing and wavelength assignment approaches for wavelength-routed optical WDM networks,” *SPIE Optical Networks Magazine*, vol. 1, no. 1, Jan. 2000

[11] K. Murakami and H. Kim, “Joint optimization of capacity and flow assignment for self-healing ATM networks”, *IEEE ICC 1995*, Seattle, WA, pp. 216-220, June 1995

[12] I. Roudas, N. Antoniades, D.H. Richards, R.E. Wagner, J.L. Jackel, S.F. Habiby, T.E. Stern, A.E. Elrefaie, “Wavelength-domain simulation of multiwavelength optical networks”, *IEEE Journal on Selected Topics in Quantum Electronics*, Vol. 6, No. 2, pp. 348-362, Mar/Apr. 2000

[13] M. Ajmone Marsan, A. Bianco, E. Leonardi, M. Meo, F. Neri, “MAC Protocols and Fairness Control in WDM Multi-Rings with Tunable Transmitters and Fixed Receivers”, *IEEE Journal of Lightwave Technology,* Special issue on “Multiwavelength Optical Technology and Networks,” Vol. 14, No. 6, pp. 1230-1244, June 1996

[14] **A. Carena, V. De Feo, J. Finochietto, R. Gaudino, F. Neri, C. Piglione, P. Poggiolini, ”RINGO: An Experimental WDM Optical Packet Network for Metro Applications,” submitted to *IEEE Journal on Selected Areas in Communications (JSAC)*, special issue on “Advances in Metropolitan Optical Networks (Architectures and Control),” 2003**

[15] M. Pinedo, *Scheduling : theory, algorithms, and systems*, Prentice Hall, 2002

Chapter 4

Università di Trento Research Unit

R. Battiti, E. Salvadori, Y. Guo
Dipartimento di Informatica e Telecomunicazioni - Università degli Studi di Trento
Via Sommarive 14 - 38050 Povo (TN), Italy
{battiti,salvador,guo}@dit.unitn.it

Abstract

During the first year of the ADONIS project, the University of Trento Research Unit concentrated its activities on these two main topics:

- the proposal of resource allocation and contention resolution algorithms suited for the optical technology and compatible with the dominant IP paradigm
- the development of efficient algorithms and protocols for establishing lightpaths in wavelength-routed networks with dynamic traffic demands

In particular, the activity focused on the study of the network control and management protocols needed to efficiently realize dynamic management of wavelength routed networks.

A first goal of the activity was the study of control protocols based on the emerging (Generalized) Multi Protocol Label Switching framework. In particular, two proposals are described in this report: the first is a load balancing mechanism for MPLS networks which optimizes the allocation of the so-called Label Switched Paths (LSPs) in order to minimize the blocking probability. Such a scheme could be adopted in next-generation optical networks, whose control plane would most probably be based on G-MPLS. The second is a Traffic Engineering (TE) scheme which routes sub-wavelength requests with QoS requirements in an IP over WDM optical network.

The second part of the activity is referred to protocols which allow to dynamically establish lightpaths in a wavelength routed network. On the basis of investigation of existing control approaches, a distributed control protocol for establishing lightpath is presented, which utilizes a hybrid resource reservation scheme from Source Initiation Reservation (SIR) and Destination Initiation Reservation (DIR), and an efficient data transmission scheme, namely Conditional One-way.

4.1 Introduction

The research activity developed by the Research Unit of the University of Trento focused on the following subjects:

- Traffic Engineering schemes based on (G-)MPLS
- Distributed Lightpath Control Protocol

Generalized Multi-Protocol Label Switching (G-MPLS) is emerging as the control-plane solution for next-generation IP over WDM optical networks [1]. G-MPLS is an extension to MPLS which enables Generalized Label Switched Paths (G-LSPs) such as lightpaths to be automatically set-up or torn down by means of a signalling protocol. One of its most interesting application is Traffic Engineering (TE), whose main objective is to optimize the performance of a network through an efficient utilization of resources and to provide for Quality of Service (QoS) guarantees.

In these networks, optical nodes can be optical cross-connects (OXC) only or IP/MPLS routers, namely Label-Switched Routers (LSRs), connected to OXCs or even with WDM interfaces on their ports. A sub-wavelength connection request (an *electronic* LSP) could be routed over a direct lightpath (a single-hop path at the IP level) connecting an ingress router to an egress router or over a sequence of lightpaths (a multi-hop path at the IP level), crossing many intermediate LSRs along its route.

G-MPLS inherits most of the MPLS mechanisms to properly route sub-wavelength requests over optical connections [1]. As in MPLS, the ingress routers use the network resource information (periodically updated through specific link-state routing protocols, e.g. OSPF-TE [2]) to perform explicit routing of electronic LSPs, by using some constraint-based routing (CBR) scheme. When new lightpaths must be set-up in the optical network to route incoming requests, G-MPLS runs some routing and wavelength assignment (RWA) algorithm. Once the LSP path is decided, the ingress router uses a signalling mechanism such as RSVP-TE to effectively route the connection [3].

Inside this framework, our research activity focused first on the IP layer only, by studying how to optimize the load in order to minimize the network congestion. Novel schemes for Traffic Engineering in both Optical Packet Switching networks and MPLS networks have been recently proposed, which main objective is to exploit network resources by performing near-optimal allocation of the requests through local search techniques [4, 5, 6]. The proposed algorithms and the main results obtained with these new schemes are reported in the next section.

When considering the routing of sub-wavelength requests in optical networks, the problem is called traffic grooming and it has been proved to be NP-hard [7, 8]. Many dynamic grooming algorithms have been proposed recently, among them [7, 9], but unfortunately the QoS requirements are not considered in none of them, thus no attention is paid on the impact of the route selection both in term of delay and signal degradation. In the next section a novel scheme to guarantee QoS requirements to dynamic arriving connection requests in an IP over WDM optical network is presented, based on [10].

A critical issue in control and management of WDM networks is the control protocol for lightpaths establishment. In order to establish and terminate such connections, we need appropriate control mechanisms to select routes and wavelengths for them. This problem is referred to as Routing and Wavelength Assignment (RWA) [11, 12]. In addition RWA, we also need provisioning protocols to exchange control information among nodes and to reserve resources along the selected route. These mechanisms are expected to be scalable and be able to support highly dynamic and bursty traffic, meanwhile with a low degree of blocking for connection requests. In the next section a new protocol called Source and Destination Cooperative Reservation with Conditional One-way (SDCRCOW) is proposed, based on [13, 14, 15].

4.2 Description of the activity

4.2.1 Traffic Engineering based on (G-)MPLS

The description of the activity related to the proposal of Traffic Engineering (TE) schemes for IP/MPLS over WDM networks is divided in two parts. In the first part an on-line load balancing mechanism for MPLS networks is described, which optimizes the LSPs allocation to minimize the blocking probability. Then in the second part a novel scheme to route sub-wavelength requests with QoS requirements in a wavelength routed network is described.

Load balancing in IP/MPLS networks

According to IETF RFC 3272, TE schemes for congestion control in Internet can be classified according to the their congestion management policies: preventive schemes allocate paths in the network in order to prevent congestion, while reactive schemes act only when a congested network state is detected. Preventive schemes are usually based on Constraint-Based Routing (CBR) algorithms, which are characterized by complex path computation algorithms, which try to exploit the network resources according to the load already installed. When LSPs are set-up and torn-down dynamically, these schemes are characterized by a common drawback: they can lead to inefficiently routed paths and to future blocking conditions over specific routes.

A novel reactive scheme which uses load balancing mechanisms based on different Local Search heuristics to reduce the congestion in an MPLS network is presented in [5, 6]. The load balancing routine is based on the idea to efficiently re-route LSPs from the most congested links in the network, in order to balance the overall links load and to allow a better use of the network resources. Network congestion can be detected in two main ways: either when the load on some network links is dangerously close to the link capacity, or when a new LSP demand request cannot be satisfied. Two specific schemes are then considered:

1. First-Improve Dynamic load balancing, called $\mathtt{FID}(x)$. The parameter x indicates the threshold for the link residual bandwidth measured as a fraction of the link capacity, which determines when a link is considered congested. Each new request is routed with Weighted-shortest Path (WSP), a modified Shortest-Path algorithm which runs on a graph where link weights are defined as $w_l = 1/c_l$, where c_l is the residual available link capacity, ∞ if c_l is zero. After routing, if less than x residual bandwidth is left on some link, the dynamic load balancing algorithm is executed. As soon as the alternate route for one LSP has more residual bandwidth than the original route, the search stops and the LSP is rerouted. If the search cannot find any improving alternate LSPs in the network for all congested links, rerouting is not performed.

2. Lazy First-Improve Dynamic load balancing, called `LFID`. In this version, the dynamic load balancing is activated when a new LSP request arrives that cannot be satisfied. Now only links having the smallest residual bandwidth are considered congested. The algorithm goes through LSPs crossing them until any improving alternate route computed with WSP is found. If the search is successful, the LSP is rerouted over the path found and another attempt is made to establish the new LSP request. If it fails, the new request is rejected.

Traffic Engineering in G-MPLS networks with QoS guarantees

The second part of the activity is related to the proposal of novel schemes to dynamically route sub-wavelength connection requests by guaranteeing their QoS requirements. When an optical network with traffic grooming is considered, very little attention has been paid so far on the effects of both the physical constraints characterizing the optical layer and the delay restrictions of a multi-hop path over the traffic carried on the wavelengths. If for example some packet-loss sensitive traffic (e.g. real-time video) is carried over a lightpath experiencing strong transmission impairments, the resulting SNR degradation could be so high to compromise the signal quality requirements. If a connection carrying delay-sensitive traffic (e.g. Voice-over-IP) is routed over a multi-hop path, the output signal could suffer very high end-to-end delay due to optical-to-electronic (o-e-o) conversions and queueing delays experienced along the path. Traffic can then be divided into High Priority (HP) or Low-Priority (LP).

A novel Traffic Engineering scheme for IP over WDM networks is proposed in [10], which is based on a dynamic grooming algorithm which finds a route to fulfill the QoS requirements of the incoming request and on a preemption mechanism which guarantees lowest blocking probability to HP traffic. Furthermore, the proposed preemption mechanism is based on an algorithm executed locally in the ingress router which minimizes the network disruption and the signaling complexity.

When a new traffic request arrives in an IP over WDM network, it can be routed on the current set of lightpaths (the so-called Virtual Topology - VT) or on a new lightpath (which would modify the VT). As a result, a request can be routed over a direct lightpath (a single-hop path at the IP level), if it crosses only optical cross-connects (OXCs) nodes between an ingress and an egress router, or over a sequence of lightpaths (a multi-hop path at the IP level), if it crosses intermediate routers. Then a connection request carried into a single or multi-hop path in the virtual topology can experience delay and packet loss. Most of delay suffered by a request derives from the queuing delays in IP/MPLS routers and from o-e-o conversion delays in regenerators and electronic wavelength converters. In the following we assume that *delay-sensitive* applications can cross no more than C_{max} router nodes. The transmission impairments that digital transmission experiences along a lightpath can impact the packet loss ratio of the connection carried over the optical path. In the following we assume that packet-loss sensitive traffic can be routed only over lightpaths characterized by stringent BER requirements. Furthermore, by assuming the simplified hypothesis that all the fiber links introduce the same level of transmission impairments, the problem of selecting a good lightpath for packet-loss sensitive traffic can be reduced to the problem of limiting the maximum number of hops for the lightpath which carries this type of traffic. Then we assume that *packet-loss sensitive* applications can be routed over lightpaths whose route is made of no more than H_{max} fiber links each.

In the rest of the description we consider two Class Types (CT) only: a HP class, characterized by minimum end-to-end delay and low packet loss probability (e.g. high-quality real-time video), and a LP class with no QoS requirements (e.g. Best-Effort). Given an IP over WDM network with connection requests belonging to different CT arriving dynamically, the objective of our on-line TE scheme is twofold: first, it must route the request according to the specific QoS requirements and second, it must balance the allocation of the already established LP connections in the network to maximize the success probability for HP traffic demands. The TE scheme is then divided in two stage: a "QoS-aware" grooming algorithm and a preemption mechanism which is triggered only when a HP request is blocked.

Two grooming policies are proposed for the "QoS-aware" grooming algorithm: one is called `VT-first`, it is characterized by always trying to route the request over the existing lightpaths before modifying the VT: while the utilization of the optical resources is definitely improved, the required electronic processing at intermediate hops is also increased. `PT-first` instead forces the set-up of new lightpaths then increasing the logical connectivity at IP level, but leading to a heavy usage of the available optical resources.

In pure IP/MPLS networks the preemption is implemented in a distributed way by using the RSVP signaling protocol. Basically, when an LSP having priority p needs to be set-up, its priority is sent in the PATH message along the route selected in the ingress router. Each time this message reaches an intermediate LSR, if the available capacity on the outgoing link is not sufficient to carry the request, a local selection of the lower-priority LSPs to be preempted is performed by the router, and then the proper notification messages is sent upstream to the routers which should try to reroute (or block) these LSPs. This mechanism can potentially involve all the edge LSRs in a network to finalize the set-up of one high-priority LSP. Many distributed algorithms for routers implementing a preemption mechanism have been proposed in literature to select the best LSP to reroute, according to different objectives (typically: minimizing the number of reroutings in the network). In [16, 17] the proposed algorithms are optimal with respect to their objective functions: we can indicate them as global preemption algorithms (GPA), to highlight this optimal behavior and the involvement of many edge routers in the network. Thus from the point of view of the signaling, they could result in a large amount of RSVP messages around the network which increase the amount of overhead. When considering a network based on G-MPLS, the complexity of this mechanism is even more complex, because the signaling must take into account much more information regarding the optical layer.

The proposed TE scheme considers a preemption mechanism which is based on a simpler implementation both from the algorithmic and the signaling point of view. In the following, the proposed algorithm is called local preemption algorithm (LPA) to distinguish it from the optimal one (GPA). The idea comes from the property of LSR routers in MPLS to have complete information about the crossing LSPs (each node maintains information about all the LSPs that originate, terminate or cross the node itself [18]). Then, by focusing only on the ingress routers of the network, the simplest suboptimal preemption mechanism to apply is to perform local selection of one or few low-priority LSPs to preempt when a high-priority LSP is blocked. An important function of the algorithm is the one which assigns the weights to all the potential LP LSPs to reroute [10]. If some LSPs to reroute is found, then the incoming HP request must be routed over its path, while the LP LSPs must be rerouted by using some dynamic grooming algorithm (or blocked if necessary). Compared to GPA, LPA allows to minimize the number of LSPs to reroute, which are more than one only in the case we need to preempt some low-bandwidth LSP, with great benefit for the network disruption. Furthermore, by limiting the execution of the preemption algorithm to the ingress router s only, instead of having an algorithm executed in many LSRs along the request path as in GPA, very few RSVP messages would flow through the network to manage the preemption. In particular the signaling involves the ingress router and some edge LSRs only if the low-priority LSPs to reroute are not originated in s.

4.2.2 Distributed Lightpath Control Protocol in WDM Networks

In the near term, most applications will run over optical network indirectly (e.g. though SONET over WDM, IP over WDM). In this case, the traffic pattern is static and slow-changing in which the arrival rate of connections is low and the connection duration is relatively long. Therefore, the main attention of lightpath control protocols is focused on the implementation in a simple way and the efficiency of them is not critical. For establishing lightpaths in such situation, many control protocols have been studied in centralized way [19, 11, 12] and distributed way [20, 21, 22, 23], respectively. However, in the long term, it is believed that, more and more applications will run over optical networks directly and transparent optical networks with wavelength routing will prevail. Such networks will provide the capacity of direct lightpath among edge nodes. At the same time, the traffic pattern will change a lot with the rapid appearance of new applications and services. It can be forecasted that the connection pattern will be dynamic and bursty in nature. As a result, compared with the static case, the strategy of control protocols for connection establishment in dynamic situation needs to be re-considered, and the main effort should be put on the efficiency of control schemes, which provide low blocking probability and high scalability.

The control protocol, proposed in [21], is believed to design for static connection pattern. In this protocol, the source will not transfer data until it receives all ACK messages from intermediate nodes and the destination nodes. In large networks, the propagation delay will be significant, and in turn it results in high connection setup time. In addition, in order to keep track of network state, each node maintains global network information and broadcasts this information to all other nodes periodically or in event-driven manner. Whereas, the lager networks are, the more inaccurate such information is, in turn, it will increase the blocking probability caused by the outdated information. At the same time, this approach will result in high control overhead for the case of dynamic connection requests and large networks because the update progress needs to be done quite frequently. Finally, in this protocol, wavelengths are reserved by the source in the forward direction, and therefore bandwidth on the reserved wavelengths is wasted during the reservation period, which may become remarkable in the high-speed WANs.

In [22], based on a *distance-vector* routing approach, a new connection setup mechanism is present. Compared with link-state approach [21], each node in [22] need not keep the global network information; instead they maintain a routing table for each wavelength, which specifies the next hop and the cost associated with the shortest path to each destination on this wavelength. As a result, routes and wavelength for ligthapths are selected in a distributed fashion, which obtains improvement in blocking probability and scalability to a certain extent. However, in this scheme, control packets still have to travel round trip between the source and the destination before the data transferring, and hence the connection setup time will still be significant in larger networks. Moreover, wavelengths are reserved in the forward direction to the destination and therefore the problem of bandwidth wasted during reservation period still exists. Also, in both [21] and [22], it takes a certain amount of time to update each node whenever a connection changes. During this period, nodes may have outdated information leading to higher connection blocking probability.

In order to reduce bandwidth wasted and make more intelligent routing and wavelength decision, the work [23] presents a distributed lightpath control scheme based on destination routing. In this scheme, upon the arrival of a connection request, the source node sends a PROB message that does not reserve any wavelength to the destination along the shortest path. It is the destination node that decides a route and selects a wavelength based on the current network state information it maintains. However, the control overhead will still be significant in large networks because each node has to maintain global network information and keep this information being updated periodically. Finally, the source still begin to transfer data only when it receives ACK message, as a result, the connection setup time is at least twice as much as the propagation delay between the source and the destination.

On summarizing, all the distributed lightpath control approaches proposed so far suffer from both or at least one of the following disadvantages: 1) Some of them are not appropriate for emerging networks in which traffic pattern are more dynamic and bursty in nature because they are designed for a fairly static network in which connections are not changing rapidly, while some are appropriate for dynamic networks only, but not for both static and dynamic networks. 2) The performance of current approaches is low, when evaluation metrics e.g. blocking probability, control overhead, and connection setup time are calculated together at the same time. On the basis of above observation, we propose an efficient control protocol for lightpath establishment in wavelength-routed WDM networks that can guarantee good performance based on several performance evaluation metrics simultaneously in wide ranges from static to fairly dynamic networks.

Proposed protocol framework

In this section, we firstly present a novel wavelength reservation scheme, Source and Destination Cooperative Reservation (SDCR), and an efficient data transmission scheme, namely Conditional One-way. Furthermore, based on above two schemes, we propose to introduce an efficient control protocol, Source and Destination Cooperative Reservation with Conditional One-way (SDCRCOW), for lightpath establishment. In our considered network, each node maintains global network information that is updated periodically, or whenever there is a change in information database. In our protocol, the routing algorithm is Adaptive Routing algorithm [12] and we use First-Fit scheme for wavelength assignment. The details of proposed protocol and results discussion can be found in [13, 14, 15].

In our protocol, a wavelength on a given link can be in one of the following three states: **Available**, **Busy**, and **Suggested**which indicates this wavelength has been suggested to the destination for a connection on that link. In addition, in order to exchange information among nodes, some basic kinds of signaling messages are needed as follows:

CAS (Collection And Suggestion) collecting resource information along a selected route without actually reserving any wavelength and suggesting wavelength which can be used to the destination; *ACK* informing the source that wavelength suggestion is successful and occupying this wavelength (by setting the state of wavelength from Suggested to Busy) on each link along the path; *NAK* confirming that the suggested wavelength has been used by other lightpaths, and informing the source to stop one-way transmission and wait for response from the destination; *ReWA (Re-Wavelength Assignment)* sent back to the source by the destination to re-assign a wavelength for lightpath when the destination is informed that the wavelength suggestion initiated by the source failed.

Source and destination cooperation reservation

The principle of SDCR that has been implied in lightpath setup procedure can be briefly stated as follows. When a lightpath request arrives, utilizing global topology and wavelength usage information, the source node *first tries* to select a route and a wavelength, and then sends out the *CAS* message to suggest this selected wavelength along the route. If this suggestion succeeds, the Conditional One-way transmission starts, which is similar to SIR. However, if this suggestion fails, this request won't be blocked like in SIR; instead the destination will retry to perform route and wavelength determination to select a route with available wavelength for this request. At this point, it is like DIR. Briefly speaking, SDCR is a hybrid scheme of SIR and DIR that combines advantages of SIR and DIR, at the same time overcomes their drawbacks.

Conditional One-way transmission

In two-way scheme, the source does not start data transmission until it receives *ACK* from the destination. As a result, the setup time is at least twice as much as the propagation delay between the source and the destination. In one-way scheme, the source first sends out *REQ* to reserve wavelength and then starts data transmission before the *ACK* comes back. It is clear that, if the reservation fails frequently, the source will transfer much useless data, which wastes the lightpath utilization. Based on this observation, we proposed the Conditional One-way transmission scheme in which the source starts data transmission on the suggested wavelength T time units *(T=HP, where H is the number of hops for a lightpath, and P is the signalling processing time at one node)* after it sends out *CAS* message if and only if the source does not receive any *NAK* message within T time units.

Lightpath setup procedure

- Upon the arrival of a lightpath request, the source performs route and wavelength determination procedure to find a route and wavelength for the requested lightpath. Then, the source sends *CAS* massage to collect resource information and suggest an wavelength along the selected route.

- When receiving *CAS* message, each intermediate node will set this suggested wavelength's state to **Suggested** and forward the *CAS* message to the next node if this suggested wavelength is **Available**. If this suggested wavelength is either **Suggested** or **Busy**, the node will include a note indicating this suggested wavelength has been occupied by other lightpaths into the *CAS* message for the destination, and then forward the *CAS* message. At the same time, the intermediate node sends back a *NAK* message to notify the source the suggestion failure as well as stopping Conditional One-way transmission and waiting for the reply from destination.

- Once the destination node receives the *CAS* message, if it finds that the suggested wavelength is available on the entire path, it will send back the *ACK* message along the reverse route. If the suggested wavelength is not available on any link alone the path, it then re-performs the routing and wavelength determination procedure to decide a route and select a wavelength based on the *CAS* message and network information it maintains. In this case, if a route with an available wavelength can be found, it sends *ReWA* message to the source. If there are no routes with available wavelengths available, the destination will send back *NAK* message to inform the source that this lightpath setup attempt fails and try it later.

- On the other hand, at the source node, data transmission will start on the suggested wavelength T time units after the source sends out *CAS* message if and only if the source does not receive any *NAK* message within T time units. If, however, the source receives any *NAK* message within T time units, it won't start one-way transmission and will wait for messages from the destination.

4.3 Discussion of results

4.3.1 Traffic Engineering based on (G-)MPLS

The description of the numerical results related to the proposed schemes is divided in two parts. In the first part we analyze the main results regarding the new load balancing mechanism for MPLS networks, while the second part refers to the TE scheme for IP over WDM networks with QoS requirements.

Results on the proposed Load balancing scheme in IP/MPLS networks

The simulations are carried out by using an MPLS network with the same topology of [24]. In the following the performance of our reactive schemes are compared to Minimum Interference Routing Algorithm (MIRA) [24], one of the better performing preventive TE schemes, based on an on-line CBR. Then, traffic requests are limited only to 4 ingress and egress router pairs (S_1,D_1), (S_2,D_2), (S_3,D_3) and (S_4,D_4); however, it is important to highlight that our algorithms allow to relax this strong constraint. The experiments compare our algorithms with Minimum-Hop Algorithm (MHA) and MIRA. Connection requests arrive between each ingress-egress pair according to a Poisson process with an average rate λ, and their holding times are exponentially distributed with mean $1/\mu$. Ingress and egress router pairs for each LSP set-up request are chosen randomly. In order to evaluate in more detail the proposed algorithms, different set of experiments have been performed. The first set of experiments considers a uniform distribution of traffic among all the ingress-egress router pairs (symmetric traffic). The second set of experiments considers a non-uniform distribution (asymmetric traffic): in particular, it is assumed that ingress-egress pair (S_1,D_1) generates a traffic rate which is four times higher than the other pairs, on average. This set allows us to highlight MIRA limitations regarding the traffic load condition over the MPLS network. Two experiment subsets are performed for two different bandwidth distributions per LSP (maximum bandwidth 3 and 6). Blocking probability is evaluated by considering the number of LSPs blocked by the algorithm. The network is loaded with 20.000 requests during one trial.

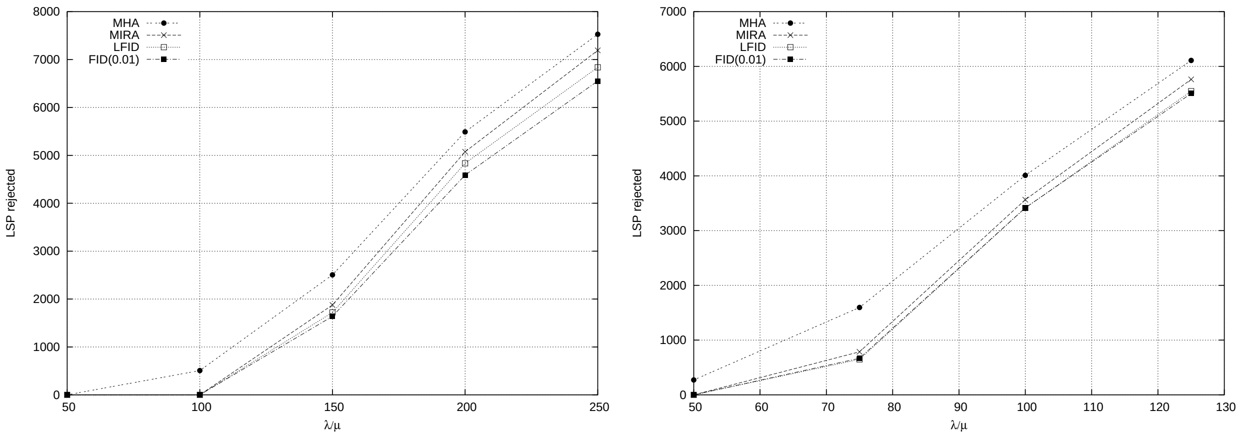

Figure 4.1: Symmetric traffic: number of rejected LSPs vs. λ/μ for maximum bandwidth equal to 3 (right) or equal to 6 (left)

Figure 4.1 presents the results for symmetric traffic. Two different bandwidth distributions per LSP are considered: the first has a maximum bandwidth of 3 units, while the second 6 units, thus simulating the case of bigger connections on average. Rejection values are calculated over 10 runs. Our algorithms perform slightly better than MIRA in these traffic conditions. The first plot shows that `FID(0.01)` performs slightly better than `LFID` for high values of network load, due to the implicit reaction mechanism: the load balancing algorithm is triggered whenever some link overcomes the congestion threshold, thus keeping a better distribution of the load over the network.

Figure 4.2 presents the results for asymmetric traffic. Two different bandwidth distributions per LSP are considered as above. The plots show that these traffic conditions lead to a higher blocking probability on average compared to the previous set of experiments. Capone et al. proved that MIRA does not perform well when asymmetric traffic is applied to the ingress-egress pairs [25]. The proposed schemes perform much better than MIRA mainly thanks to the

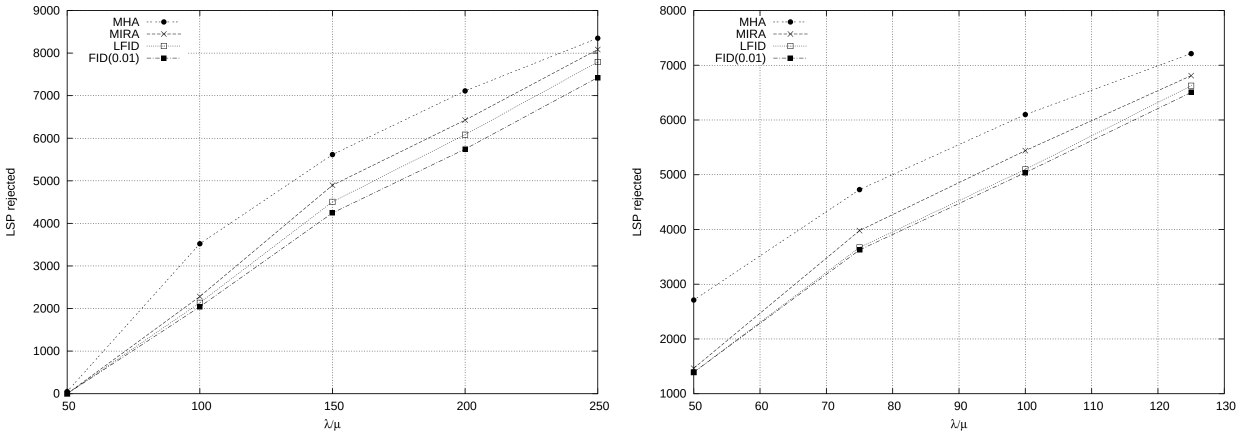

Figure 4.2: Asymmetric traffic: number of rejected LSPs vs. λ/μ for maximum bandwidth equal to 3 (left) or equal to 6 (right)

independence of our algorithms from the traffic conditions. As in the first set of experiments, FID(0.01) performs slightly better than LFID. Finally, it has also been proved that FID(0.01) leads to an high number of rerouted LSPs on average compared to LFID due to the implicit mechanism of triggering. Further details on the impact over the network disruption in term of rerouted LSPs can be found in [5, 6].

Results on the proposed Traffic Engineering scheme with QoS guarantees

When instead the proposed TE scheme to guarantee QoS requirements to sub-wavelength connection requests is considered, an extensive set of experiments have been executed on different network topologies, both with low and high number of nodes: thanks to the consistency of the results, only the graphs relating to the well-known topology of [9] are shown. Here the number of wavelengths considered in all the tests is 4, where each wavelength has a full capacity equal to 10 units, and connection requests have bandwidth demand distributed uniformly between 1 and 3 units, independently from their priority. Requests arrive between each ingress-egress pair according to a Poisson process with an average rate λ, and their holding times are exponentially distributed with mean $1/\mu$. Ingress and egress router pairs for each LSP set-up request are chosen randomly. The network is loaded with 50.000 requests during one trial, and the performances are evaluated by considering average values calculated over 10 runs. The percentage of traffic routed in the network is 60% for LP traffic and 40% for HP traffic. These percentages have been chosen in order to have appropriate results from the point of view of the blocking probability when a dynamic grooming algorithm is applied. Regarding the physical constraints used in the experiments, $H_{max} = 4$ and $C_{max} = 1$ have been chosen for the High-Priority Class (these values are very dependent on the topology of the optical network).

The first set of tests compare the "QoS-aware" grooming algorithms VT-first, PT-first and the Minimum Open Capacity Algorithm (MOCA) proposed in [9]. Traffic requests are limited only to some specific ingress and egress router pairs because MOCA can work only when this strong assumption is considered. In this first test, the same pairs as in [9] have been considered for comparison. However it is important to highlight that our scheme allows to relax this constraint.

Figure 4.3 shows the success probability of MOCA, VT-first and PT-first for each CT. MOCA performs the better for LP traffic, when no requirements are needed to route successfully a connection, while its performance is much worse than our grooming algorithms for HP traffic. This behavior is due to the fact that MOCA is a dynamic grooming algorithm whose main objective is to perform load balancing distribution of the traffic: because the average number of physical hops crossed by each request is very high, it performs very badly when HP traffic must be routed over the network.

As expected, for all the proposed dynamic grooming algorithms a higher blocking probability is experienced by high-priority traffic. In the following, we analyze the impact of the proposed LPA preemption mechanism. The assumption on the position of the ingress-egress router pairs, which are randomly selected every time a new request is

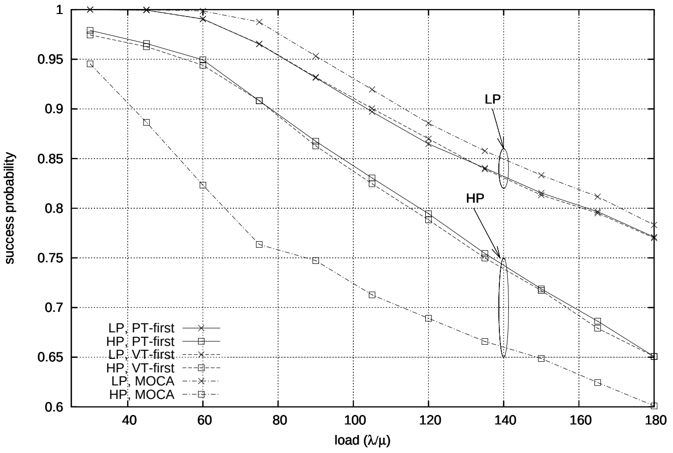

Figure 4.3: Success probability for MOCA vs. `PT-first` and `VT-first`

loaded in the network, can now be relaxed.

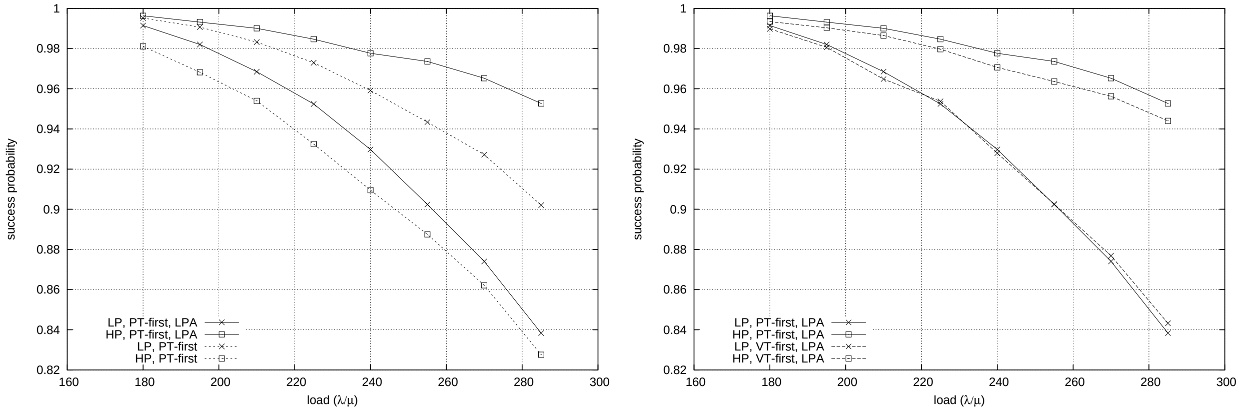

Figure 4.4: Success probability (left) with and without LPA (`PT-first`), (right) for `PT-first` and `VT-first` with LPA

Figure 4.4 (left) shows the success probability for the `PT-first` dynamic grooming algorithm when a "local" preemption mechanism (LPA) is applied. `VT-first` performs very similarly, thus results are not included: in both cases the obtained gain is quite high, and in particular it can be noticed that by using LPA, the success probability for HP traffic is increased by about 14%, while it decreases dramatically for LP traffic. Figure 4.4 (right) shows the performance of LPA when `PT-first` and `VT-first` is applied. Compared to Figure 4.3, the success probability is increased dramatically for HP traffic to the detriment of LP traffic. It can be noticed that when the `PT-first` grooming algorithm is applied, the success probability for the HP traffic class is always higher than the one obtained with `VT-first`. This can be explained by the implicit mechanism used in PT-first, which always tries to set-up a new lightpath (a direct one) when a new request is arrived, thus guaranteeing a a higher connectivity in the VT and then more available routes for HP traffic.

Figure 4.5 shows a comparison of the success probability when the proposed LPA and the optimal preemption algorithm (GPA) are applied: only the results when the `PT-first` grooming scheme is applied are indicated, because `VT-first` performs very similarly. The GPA mechanism considered in these simulations is implemented by using the mechanism proposed in the MinConn algorithm [16]. LPA performs quite well compared to the optimal algorithm,

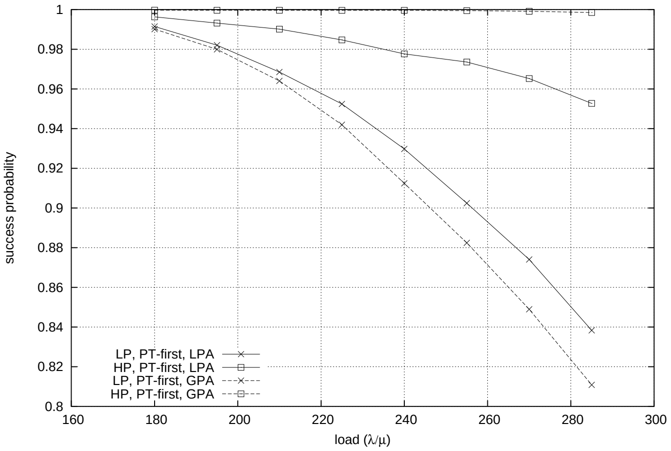

Figure 4.5: Success probability with LPA and GPA: `PT-first`

which always find a route for HP requests. In fact the success probability is quite high (more than 90%) even at high network loads, when the LP traffic experience a higher blocking probability.

When considering the impact of the proposed TE scheme in term of network disruption there are two main parameters to consider: the percentage of rerouted or blocked LP LSPs and the number of lightpaths which are set-up when LP LSPs must be rerouted in the network to leave room for incoming HP connection requests.

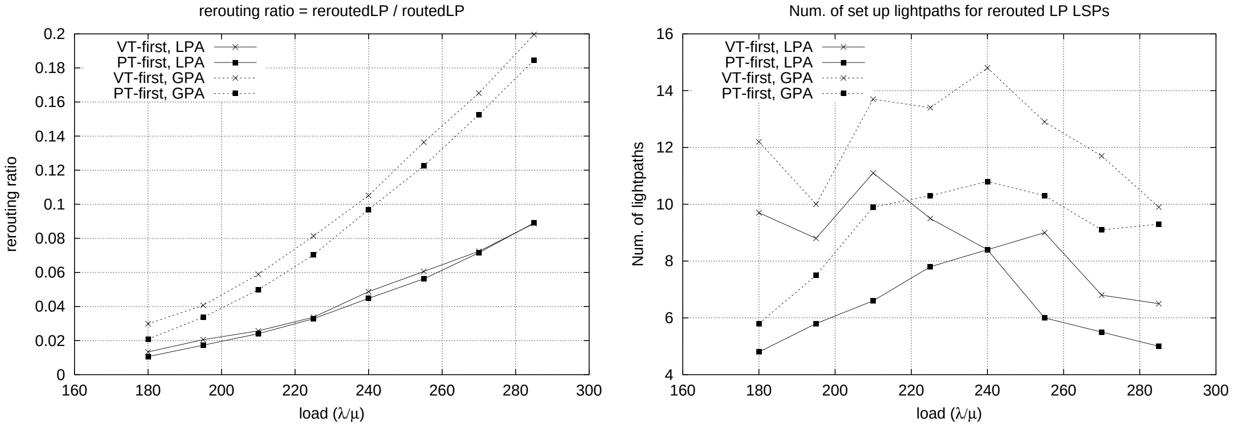

Figure 4.6: Rerouting ratio for `PT-first` and `VT-first` (left) and number of set-up lightpaths due to reroutings (right)

Figure 4.6 (left) shows the percentage of rerouted LSPs, calculated as number of rerouted LSPs over the total number of LP LSPs routed with success. As expected, when using the proposed LPA this ratio is much lower (roughly the half!) than in the optimal case. The same behavior has been verified when the ratio of blocked LP LSPs is considered. In both cases, the lowest ratio of preempted LSPs is obtained by using the `PT-first` grooming policy instead of `VT-first`.

When instead the absolute number of set-up lightpaths when LP LSPs are rerouted is considered (Figure 4.6, right), GPA sets up between 40% and 60% more lightpaths on average compared to LPA. The lower number of new lightpaths is reached when `PT-first` is used. In fact, by using `VT-first` the Virtual Topology would be highly loaded on average, thus forcing the set-up of new-lightpaths when a LP LSP needs to be rerouted. Further details on the impact over the network disruption in term of number of set-up lightpaths when LP LSPs are rerouted can be found in [10].

4.3.2 Results on the proposed Distributed Lightpath Control protocol

In this section, we evaluate our proposed protocol comparing with two current approaches, Link-state approach [21] and Distributed-routing approach [22], in terms of blocking probability, lightpath setup time and network utilization via simulation of the mesh network [26] in Path Multiplexing (PM) scenario. In our simulations, we assume: 1) The simulated mesh network has 16 nodes and 25 links (the number on the links represents link distances in tens of kilometers and the number of wavelengths on each link is 8, Figure 4.7 (left), 2) New connections will arrive according to the Poisson process, 3) The destination of each connection is uniformly distributed, 4) Connection duration has an exponential distribution with mean 100 ms, 5) Signaling message processing time at a node, P, is 10μs; the time to configure across-connect, C, is 500 μs; the average propagation delay between two nodes D is 14.7ms, and the average hop distance is H_a = 2.28.

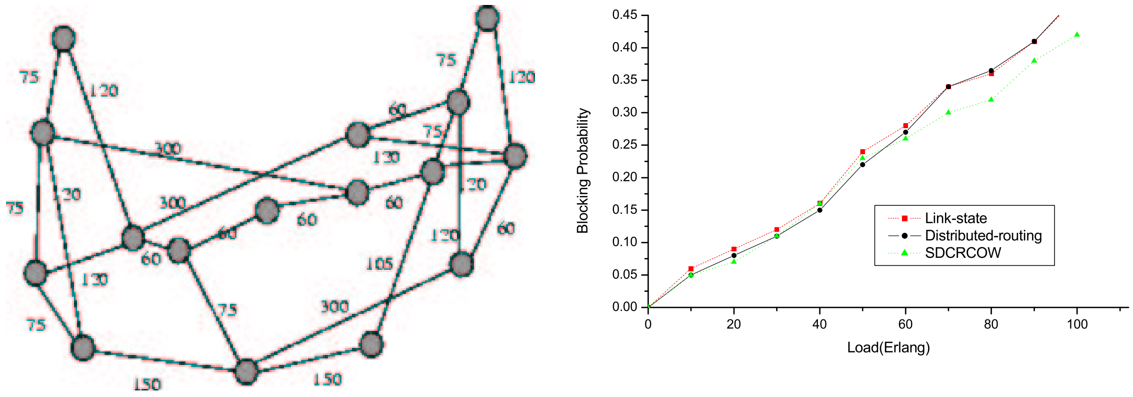

Figure 4.7: A nationwide backbone network (left), Blocking Probabilities vs. Load (right)

Figure 4.7 (right) plots the blocking probability vs. load for the three approaches. It shows that blocking in all the three approaches is quite similar under low load, but our approach has lower blocking probability than other two approaches under high load. These differences are due to the facts that under low load, our approach works in the similar way to Link-state and Distributed-routing approaches (the source first-try with quite high success probability in the forward direction). As the load increase, the reservation initiated by the source will fail frequently. Therefore, Link-state and Distributed-routing approaches have higher blocking probability. However, in our approach, the destination will retry to search for a route with available wavelength after the failure of source suggestion, which can decrease the lightpath blocking probability.

Lighpath setup time is the time required to establish a lightpath once a lightpath request arrives. When the network traffic load is light and there are no reattempts, the Average Setup Time (AST)can be formulated AST = 2D + (2H+1)P + C = 30.2 ms for Link-state approach and Distributed-routing approach, which are plot in Figure 4.8 (left). When the network load is light, we observe that the setup times in Link-state approach and Distributed-routing approach are fairly close to this bound (30.2 ms), but the setup time in our proposed protocol is significant low. This is due to the fact that, under low network load, the first-try of wavelength reservation initiated by the source succeeds with quite high probability, which in fact is one-way transmission scheme. It is clear that the setup time in one-way scheme is less than the one in two-way scheme. As the load increases, since the first-try will fail frequently, the transmission scheme used in our proposed protocol is essentially two-way scheme, in which the setup delay is at least twice as much as the propagation delay between the source and the destination. Therefore, the setup delay difference between Distributed-routing approach and our protocol is not remarkable.

From Figure 4.8 (right), we can see that networks reach the saturable state where the network utilization is around 50 percent for a load of 160 Erlangs. We also find that our protocol's performance in term of network utilization is different under different network load. Under light network load situation, our protocol obtain higher network utilization than link-state and distributed-routing approaches because the overhead in SDCRCOW is smaller than overhead in SIR, even in DIR when the *first-try* of suggestion initiated by the source succeeds. However, under heavy network load, the *first-try* of wavelength suggestion fails frequently, which causes that lightpath establishment must be

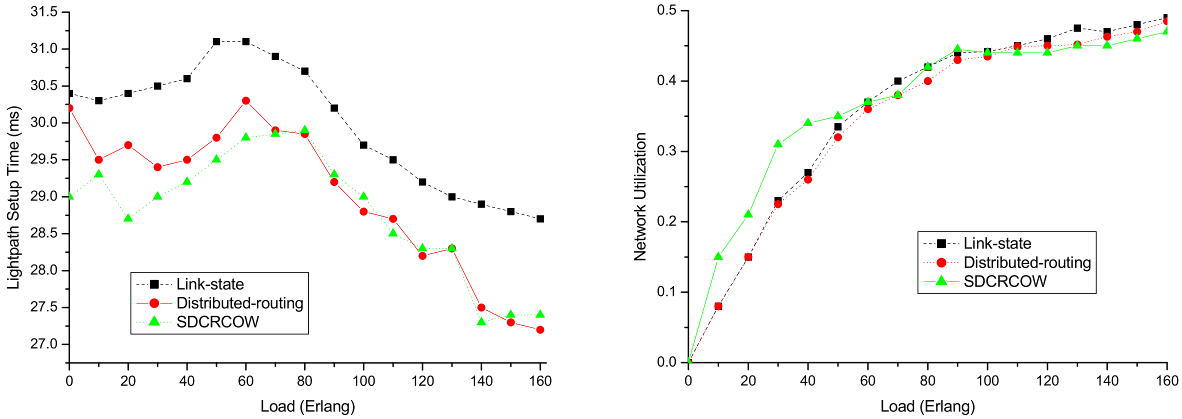

Figure 4.8: Lightpath Setup Time vs. Load (left), Network Utilization vs. Load (right)

implemented by the retry initiated by the destination. As a result, the overhead in our protocol is similar to overhead in link-state and distributed-routing, in turn, the difference of network utilization in the three approaches is not significant in heavy load situation.

4.4 Conclusions and future work

The first year of research in control and management protocols for dynamic optical networks can be divided into two main activities, one related to the proposal of new Traffic Engineering schemes for G-MPLS based networks, the other related to the study of a new distributed control protocol to dynamically establish lightpaths in an optical network.

Two new Traffic Engineering algorithms have been presented. The first is based on a load balancing mechanism for MPLS networks which optimizes the allocation of LSPs arriving dynamically in order to minimize the blocking probability. The second is a novel on-line scheme to route sub-wavelength requests with QoS requirements in a G-MPLS based optical network. Both schemes perform very well compared to existing schemes in the literature. In particular, the `LFID` scheme for load balancing in MPLS allows to reduce blocking probability by reducing the amount of rerouted LSPs in the network at the same time [5, 6]. When considering instead an optical network with on-line sub-wavelength connection requests, the proposed local preemption algorithm, associated to a dynamic grooming policy based on the maximization of the optical resources, allows to keep the blocking probability to very low values in a heavy loaded network, by guaranteeing QoS requirements to the highest-priority connections [10]. Future improvements of the proposed TE scheme should consider a more accurate representation of the network which considers the number of ports per LSR in the network or even the specific physical properties of the optical devices (fibers, OXCs, amplifiers,...) in order to better evaluate the real impact of specific lightpaths to guarantee the transmission quality of high-priority connections.

As a further work on integrated routing in IP over WDM networks, we plan to study the impact of different grooming policies when IP elastic traffic is considered instead of fixed duration traffic. When the duration of a connection depends on the amount of resources available during the connection lifetime as in Internet, the traffic is called *elastic*, because the closing time is dynamically evaluated depending on the evolving congestion level of the network [27, 28].

Furthermore, a new control protocol SDCRCOW for lightpath establishment has been proposed, which can be used in both static and dynamic wavelength-routed networks. We compared our proposed protocol with Link-state approach and Distributed-routing approach and simulated their performances in a mesh network in terms of blocking probability, lightpath setup time and network utilization. It is shown that, our protocol, SDCRCOW, obtains the gain of above three metrics.

In the following work, we plan to extend our protocol to control lightpath in LM scenario in which a lightpath can occupy different wavelengths on different links along the path. We also intend to implement the distributed restoration capacity of our protocol for handling single link failure and wavelength failure on a link. Specifically, the LST update procedure guarantees that the entire lightpath can be taken down when there is a link or a wavelength fails. Then, the originator will perform routing and wavelength determination procedure to resume this lightpath. Besides we are aware that, by maintaining global network state information, each node can make more intelligent decision on route selecting and wavelength reservation; however, this obtaining of global information also results in high control overhead and increase in blocking probability caused by outdated information due to propagation delay in large networks. High control overhead may be one drawback of our proposed protocol. On the other hand, if each node only maintains its local information, the control overhead and probability of outdated information can be decreased; but, there is a high possibility that nodes will make wrong decision on route selecting and wavelength reservation. Hence, we postulate that there is an equilibrium value for the range in which each node collects the network state information and exchanges control messages. In the following work, one issue need to be addressed is to determine the optimal range. In this range, each node does not maintain the global network state information any longer; instead, every node obtains only the wavelength information from its neighboring nodes within this optimal distance.

Bibliography

[1] A. Banerjee, J. Drake, J.P. Lang, B. Turner, K. Kompella, and Y. Rekhter. Generalized Multiprotocol Label Switching: an Overview of Routing and Management Enhancements. *IEEE Communications Magazine*, 39(1):144–150, January 2001.

[2] K. Kompella and Y. Rekhter. OSPF Extensions in Support of Generalized MPLS. IETF Draft, *draft-ietf-ccamp-ospf-gmpls-extensions-09.txt*, work in progress, December 2002.

[3] L. Berger. GMPLS Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions. IETF RFC 3473, January 2003.

[4] M. Brunato, R. Battiti, and E. Salvadori. Dynamic Load Balancing in WDM Networks. *Optical Networks Magazine*, September 2003.

[5] E. Salvadori and R. Battiti. A Load Balancing Scheme for Congestion Control in MPLS Networks. In *Proceedings of IEEE Symposium on Computers and Communications - ISCC*, volume 2, pages 951–956, Antalya - Turkey, July 2003.

[6] E. Salvadori, R. Battiti, and F. Ardito. Lazy Rerouting for MPLS Traffic Engineering. Technical report, Università di Trento, Dipartimento di Informatica e Telecomunicazioni, April 2003.

[7] H. Zhu, H. Zang, K. Zhu, and B. Mukherjee. A novel generic graph model for traffic grooming in heterogeneous WDM mesh networks. *IEEE/ACM Transactions on Networking*, 11(2):285 –299, April 2003.

[8] M. Brunato and R. Battiti. A multistart randomized greedy algorithm for traffic grooming on mesh logical topologies. In *Proceedings of the 6th IFIP ONDM*, Torino, Italy, February 2002.

[9] M. Kodialam and T.V. Lakshman. Integrated Dynamic IP and Wavelength Routing in IP over WDM Networks. In *Proceedings of INFOCOM*, volume 1, pages 358 –366, Anchorage - Alaska, 2001.

[10] E. Salvadori and R. Battiti. Traffic Engineering in G-MPLS networks with QoS guarantees. Technical report, Università di Trento, Dipartimento di Informatica e Telecomunicazioni, September 2003.

[11] R. Ramaswami and K.N. Sivarajan. Optimal routing and wavelength assignment in all-optical networks. In *Proceedings of INFOCOM* , pages 534–543, Toronto, Canada, June 1994.

[12] H. Zang, J.P. Jue, and B. Mukherjee. A Review of Routing and Wavelength Assignment Approaches for Wavelength-routed Optical WDM Networks. *Optical Networks Magazine*, 1:47–60, January 2000.

[13] Guo Yinghua and Yuan Cong. An Efficient Control Protocol for Dynamic Lightpath Establishment in Wavelength-Routed WDM Networks. In *Proceeding of SPIE APOC2003*, volume 5281, 2003.

[14] Guo Yinghua. Lightpath Management in Wavelength-Routed WDM Optical Networks. In *8th International Conference on CIC.*, Korea, October 2003.

[15] Guo Yinghua. Distributed Connection Control Protocols in All-optical Networks. In *5th International Workshop on Distributed Computing (IWDC2003)*, India, December 2003.

[16] M. Peyravian and A.D. Kshemkalyani. Connection preemption: issues, algorithms, and a simulation study. In *Proceedings of INFOCOM*, volume 1, pages 143 –151, 1997.

[17] J.C. de Oliveira, C. Scoglio, I.F. Akyildiz, and G. Ulh. A new preemption policy for diffserv-aware traffic engineering to minimize rerouting. In *Proceedings of INFOCOM*, pages 695 –704, 2002.

[18] A. Bosco, A. Botta, M. Intermite, P. Iovanna, and S. Salsano. Distributed Implementation of a Pre-Emption Mechanism for a Network Control Based on IP/MPLS Paradigm. In *Proceedings of the 7th IFIP ONDM*, Budapest, Hungary, February 2003.

[19] I. Chlamtac, A. Ganz, and G. Karm. Lightpath communications: An approach to high-bandwidth optical WAN's. *IEEE Transactions on Communication*, 40:1171–1182, 1992.

[20] X. Yuan, R. Melhem, and R. Gupta. Distributed Control Protocols for Wavelength Reservation and Their Performance Evaluation. *Photonic Network Communications*, 1(3):207–218, 1999.

[21] R. Ramaswami and A. Segall. Distributed Network Control for Optical Network. *IEEE/ACM Transactions on Networking*, 5(6):936–943, 1997.

[22] H. Zang, L. Sahasrabuddhe, J.P. Jue, S. Ramamurthy, and B. Mukherjee. Connection Management for Wavelength-Routed WDM Networks. In *Proceedings, IEEE Globecom*, volume 2, pages 1428–1432, Rio de Janeiro, Brazil, Dec. 1999.

[23] Jun Zheng and H.T Mouftah. Distributed lightpath control based on destination routing for wavelength-routed WDM networks. In *Proceedings, IEEE Globecom* , volume 3, pages 1526–1530, 2001.

[24] K. Kar, M. Kodialam, and T.V. Lakshman. Minimum Interference Routing of Bandwidth Guaranteed Tunnels with MPLS Traffic Engineering Applications. *IEEE Journal on Selected Areas in Communications*, 18(12):2566 –2579, December 2000.

[25] A. Capone, L. Fratta, and F. Martignon. Virtual Flow Deviation: Dynamic Routing of Bandwidth Guaranteed Connections. In *Proceedings of IEEE Workshop on Quality of Service in Multiservice IP Networks*, pages 608–620, Milan, Italy, February 2003.

[26] H. Zang, J.P.Jue, and L. Sahasrabuddhe. Dynamic Lightpath Establishment in Wavelength-Routed WDM Networks. *IEEE Communication Magazine*, 39(9):100–108, September 2001.

[27] C. Casetti, G. Mardente, R. Lo Cigno, M. Mellia, and M. Munafo. On-line routing optimization for MPLS-based IP networks. In *IEEE Workshop on High Performance Switching and Routing (HPSR)*, pages 215–220, Torino, Italy, June 2003.

[28] C. Casetti, R. Lo Cigno, M. Mellia, and M. Munafo. A new class of QoS routing strategies based on network graph reduction. In *Proceedings of INFOCOM*, 2002.