uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Abdelilah Maach
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

Ph.D. (Computer Science)
GRADE / DEGREE

School of Information Technology and Engineering
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Contention reduction and reliability enhancement for optical burst switching

TITRE DE LA THÈSE / TITLE OF THESIS

Gregor Bochmann
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

Hussein Mouftah
CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

Chadi Assi

Trevor Hall

Chung-Horng Lung

Hussein Mouftah

Gary W. Slater
LE DOYEN DE LA FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTDOCTORALES /
DEAN OF THE FACULTY OF GRADUATE AND POSTDOCORAL STUDIES

Contention reduction and reliability enhancement for optical burst switching

By

Abdelilah Maach
amaach@site.uottawa.ca

A thesis submitted to the Faculty of Graduate and Post-Doctoral Studies in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Computer Science

Ottawa-Carleton Institute for Computer Science
School of Information Technology and Engineering
University of Ottawa
Ottawa, Ontario, Canada

Your file Votre référence
ISBN: 0-494-10989-0
Our file Notre référence
ISBN: 0-494-10989-0

Abstract

The rapid growth of the Internet and the advent of WDM in the second generation of optical network systems have led to the extensive use of optical resources available for switching and routing. Such growth has boosted the research activities that focus on the most efficient techniques to make better use of the enormous speed and bandwidth of all-optical networks. Following the trend of carrying IP traffic over WDM, many research initiatives started under two major strategies: Wavelength Routing and Optical Burst Switching.

Wavelength Routing is a technique that allows the establishment of direct end-to-end light channels between two nodes, known as light-paths. Whereas Optical Burst Switching is a forwarding technique employed with a transparent optical backbone aiming to keep a large part of the information in the optical domain, and reduce the opto-electronic conversion overhead.

Optical Burst Switching is an attractive hybrid approach between coarse-grain optical circuit switching and fine-grain optical packet switching. However, a major concern for OBS networks is contention on outgoing data channels, which can result in burst loss. Therefore, contention resolution is necessary in order to handle the case where more than one burst are destined to go out of the same output port at the same time.

This work explores the contention problem that occurs in optical burst switching. In order to control the contention and pave the way to optical burst switching to be more suitable for Internet traffic, we propose to use OBS with segmentation where a burst is broken into many small (in size) segments. In case of contention we remove only the segments involved in the contention. We also propose a hybrid architecture where part of the wavelengths is used for OBS and the others are used for wavelength routing technique. The edge node can use both of them and send the data according to the class of traffic.

Another solution consists of controlling the network load. Indeed, under heavy traffic, the loss in OBS increases. Therefore one needs to control the traffic and keep the network away from congestion. Here we propose a protocol that reports the losses in the network to the edge nodes, so they can adjust the traffic accordingly. We propose also another protocol to retransmit the lost bursts.

Although these propositions could improve OBS and keep the loss under control, the loss is still there, and some applications may not tolerate it. In a second part of the thesis we propose a loss free architecture. This scheme uses time division multiplexing with reservation. A flow of slots is established between a source and a destination on request of the source. This flow is used to send bursts without any loss.

In the last part of the thesis we propose to enhance the network reliability by protecting the flows carrying the traffic inside the network. The shared risk is analyzed and the optimal protection capacity is determined. This will lead to a robust network where no loss is observed whether due to the contention problem or to the network failure.

Acknowledgments

I would like to express my deep and sincere gratitude to my supervisor, Dr. Gregor von Bochmann. His wide knowledge and his logical way of thinking have been of great value for me. His detailed and constructive comments and his important support have provided a good basis for the present thesis.

I am deeply grateful to my co-supervisor, Dr Hussein Mouftah, for his understanding, encouragement and for his personal guidance throughout this work.

I warmly thank Dr Trevor Hall for his valuable advice. His extensive discussions in our group meetings have been very helpful for this study.

My sincere thanks are due to the committee members, for their review,

During this work I have collaborated with many colleagues for whom I have great regard, and I wish to extend my warmest thanks to all those who have helped me with my work

Finally, I would like to express my deepest gratitude for the constant support, and understanding I received from my mother, my family and all my friends.

Table of Content

Abstract ii
Acknowledgments iii
Table of Content iv
List of Figures vi
List of Tables viii

1 Introduction
1.1. Background 1
1.2. Motivation and objective 3
1.3. Thesis contribution 5
1.4. Thesis outline 6

2 Techniques for optical network engineering
2.1. Network architecture and signaling protocols 8
2.2. Optical burst switching 13
2.3. Protection and restoration 18
2.3.1 MPLS Fault management 22
2.3.2 Protection and restoration in optical burst switching 31

3 Segmented optical burst switching
3.1. Introduction 36
3.2. Segmented optical burst switching 37
3.3. Performance evaluation 40
3.4. Summary 47

4 A hybrid architecture using optical burst switching and routed wavelengths
4.1. Introduction 49
4.2. Wavelength routing network architecture 49
4.3. Shortcomings of existing approaches 54
4.4. Hybrid architecture 55
4.5. Simulation results 58
4.6. Summary 60

5 Contention avoidance using congestion control
5.1. Introduction 62
5.2. Congestion avoidance in optical burst switching 63
5.3. Burst retransmission approach 69
5.4. Analytical model in star network 71
5.5. Simulation results and analysis 73
5.6. Summary 78

6 A Bandwidth allocation scheme in optical TDM network
6.1. Introduction ... 80
6.2. Resource allocation and switch architecture ... 82
6.2.1. Switch architecture ... 84
6.2.2. Flow reservation ... 88
6.2.3. Basic scheduling ... 89
6.3. Optical wavelength conversion and optical slot delaying ... 92
6.4. Simulation result and analysis ... 94
6.5. Summary ... 99

6 Shared protection
7.1. Introduction ... 101
7.2. Network architecture and shared protection scheme ... 104
7.2.1. Shared protection ... 104
7.2.2. Single link failure protection ... 108
7.2.3. Single node failure protection ... 111
7.2.4. Multiple failures protection ... 114
7.2.5. A recovery from a failure ... 116
7.3. Cases studies ... 117
7.3.1. Optical network deploying time division multiplexing ... 117
7.3.2. Star network ... 118
7.4. Simulation result and analysis ... 122
7.5. Summary ... 129

7 Conclusion
8.1. Summary and conclusive remarks ... 130
8.2. Future research ... 133

References ... **136**

List of Figures

Figure 2.1: The optical network evolution 8
Figure 2.2: An optical network with IP edge routers 9
Figure 2.3: Optical burst with offset between the header and the burst 16
Figure 2.4: Optical burst without offset between the header and the burst 16

Figure 3.1: Contention in optical segmented burst switching 38
Figure 3.2: Simulated Switch 41
Figure 3.3: Switch with light traffic 43
Figure 3.4: Switch with high traffic 44
Figure 3.5: Switch with light traffic and 2 fibers in each output 45

Figure 4.1: Virtual topology 50
Figure 4.2: Hybrid architecture 56
Figure 4.3: Loss probability for OBS and the hybrid architecture 59
Figure 4.4: Average delay (ms) of OBS, wavelength routing and the hybrid architecture 60

Figure 5.1: Performance as function of traffic load 64
Figure 5.2: A star optical network with retransmission scheme 72
Figure 5.3: A star network with 8 edge nodes 73
Figure 5.4: Loss rate as function of load 74
Figure 5.5: Loss rate as function of load with and without congestion control 75
Figure 5.6: Average number of transmissions 76
Figure 5.7: Distribution of the number of retransmissions 77
Figure 5.8: Average delay per burst 77

Figure 6.1: Switch Architecture 85
Figure 6.2: Sequencer 87
Figure 6.3: Alternative Switch Architecture 88
Figure 6.4: Scheduling Example 92
Figure 6.5: NSFNET topology with 14 nodes 95
Figure 6.6: Blocking rate of TSRS, TSR and WR 96
Figure 6.7: Average number of paths for TSRS, TSR and WR 97
Figure 6.8: Average number of hop versus the load for TSRS, TSR and WR 97
Figure 6.9: Resource used by TSRS, TSR and WR 98

Figure 7.1: A primary traffic (continued lines) with its protection (dashed lines) 107
Figure 7.2: A network with 4 groups 115

Figure 7.3 : An overlay star network 119
Figure 7.4: A flow of time slots between two nodes 119
Figure 7.5: A flow of time slots between two nodes with a protection 120
Figure 7.6: NSFNET topology with 14 nodes 123
Figure 7.7: Protection efficiency for SPFS 124
Figure 7.8: Protection efficiency for DFS 125
Figure 7.9: Blocking rate for DFS 126
Figure 7.10: Protection blocking rate for DFS 127
Figure 7.11: Ratio of the traffic protected for DFS 128

List of Tables

Table 2.1: Protection scheme comparison 19

Table 3.1: A comparison between the 5 optical switching methods..... 40
Table 3.2: Dropping probability of the last segment with low load 46
Table 3.3: Dropping probability of the whole burst and the last segment with high load47

Table 5.1: admission control 67

Chapter 1

Introduction

1.1.Background

Optical networks are high-capacity telecommunication systems based on photonic technologies, and consist of the interconnection of many optical elements [1] such as transmission links, amplifiers, cross-connects, wavelength-add/drops, multiplexers and other devices. The novel idea of this kind of networks is to keep the information in the optical domain as long as possible.

The extensive bandwidth of the optical fiber can be used more efficiently by using the dense wavelength division multiplexing (DWDM) technology which consists of the use of several light-paths in the same fiber [2,3,4,5]. This increases tremendously the capacity of the fiber, and hence provides a good solution to face the increasing demand for bandwidth.

The Optical Technology is providing more equipment and components that process and handle optical information. The real challenge is using the whole capacity and building an optical carrier that shares the resources efficiently between all users while providing good quality of service. Basically the trend in optical networks is to carry IP traffic directly over DWDM. The efforts towards this goal have resulted in two kinds of strategies:

- The development of optical burst switching (OBS) [6,7,8,9,10] as a forwarding technique to build a transparent optical backbone. This technique aims to keep a large part of the information in the optical domain and hence, reduce the burden of opto-electrical conversion. However, this technique suffers from a lack of intelligence inside the network resulting in possible loss of bursts as a result of contention. Indeed, whenever two or more bursts compete for the same output at the same time, only the first is sent and the others are dropped.
- The development of wavelength routed networks [11,12,13,14]. With developments being made now in WDM which provides more wavelengths in the same fiber, and the capability of cross-connects to switch a wavelength separately, it is possible to establish a direct light-path between two edge nodes. This pipe of light can be seen as a separate link operating autonomously without interaction with any other links or awareness of any crossed components. This path, also called a virtual link, can carry information transparently in one hop from a source to a destination without the need for conversion to the electrical domain, and hence avoiding the bottleneck that may appear in the intermediate nodes (because of the store and forwarding process). Even better, this technique can be used to build a new topology by setting up many paths inside the network. This new architecture aims to build a virtual topology [11,12,15], completely different from the physical one, and hence reduce the complexity related to optical technology, especially the opto-electrical conversion and optical buffering. The presence of such a layer enhances the transport network and bridges the gap between the conventional and optical networks by providing a flexible high-capacity interconnection; indeed the reconfiguration provides a new topology, which can be

used with well-known protocols and principles. The virtual topology can be completely different from the underlying physical topology.

1.2.Motivation and objective

Used separately, neither OBS nor wavelength routed networks are suitable for all classes of traffic or to guarantee the quality of services needed for all applications. Each method has its own limitations and suffers from many drawbacks. With OBS, the burst loss due to contention, which is inherent to this technique, decreases the performance in terms of throughput and delivery delay, especially with high loads. Several methods have been proposed in the literature to decrease the loss ratio. Some of these techniques are realized purely in software, such as deflection routing [16,17,18,19] and segmented bursts [20,21,22,23] where the others require specific hardware such as burst buffering [24,25,26] or wavelength conversion [27,28,29,30]. These methods may reduce the contention, but they all remain sensitive to the traffic load. Indeed, according to [6], it is clear that even in an ideal network, where the switches use a number of buffers and can perform wavelength conversion, contention still occurs when the load gets heavier. This means that the delay and the delivery are not guaranteed and, hence, this type of network may be useless for some kinds of applications, especially real time and delay sensitive applications. On the other hand, wavelength routed networks have seen some improvement especially in wavelength assignment algorithms. Unfortunately, this technique is still facing many problems; namely:

- The complexity of the wavelength assignment algorithm increases with the network size and the number of wavelengths, which can hinder the future expansion of the network.

- Due to resources limitations, it is sometimes impossible to establish a direct light-path between all the edge nodes. Therefore intermediate nodes must be used as a tandem, which can lead to additional delay and routing complexity
- Even when a light-path is established between two edge nodes, it is not necessarily the shortest light-path.
- Edge nodes may not have enough loads to fill the capacity of the established path, and hence a part of the bandwidth may remain unused.
- To establish a new light-path, the manager may need a long time to analyze the resources available.
- Some of the wavelengths are left unused on some links.

Finally, wavelength-routed optical burst switching technique appears as an intermediate solution between wavelength routed and optical burst switching and aims to overcome some limitation of both techniques, especially the contention losses of OBS. However, many problems arise with this method. The following are some of them:

- Even if the loss has disappeared in the network, the edge nodes drop the burst because of the shortage of resources.
- The centralized solution (reservation done through the network control node) could be a handicap to this solution since it incurs additional traffic and the service time will increase with the propagation delay (especially in case the network is covering a large geographical area).
- This technique has an overhead similar to that of circuit switching, and it is justified only if the session is long enough, which is not the case for bursts.

Another concern that arises with these two techniques is reliability and their capacity to survive physical network failure. Without strong defense against failure, a potential problem could be fatal. Indeed, high speed and high connectivity networks carrying a large amount of information are very vulnerable to any kind of failure inside the network (link or node failure); in case of link failure, very large amounts of traffic can be affected leading to delay and disruption in the network. One should enhance the network by sophisticated network survivability features. In this context network survivability [31,32,33,34,35] becomes an important issue in order to build an intelligent network that can face these failure situations efficiently.

1.3. Thesis contribution

In this work I focus on optical burst switching as a technique to build an optical backbone network. In this context, the network will be considered as a distributed switch where the information is exchanged between the edge routers, which are considered as the switch ports. This work aims to enhance the network reliability by decreasing the loss due to contention and by enforcing survivability against network failures.

This work explores the contention problem that occurs in optical burst switching. In order to control the contention and pave the way for optical burst switching to be more suitable for Internet traffic, we propose three steps; the first step aims to reduce the loss and keep it under an acceptable level. This step consists of the following techniques:

- Segmented OBS: where a burst is broken into many small (in size) segments; in case of contention we remove only the segments involved in the contention. Simulation proves that this approach improves the network utilization and decreases the loss.

- Hybrid architecture: in this architecture, some wavelengths are used for OBS and the others are used for wavelength routed technique. The edge node can use both of them and send the data according to the class of traffic. This architecture increases the network connectivity.
- Congestion control: under heavy traffic, the loss of bursts in OBS increases. Therefore one needs to control the traffic and keeps the network away from congestion. Here we implement a protocol that reports the state of the network to the edge nodes so they can adjust the traffic accordingly. We also propose another protocol to retransmit the lost burst.

These propositions can improve OBS and keep the loss under control, nevertheless the loss is still there, and some applications may not tolerate this loss. In the second step we propose a loss-free architecture. This scheme uses time division multiplexing with reservation. A flow of slots is established between a source and a destination on request of the source. This flow is used to send bursts without any loss.

In the last step, we enhance the network reliability, by protecting the flows carrying the traffic inside the network. This will lead to a robust network where no loss is observed whether due to the contention problem or to the network failure.

1.4.Thesis outline

Chapter 2 gives an overview of traffic engineering techniques; optical burst switching and protection/restoration issues. Chapter 3 gives more details about segmented optical burst switching. Chapter 4 proposes a hybrid architecture which is a combination of optical burst switching and wavelength routed. Chapter 5 presents a protocol that informs the edge nodes about the loss inside the network, so that they can adjust their traffic and

also retransmit the dropped bursts. Chapter 6 presents the time division multiplexing scheme with reservation. Chapter 7 proposes a protection scheme that could be used to enhance the survivability of the network. Chapter 8 provides the summary, conclusive remarks and proposals for future research directions founded on this thesis.

Chapter 2

Techniques for optical network engineering

2.1.Network architecture and signaling protocols

In its first stage, optical networks consisted of electronic switches connected by optical fibers. At each intermediate node, the data has to be converted to the electronic domain and stored temporarily while the header is being processed. Unfortunately, the electronic processing and conversions limit the speed of the network, and do not exploit all the advantages of DWDM. Furthermore, the network cost was very high due to the components used to perform the opto-electronic conversions and storage.

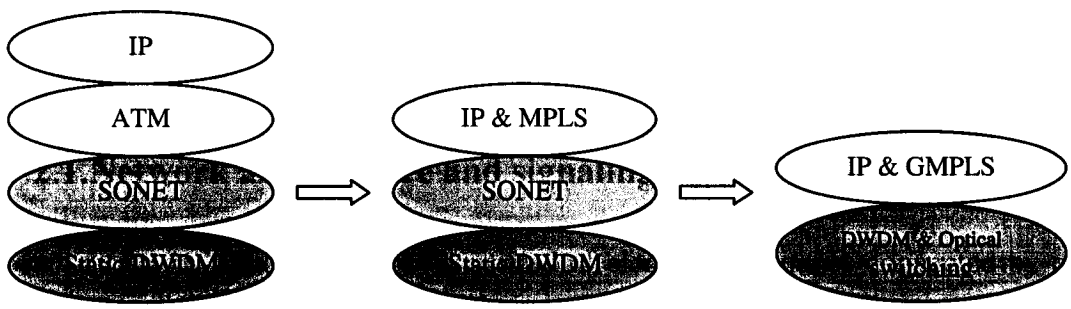


Figure 2.1: The optical network evolution

DWDM has been statically employed, as shown in Figure 2.1, to support IP over Asynchronous time multiplexing (ATM) [36,37] over synchronous optical network (SONET) [38,39]. In this architecture with 4 layers, DWDM is deployed as a capacity

provider (point-to-point links), SONET as a reliable transport layer, ATM as a traffic engineering and IP as an intelligent packet forwarder. Furthermore IP has been enhanced and empowered by the high forwarding capacity of MPLS [40,41,42]. However, this architecture faced many challenges, especially those related to the cost of such a solution and its capacity to scale to very large volumes of traffic. The next step, as shown in Figure 2.1, is to carry IP traffic over SONET to give the network more flexibility and subtract the overhead incurred by ATM. A special SONET layer called "thin SONET" was developed to be used with DWDM. In these architectures, the network protection/restoration was trivial. The survivability was strongly enhanced by SONET.

In order to provide a viable solution that meets the new requirements (high speed, high transparency, low cost), DWDM technologies should combine broadband switching and optical transport technology to carry IP traffic (without intermediate layers) in order to overcome the problems of stranded bandwidth and high growth costs of existing transport solutions. This architecture requires that the functions performed by SONET and ATM must move to IP with MPLS and DWDM levels (as layers).

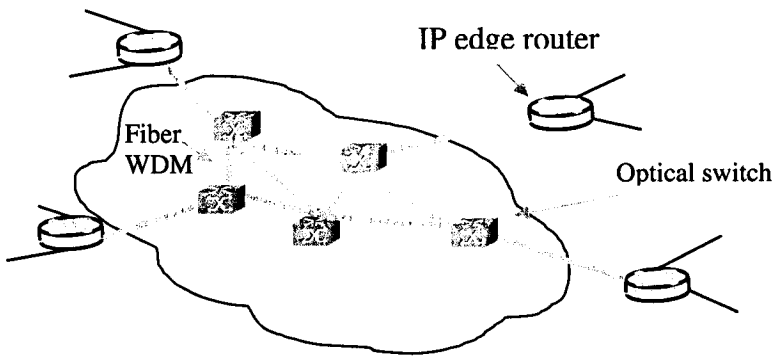


Figure 2.2: An optical network with IP edge routers

Figure 2.2 shows an IP network operating over an optical backbone. So far all-optical networks are not available, but the progress made in optical technology, especially in optical components such as add/drop modules and optical cross connects, makes it possible to have what is called an almost-all optical network [4]; in this kind of networks, an intermediate node does not need to convert the whole information to the electrical domain and store it. Instead, it converts only the part that carries pertinent information. That means, only a small part of the traffic is converted to the electrical domain, whereas the larger amount remains in the optical domain. Several architectures and methods have been proposed to carry information using this kind of networks; among them we find wavelength routed networks [11,12], which consist of a logical topology built over the physical one by the assignment and allocation of wavelengths. This configuration is performed dynamically [13] to meet the traffic load. In general, the goal of reconfiguration is to minimize the average number of hops to reach the destination. Another method, very similar to fast circuit switching, consists of separating the control information and the data. The control information (burst header) is sent over a dedicated wavelength followed by the data using a different wavelength. The burst header is converted and processed at all the nodes crossed, however, the information will cut-through a cross-connect transparently. This technique is called optical burst switching.

Optical burst switching offers fine granularity between circuit switching and packet switching. In a circuit-switched network, a dedicated circuit must first be connected. Once the circuit has been physically established, transmission can begin. When the transmission is complete, the circuit is released for the next transmission. The establishment and tearing down of circuits uses a relatively simple signaling protocol that

assumes global knowledge of the network resources. In packet switching there are no dedicated circuits. Each circuit carries many different transmissions at the same time. The only rule is that every data unit sent through a packet-switching network must have enough information in the header so that the nodes in the network can determine how to route the data unit.

OBS could use the control plane of multi-protocol label switching (MPLS) to establish virtual paths and handle signaling traffic. The deployment of MPLS will bridge the optical layer to the IP layer, in order to allow interoperable and scalable parallel growth in the IP and photonic dimensions. The data (or burst) is carried in the data plane and forwarded according to the label carried in the control plane. These labels are assigned at the ingress router and are used to identify the forwarding equivalence class (FEC). Consequently, the data (or burst) belonging to the same FEC will be forwarded over the same path through a network; these paths are also called label-switched paths (LSPs).

The MPLS control plane uses standard protocols to exchange information between the different routers of a network to build and maintain a forwarding table. This plane is completely separated from the data plane. Therefore they can be independently developed and modified. They continue to communicate through the forwarding table which is updated by the control plane and used by the data plane.

MPLS components include signaling protocols for setting up the LSP and routing protocols (Open Short Path First OSPF or Intermediate System- Intermediate System IS-IS) with appropriate extensions for advertising the network topology and available link resources.

There are two signaling protocol families proposed for LSP establishment; (1) Label Distribution Protocol (LDP) [43] and its extension, Constraint-based Routing-Label Distribution Protocol (CR-LDP) [44] and (2) Traffic Engineering extensions for Resource ReSerVation Protocol (TE-RSVP) [44].

MPLS supports devices that can perform packet and maybe burst switching only. However, to support those performing switching in time, wavelength and space, an extension has been proposed. That is Generalized Multi-protocol Label Switching (GMPLS) [45,46, 47]. It defines four networking layers:

- Packet Switching capable, which can handle packets, cells or bursts (e.g. IP, ATM, and OBS).
- Time Division multiplexing layer capable, which can handle larger TDM frames.
- λ-switching capable layer, which can handle the light-path establishment.
- Fiber switching capable layer, which handles fibers with all traffic over them.

GMPLS helps to build multiple layers one over the other and define standard interfaces, protocols and signaling to set up connection on demand with required traffic and quality parameters.

The performance and the quality of service delivered may depend on the architecture deployed and the network management. However, the class of traffic being carried on the network could be very decisive when it comes to a real success. Thus a design of any network should take the traffic parameters into consideration. Since some networks could be very suitable for a given class of traffic and not suitable for others.

In this context we consider the following classes of traffic:

- Delay sensitive: This class of applications includes real time services, such as streaming media or interactive multimedia, as well as data services requiring low latency. Multimedia applications with real-time properties require tight guarantees in terms of packet delay and packet loss. Obviously, interactive multimedia communication can only be realized with upper bounded end-to-end delays. However, some of the real-time applications may tolerate some losses as long as this loss remains below an acceptable rate (for example video less than 1%).

- Loss sensitive: this class of traffic includes applications, such as file transfer and data transmission. The loss here is not acceptable and high carrier reliability is required. However, some applications in this class may tolerate an additional delay. The major concern is the loss and the performance metrics for this class are focusing on the loss rate. Loss no sensitive applications are those that can tolerate some loss.

- Another class could be both loss and delay sensitive, which requires high control of both the loss and the delay.

The idea of classifying and prioritizing traffic goes hand in hand with the notion of network consolidation, which is used to carry many different kind of traffic (voice, data and video). So it is important for network architects to decide what their traffic classes should be and how they will be signaled at a technical level. In this work I assume that the edge nodes are able to identify the class of all incoming traffic.

2.2.Optical burst switching

Optical burst switching (OBS) is a technique for transmitting information by setting up the switches and reserving resources only during the time the burst is crossing the network. In OBS, the data enters the optical cloud via an edge router where it is

converted to an optical burst to be sent to the core network. The principle is similar to the one used in conventional packet switching, however the information is separated into two parts; a header and a payload. The main goal of this separation is to minimize the opto-electrical conversion and avoid the limitation incurred by the electronic technologies such as the processing time and conversion. The header is converted (at every intermediate node) to the electrical domain, where it is processed and converted back to the optical domain. The payload is simply forwarded in the optical domain according to the information transported by the header. In this technique, the concept of the packet is replaced by a burst; this constitutes an interesting step towards an all-optical network where the largest part of the information remains in the optical domain.

In an optical network using the optical burst switching technique, the edge nodes are able to store and process IP packets whereas the intermediate nodes will perform forwarding of burst according to the destination. In this architecture, the incoming packets are buffered in the source edge routers to form bursts. Bursts are collected according to their destination and class of service. Then, a control packet is sent over a specific optical wavelength channel to announce an upcoming burst. The control packet, also called an optical burst header (OBH), is then followed by a burst of data, over another optical channel without waiting for any confirmation. The OBH is converted to the electrical domain at every node in order to be interpreted and transformed according to the routing decision taken at each switching node. Also, pertinent information is extracted such as the wavelength used by the following data burst, the time it is expected to arrive, the length of the burst and the label, which determines the destination. This

information will be used by the switch to schedule and set-up the cross-connect for the coming data burst.

There are several approaches to practically implement OBS. The main difference is related to the timing issues concerned with synchronization between the data bursts and their headers. These techniques fall into two categories:

OBS with offset: In this category, the data burst is delayed at the source and sent after a short delay as shown in Figure 2.3. This offset must be long enough to cover the processing time at all intermediate nodes. This assumes that the source knows the number of hops needed to reach the destination and the processing time at each node. The offset can be formulated as a function of processing and switching time at each node and the number of nodes crossed by the burst as follows. Let Tp be the processing time of the optical header at each of the nodes and Ts the switching time at each node; N is the number of nodes crossed by the burst, the offset for that burst must be greater than N*Tp + Ts with the assumption that the processing time and the switching time is the same at all the nodes. This technique is called in the literature "just enough time" [48,49,50]. The advantage of this class of methods is that no buffers are needed at the intermediate nodes. Besides, this offset can be used to define several classes of service [51,52,53,54,55]; indeed an extra offset is added to the burst to give it higher priority, the longer the offset, the higher its priority. The switch has enough time to reserve the bandwidth for a burst with longer offset.

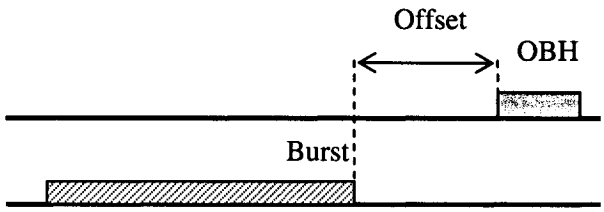


Figure 2.3: Optical burst with offset between the header and the burst

This technique is very similar to fast circuit switching where no acknowledgement from the network is needed prior to sending the data. Thus the pre-transmission delay of the burst is reduced.

OBS without offset: In this category, the data burst is sent at the same time as the OBH, as shown in Figure 2.4, and delayed at the switch by a fiber delay line (FDL). The FDL in this case is different from a buffer since we do not need to store the burst. The delay must be long enough to cover the header processing time and the switching time. The feasibility of this technique depends upon switching speed. Indeed, the length of the FDL used to delay the burst, can range from a few hundreds meters for fast switches to hundreds of kilometers for slow switches. This method has the advantage of freeing the edge from the burden of calculating the offset. This method can also be used with routing methods where the path is determined hop by hop.

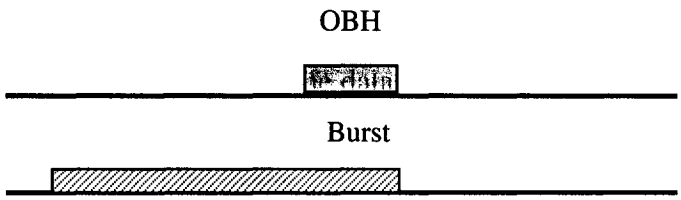


Figure 2.4: Optical burst without offset between the header and the burst

This technique is very similar to packet switching using store and forward. However, the burst can be stored (delayed) only for a fixed and very short time.

From another point of view, OBS could be seen as an architecture using asynchronous light-path establishment for a very short duration using one way reservation. Indeed, in this scheme, the construction of a light-path starts slightly before the burst is sent and continues as the burst crosses the network. At the same time and as soon as the burst leaves the source, the operation of tearing down starts. This technique may use three variations for releasing bandwidth:

- Tell-and-go (TAG): as soon as the burst is transmitted, the sender sends an explicit release message to tear down the circuit (like circuit switching).
- Reserve-a-fixed-duration (RFD): each set-up request specifies the duration for the circuit.
- In-band-terminator (IBT): a burst contains a header and a tail (terminator) to indicate the end of the burst (like packet-switching). However, this technique is hard to use because it is difficult to recognize the end of the burst.

The OBH is processed electronically and it may be processed optically in the future, but whether in optic or in electronic form, the most important feature is the switching time since this switching time overhead is considered a waste of bandwidth. Therefore the switching time is critical for the success of OBS.

Optical burst switching combines the intelligence and the capacity of processing and buffering of the electronic edges with the capacity of optical forwarding of the core. Thereby the network can meet the demands of increasing traffic volumes and accommodate different kinds of information. Basically, OBS is designed to avoid the long end-to-end setup times of conventional virtual circuit configuration with no need for memory at intermediate nodes. However, the main problem is contention, that may occur

when one or more bursts arrive at the same time and try to leave over the same output fiber on the same wavelength. The way this contention is resolved can affect tremendously the performance of the network. Indeed, when the packet-loss ratio increases, the efficiency as well as the delivery rate decreases and hence the throughput. In other words the contention makes the delivery not guaranteed and not deterministic.

Papers [56,57,58] propose another network architecture combining OBS with dynamic wavelength routing. This technique is called wavelength-routed optical burst switching (WR-OBS) because of its capability of switching light-paths on burst by burst basis. Unlike OBS, WR-OBS requires a two-way reservation scheme, where the source should receive an acknowledgement prior to the burst release. It relies on an end to end reservation. A burst is aggregated at the edge router and a request for a path is sent to the network in order to assign one wavelength. This technique assumes that during the aggregation of different packets into a burst, a reservation of resources along a path is made. No burst is sent without wavelength reservation and hence no loss is observed inside the network. This guarantees the delivery and controls the delay. However, when the network control fails to find a wavelength or the waiting time at the edge node exceeds some limit, the burst is dropped to avoid a source-destination time out.

2.3.Protection and restoration

WDM networks are prone to failures of components, such as links and nodes. Failures may have severe consequences due to the high traffic volume carried by these networks. Thus, it is imperative that these networks support schemes to protect or restore the failed light-paths. A number of protection/restoration schemes [31,32,33,34,35] have been

proposed in the literature. These schemes can be classified into two classes: protection schemes and network reconfiguration/restoration schemes.

In a protection scheme, backup (also called protection) routes are computed, and resources are reserved along the backup light-paths at the time of the primary light-path setup. Upon occurrence of a failure in the primary path, traffic is immediately switched from the primary routes to the backup routes. Thus, the recovery is fast and guaranteed (assuming that primary and backup paths do not fail at the same time). Also, the protection techniques can be distinguished in two cases depending on the level of resource reservations: dedicated or shared schemes. In the dedicated case the protection paths cannot share the resources among themselves. They are typically referred to in the literature as 1+1 in the case that the protection path is occupied by the traffic itself (the same traffic is sent twice) or 1:1 in case that the protection path is used by other traffic that can be preempted in case of failure). In shared schemes, several disjoint working paths can have the same protection paths (1:n or m:n).

Each of these protection schemes has its own advantages and disadvantages, and hence causing a trade-off (resources required and restoration speed), which has to be taken into consideration when using and designing protection schemes. Table 2 1 summarizes these differences.

Protection Scheme	*Cost/complexity of OXC, operation & Management systems*	*Restoration speed*	*Wavelength resources required*	*Pre-emptable traffic that can be carried at time of failure*
1+1	Lowest	Highest	Highest	None
1:1	Medium	Medium	Medium	Medium
1:N, m:n	Highest	Lowest	Less than 1:1	Lowest

Table 2.1: Protection scheme comparison

- **Cost of OXC**

In 1+1, the OXC are less complex in design, to operate and maintain. Also, the algorithms to perform such a protection are simpler. The OXC's in such a protection scheme are least expensive. The 1:1, 1:N and m:n are progressively more expensive and complex. 1:N, m:n being the most expensive [33].

- **Restoration speed**

The restoration speed of 1+1 protection is highest, the destination nodes just have to tune to another receiver. The restoration speeds of 1:1, 1:N becoming progressively less, with 1:N protection scheme being the slowest. Also the Operation, Administration and Management systems (OAM) get more complex. The exact location of the fault has to be identified in such a protection mechanism. In addition, in shared protections mechanisms, the failed light-paths have to be identified and hence the protection switching mechanism is more complex and slower.

- **Wavelength resources required**

Due to the way each of these protection mechanisms are implemented (and sharing occurs/does not occur), the wavelength resources needed for 1+1 will typically be the maximum. Similarly it can be argued the wavelength resources needed for 1:N, m:n will be the least [33].

- **Pre-emptable traffic that can be carried at time of failure**

Again due to the way the sharing occurs and these protection schemes are implemented, it can be verified that typically 1+1 will not allow any pre-emptable traffic to be carried in the network. In a 1:1 scheme, one expects that some amount of pre-emptable traffic can be carried and the amount of such traffic which will still survive on the network (not

be thrown away) on the average, when all links are broken one at a time, will be higher than the 1:N, m:n case (as these schemes allow sharing/multiplexing of protection resources). This is due to the way sharing occurs in these two shared protection schemes.

In a restoration scheme, when an existing path fails, a new path, which does not use the failed components, is computed on the fly; then, the traffic is switched from the failed path to the new path. Restoration schemes do not guarantee successful recovery of failed paths since the computation of new light-paths may fail because of a resource shortage. Nevertheless this scheme is very efficient in resource sharing. The major challenge of optical network restoration scheme design is deciding where and how much spare capacity is needed in order to accommodate the failed traffic easily and without too much resource waste [59,60, 61].

The routing principle of OBS is similar to the one used by Multi Protocol Label Switching (MPLS) [62] in the sense that both, OBS and MPLS, use a label to forward the data, the label edge routers (LERs) are the edge electronic routers and the label switching routers (LSRs) are replaced by optical cross-connects (OXCs). An OXC is a path switching element that establishes paths for optical channels by locally connecting an optical channel from an input port (fiber) to an output port (fiber) on the switch element. Thereby optical networks can take advantage and exploit recent advances in the MPLS control plane in terms of fault and traffic management. Nevertheless there are structural differences between LSRs and OXCs. Indeed, with the former the forwarding information is carried explicitly as part of the labels appended to data packets, while with the latter the switching information is sent separately on another wavelength. Besides OXCs do not perform packet level processing in the data plane, while the LSRs are

datagram devices which may perform certain packet level operations in the data plane. These differences may require some enhancement to adapt MPLS to the new environment especially to deal with the problems of quality of services and traffic engineering. Because of this similarity (OBS uses MPLS), in the following two sub-sections I will give a small overview of protection in both MPLS and OBS.

2.3.1 MPLS Fault management

a- Introduction

MPLS can be used to support advanced survivability requirements and enhance the reliability of IP networks [63,64]. Differing from classical IP networks, MPLS networks establish label switched paths (LSPs), similar to VP/VC-ATM. This allows MPLS networks to pre-establish protection LSPs backups for the working LSPs, and achieve better protection switching times than IP networks.

The usual method to develop an MPLS protected domain involves a working path and a recovery path (backup path). Not always is a backup LSP created at an ingress-node and finalized at an egress-node. A backup LSP could be implemented at a LSP segment. In this case the node where the backup originates, is called a PSL (Path switch LSR) and where the backup ends is called PML (Path Merge LSR).

To perform a protection within MPLS [65], one needs the following components:

1.- A method for selecting the working and the protection paths.

2.- A method for bandwidth reservation for the working and the protection paths.

3.- A method for signaling the setup of the working and protection paths.

4.- A fault detection mechanism to detect faults along a path.

5.- A fault notification mechanism, to convey information about the occurrence of a fault to a network entity responsible for reacting to the fault and taking appropriate corrective action.

6.- A switchover mechanism to move traffic over from the working path to the protection path.

7.- A repair detection mechanism, to detect that a fault along a path has been repaired.

8.- An (optional) switchback or restoration mechanism, for switching traffic back to the original working path, once it is discovered that the fault has been corrected or has been repaired.

The next section explains, in more detail, all the components involved in the MPLS fault control.

b- Definitions

The next section introduces many definitions to facilitate understanding of MPLS fault management components.

MPLS Protection Domain: The set of LSRs over which a working path and its corresponding protection path are routed. The protection domain is denoted as: (working path, protection path).

- **Fault and recovery signals/messages:**

Failure Indication Signal (FIS) : A signal that indicates that a failure has been detected at a peer LSR. It consists of a sequence of failure indication packets transmitted by a downstream LSR to an upstream LSR. It is relayed by each intermediate LSR to its upstream neighbor, until it reaches an LSR that is setup to perform a protection switch.

Failure Recovery Signal (FRS) : A signal which, indicates that a failure along the path of an LSP has been repaired. It consists of a sequence of recovery indication packets that are

transmitted by a downstream LSR to its upstream LSR. Again, like the failure indication signal, it is relayed by each intermediate LSR to its upstream neighbor, until is reaches the LSR that performed the original protection switch.

Liveness Message (LM) : A message exchanged periodically between two adjacent LSRs that serves as a link probing mechanism. It provides an integrity check of the forward and the backward directions of the link between the two LSRs as well as a check of neighbor aliveness.

Link Failure (LF) : A link failure is defined as the failure of the link probing mechanism, and is indicative of the failure of either the underlying physical link between adjacent LSRs or a neighbor LSR itself. (In the case of a bi-directional link implemented as two unidirectional links, it could mean that either one or both unidirectional links are damaged.)

Loss of Signal (LOS): A lower layer impairment that occurs when a signal is not detected at an interface. This may be communicated to the MPLS layer by the lower layer.

Loss of Packet (LOP) : An MPLS layer impairment that is local to the LSR and consists of excessive discarding of packets at an interface, either due to label mismatch or due to time-to –live (TTL) errors.

- **MPLS protection components:**

Working or Active LSP : An LSP established to carry traffic from a source LSR to a destination LSR under normal conditions, that is, in the absence of failures. In other words, a working LSP is an LSP that contains streams that require protection.

Working or Active Path : The portion of a working LSP that requires protection. (A working path can be a segment of an LSP or a complete LSP) The working path is denoted by the sequence of LSRs that it traverses.

Protection Switch LSR (PSL) : An LSR that is the origin of both the working path and its corresponding protection path. Upon learning of a failure, either via the FIS or via its own detection mechanism, the protection switch LSR switches protected traffic from the working path to the corresponding backup path.

Protection Merge LSR (PML) : An LSR that terminates both a working path and its corresponding protection path, and either merges their traffic into a single outgoing LSP, or, if it is itself the destination, passes the traffic on to the higher layer protocols.

Intermediate LSR : An LSR on the working or protection path that is neither a PSL nor a PML.

MPLS Traffic Group (MTG) : A logical bundling of multiple, working LSPs, each of which is routed identically between a PSL and a PML. Thus, each LSP in a traffic group shares the same redundant routing between the PSL and the PML.

Protected MPLS Traffic Group (PMTG) : An MPLS traffic group that requires protection.

Protected MPLS Traffic Portion (PMTP) : The portion of the traffic on an individual LSP that requires protection. A single LSP may carry different classes of traffic, with different protection requirements. The protected portion of this traffic may be identified by its class, as for example, via the EXP bits in the MPLS shim header or via the priority bit in the ATM header.

Protection or Backup LSP (or Protection or Backup Path) : An LSP established to carry the traffic of a working path (or paths) following a failure on the working path (or on one of the working paths, if more than one exists) and a subsequent protection switch by the PSL. A protection LSP may protect either a segment of a working LSP (or a segment of a PMTG) or an entire working LSP (or PMTG). A protection path is denoted by the sequence of LSRs that it traverses.

- **Protection modes:**

Revertive : A switching option in which streams are automatically switched back from the protection path to the working path upon the restoration of the working path to a fault-free condition.

Non-revertive : A switching option in which streams are not automatically switched back from a protection path to its corresponding working path upon the restoration of the working path to a fault-free condition.

c- Protection types

Protection types for MPLS networks can be categorized as link protection, node protection, path protection, and segment protection.

Link Protection: The objective for link protection is to protect an LSP from a given link failure. Under link protection, the path of the protect or backup LSP (the secondary LSP) is disjoint from the path of the working or operational LSP at the particular link over which protection is required. When the protected link fails, traffic on the working LSP is switched over to protect the LSP at the head-end of the failed link. This is a local repair method that can be fast. It might be more appropriate in situations where some network elements along a given path are less reliable than others.

Node Protection: The objective of LSP node protection is to protect an LSP from a given node failure. Under node protection, the path of the protected LSP is disjoint from the path of the working LSP at the particular node to be protected. The secondary path is also disjoint from the primary path at all links associated with the node to be protected. When the node fails, traffic on the working LSP is switched over to the protect the LSP at the upstream LSR directly connected to the failed node.

Path Protection: The goal of LSP path protection is to protect an LSP from failure at any point along its route. Under path protection, the path of the protect LSP is completely disjoint from the path of the working LSP. The advantage of path protection is that the backup LSP protects the working LSP from all possible link and node failures along the path, except for failures that might occur at the ingress and egress LSRs, or for correlated failures that might impact both working and backup paths simultaneously. Additionally, because the path selection is end-to-end, path protection might be more efficient in terms of resource usage than link or node protection. However, path protection may be slower than link and node protection in general [33].

Segment Protection: An MPLS domain may be partitioned into multiple protection domains whereby a failure in a protection domain is rectified within that domain. In cases where an LSP traverses multiple protection domains, a protection mechanism within a domain only needs to protect the segment of the LSP that lies within the domain. Segment protection will generally be faster than path protection because recovery generally occurs closer to the fault.

d- m:n protection model

"m:n protection model" where m is the number of backup LSPs used to protect n working LSPs is one way to classify MPLS restoration models. Feasible protection models could be:

1:1: one working LSP is protected/restored by one protect LSP.

n:1: one working LSP is protected/restored by n protect LSPs, possibly with configurable load splitting ratio. When more than one protect LSP is used, it may be desirable to share the traffic across the protect LSPs when the working LSP fails to satisfy the bandwidth

requirement of the traffic trunk associated with the working LSP. This may be especially useful when it is not feasible to find one path that can satisfy the bandwidth requirement of the primary LSP.

1:n: one protection LSP is used to protect/restore n working LSPs.

1+1: traffic is sent concurrently on both the working LSP and the protect LSP. In this case, the egress LSR selects one of the two LSPs based on a local traffic integrity decision process, which compares the traffic received from both the working and the protect LSP and identifies discrepancies. It is unlikely that this option would be used extensively in IP networks due to its resource utilization inefficiency. However, if bandwidth becomes plentiful and cheap, then this option might become quite viable and attractive in IP networks.

Recovery Paths types with QoS requirements

Equivalent Recovery Path : Means that the recovery path preserves Working Path QoS requirements.

Limited Recovery Path : Does not preserve QoS requirements.

e- Network Survivability Layer Considerations

Best effort networks were focused primarily on connectivity. The re-routing fault management systems were enough to provide survivability. However, actual networks begin to support different classes of services (critical traffic, real-time traffic or high priority traffic), which means that slow re-routing schemes are not enough to achieve reliable fast services. The main drawback of level 3 re-routing algorithms is the amount of time that the algorithms take to converge and restore service. Actual networks need to provide highly reliable services, where the time needed to recover a failure might be of the order of milliseconds. In practice, fault restoration capabilities are implemented in

multiple protocol layers, such as automatic protection switching in the physical transmission layer, self-healing in the ATM virtual path layer, and fast rerouting in MPLS [66]. Usually, fault recovery is attempted first at the lowest layer, and then escalated to the next layer if recovery was unsuccessful or not possible.

To achieve fault management actual networks provide different schemes at different layers. At the bottom of the layered stack (optical networks) ring and mesh topology restoration functionality at the wavelength level, is provided. At the SONET/SDH layer survivability is provided at a link level in ring and mesh architectures. Similar functionality is provided by layer 2 technologies such as ATM (generally with slower mean restoration times).

Rerouting is traditionally used at the IP layer to restore service following link and node failures. Rerouting at the IP layer occurs after a period of routing convergence, which may require anything from seconds to minutes to complete.

MPLS allows new restoration mechanisms, with better performance than IP re-routing mechanisms. Recently, a common suite of control plane protocols has been proposed for both MPLS and optical transport networks under the acronym Generalized Multiprotocol label Switching (GMPLS) [67]. This new paradigm of Multiprotocol label Switching will support even more sophisticated mesh restoration capabilities at the optical layer for the emerging IP over WDM network architectures.

Developing a multi-layer survivability scheme involves providing restoration at different time scales (temporal granularity). Bandwidth granularity is another way of classifying protection mechanisms. Bandwidth granularity goes from the wavelength

level (optical level) to packet level (IP and higher layer protocols). Another vision of protection applicability is from the point of view of network services or traffic classes.

Protection and restoration coordination across layers may not always be feasible, because networks at different layers may belong to different administrative domains. Several points at which to minimize the impact of different layer protection disruption to achieve an efficient and complete protection scheme are according to [66]:

• Minimization of function duplication across layers is one way to achieve coordination. Escalation of alarms and other fault indicators from lower to higher layers may also be performed in a coordinated manner. A temporal order of restoration trigger timing at different layers is another way to coordinate multi-layer protection/restoration.

• Spare capacity at higher layers is often regarded as working traffic at lower layers. Placing protection/restoration functions in many layers may increase redundancy and robustness, but it should not result in significant and avoidable inefficiencies in network resource utilization.

• It is generally desirable to have protection and restoration schemes that are bandwidth efficient.

• Failure notification throughout the network should be timely and reliable.

• Alarms and other fault monitoring and reporting capabilities should be provided at appropriate layers.

The next section introduces several main fault management features of each network level introduced in [66]:

Optical Layer:

• "Fast fault failure detection": the loss of light or carrier signals detection and switching to a backup light-path (if configured).

• Limited at lighpath granularity.

• No discrimination between traffic types.

Sonet/SDH Layer:

• Limited to ring topologies and may not always include mesh protection.

• Cannot distinguish between different priorities of traffic.

• No vision of higher layer failures.

• Limited to link failures

ATM Layer:

• Node failure detection

• "in band OAM functionality": fast path error detection.

MPLS Layer:

• Node/link failure detection: "Path Continuity Test", "Fast Liveness Message Test"

IP Layer:

• Re-routing mechanisms (too slow).

2.3.2 Protection and restoration in optical burst switching

Although significant research has been carried out into optical network restoration and protection, few research works have investigated that issue in the context of optical burst switched networks.

Most of the works in the literature have focused on wavelength routed networks where the major concern is the restoration or the protection of an already established light-path [68,69,70,71]. In this context both protection-based and restoration-based schemes have

been proposed. In protection-based optical networks, dedicated protection mechanisms such as redundant resources (backup light-paths or backup links) are established to cope with failures. This is very similar to the techniques used in conventional networks where at the moment of establishment, the path/link could be protected by another path/link (GMPLS uses this kind of mechanism and could deploy both link-based and path-based protection). An intermediate scheme (between path and link protection) is proposed in the literature.

The Sub-path (or partial path) protection [72,73,74,75]scheme is another alternative , which is specifically designed to couple the advantages of both path protection and link protection. This technique consists of partitioning the primary path into sub-paths or several overlapped segments and, dynamically, computing their backup paths. In case of failure, the segment affected can switch to its backup. Therefore, the protection is achieved with more resource sharing than link-based protection and is faster than path-based protection.

In restoration-based networks, once the failure occurs, a computation is made on the fly to restore the light-paths affected by the failure. This process can to restore the effected light-paths only (without disrupting the existing unaffected light-paths) or can reconfiguring the whole network and switching to another virtual topology that bypasses the effected part of the network [76,77]. The latter technique, which is called the reconfiguration approach uses the resources more efficiently and has the potential to reroute more traffic than the former approach. However, the major drawback with this technique is the disruption of ongoing traffic (traffic unaffected by the failure).

Very few works have considered restoration in the context of optical burst switching. In [78], a restoration scheme is proposed for wavelength routed optical burst switching (the switching of light-paths on a burst by burst basis); in the case of a link failure, the control node updates the routing tables such that all the subsequent bursts are routed around the failed link.

Optical burst switching still relies on higher protocol layers to recover in case of failure. Nevertheless the opportunity exists to provision and restore the failed traffic in the optical domain. That adds more resilience to the network and frees the higher layers from the burden of monitoring the network for restoration purposes.

The architecture of an optical burst switching network is very simple; indeed, the topology seen by the higher layers is the same as the physical topology in opposition to the routed wavelength network where the physical and logical topology can be completely different. This suggests two classes of restorations:

- **Local restoration:** in case of link (or node) failure, the two nodes connected (or the set of nodes directly connected to the failed node) are required to find another path that bypasses the link (or the node). The computation of the new path is relatively easy since the only information needed is the network topology which is supposed to be known by all the network nodes. This kind of restoration aims to isolate the effected part of network without disturbing or involving the other nodes in the restoration process. However, this technique has many challenges to overcome:
    - it requires intelligence in intermediate nodes to compute the new route;
    - an additional protocol is required to let the other network nodes know about the new topology for future use;

- the new route may incur additional delay and this can be in contradiction with the quality of service (delay) expected by the source; (indeed, the new route may be longer than the first one and hence the propagation delay may be longer. This delay could be unacceptable for the traffic being carried on this route.
- the new route may cross more intermediate nodes than the original one which may cause a technical problem, since the source computes the offset between the burst header and the data according to the number of intermediate nodes crossed.

- **Source restoration:** in case of failure of any component (link or node), the nodes involved in this failure are requested to advertise the new state to all the source nodes. Consequently all the sources will compute a new route taking into account the failure in the network. The source nodes can also identify the lost bursts (all the bursts crossing the link or the node affected). This technique does not need any intelligence at the intermediate nodes and is very convenient for the source routing. However, this technique has some challenges to overcome:
  - a reliable protocol is needed to advertise a failure or recovery, this protocol should be fast enough in order to stop sending more burst towards the failed part;
  - all the traffic, which has been sent between the moment the failure occurred and the notification of the sources, should be sent again.

The local restoration is very similar to link-based restoration whereas the source restoration is very similar to path-based restoration, both of them need the following steps:

- failure detection;
- failure advertisement;
- new route computation (for all the traffic crossing the affected part);
- switch-over to the new route;

Chapter 3

Segmented optical burst switching

3.1.Introduction

With the growing demand of bandwidth and the development of optical components technology, the IP over DWDM using optical burst switching seems to be a most promising solution to take advantage of the huge capacity of the fiber and accommodate high information traffic. In this architecture the optical network is seen as an optical cloud with intelligent edges capable of the interpretation of the IP address and the storage of the information in the electronic domain as well as the checking and correction of errors. Optical burst switching achieves better bandwidth exploitation (compared to circuit switching technique) because all the fiber wavelengths are shared among the bursts without resource pre-allocation and the whole wavelength capacity can be used by a burst. However, with higher load, contention increases, and hence the number of dropped bursts increases leading to a serious loss of performance. Several methods can be used to lower the burst-dropping probability, such as wavelength conversion and the use of buffers (but these solutions are still not used due to the high cost and the immaturity of technology). Other simple solutions have also been proposed, such as the delayed burst method and deflection routing.

In this chapter we propose an alternate method where the burst is segmented into several parts of equal length and, in case of contention, only the parts that cause the conflict will

be discarded [20,21,79]. This method will be analyzed and compared with other practical methods: delayed burst, deflection routing. The segments of a burst can be used to carry different classes of service since the dropping probability of each segment depends on its position in the burst. The parts at the end have the smallest probability of being dropped. We prove through analysis and simulation that the segmented burst method improves the network performance and is more suitable for traffic with several class of service.

3.2.Segmented optical burst switching

The contention of bursts depends on the physical topology and resources available such as the number of wavelengths and the network connectivity, and also the burst length and the traffic load. The longer the average burst length, the higher is the dropping probability. In conventional OBS, the edge node assembles the incoming packets into a burst, and once sent, the burst will reach the destination or all the information in the burst will be lost in case of conflict. On the other hand, if the average length of the burst is short, the overhead due to the switching time (time the switch spends in setting up its cross-connect) will be more significant, which leads to a waste of bandwidth. In order to send a burst (short or long) and avoid the loss of the whole burst (in case of contention), it is proposed that the burst is divided into many segments with the same length. With the segmented burst, the ingress router will group incoming packets into a list of segments with fixed length (one to five IP packets in a segment). All the segments will be sent with the same OBH and with a short time separation between two segments (the size of the burst is identified by the number of carried segments). In case of contention instead of dropping the whole burst, only the contending segments (including those needed for the

switching time) belonging to the second burst will be discarded whereas the other segments can continue on their way as shown in Figure 3.1. The information carried by the OBH is essentially the same as in conventional OBS, it is necessary just to indicate the number of segments in a burst instead of the length of burst, and this information may change at intermediate nodes whenever some of the segments are dropped.

Figure 3 1 shows contention between two segmented bursts and how a part of the burst may be saved instead of the whole burst being dropped.

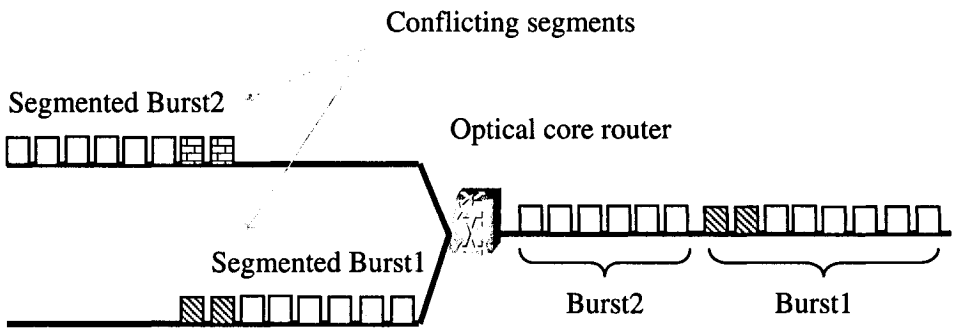


Figure 3.1: Contention in optical segmented burst switching

Segmented burst adds more flexibility to OBS and improves the throughput of the network since it reduces the dropping probability. Unlike buffering and deflection, burst segmentation introduces neither additional latency nor additional load to the network.

In buffer-less optical networks, the scheduler in OBS is constrainted to respect the arrival time of the bursts; this can lead to unused slots of time between successive bursts.

Example: The scheduler receives three optical burst headers with the following information:

- Burst 1 will arrive at 0.4 and end at 0.6

- Burst 2 will arrive at 0.5 and end at 0.7
- Burst 3 will arrive at 0.7 and end at 0.9

All the bursts want to leave through the same output port. If the node cannot delay the second burst, it will be dropped. Consequently burst 1 and burst 3 can leave without problem. However, between 0.6 and 0.7, the output port will be free and unused, which is considered a wasted resource. SOBS overcomes this problem; it will take some segments from burst 2 for transmission and fill the gap between burst 1 and burst 2.

Using the segment granularity, the traffic becomes more fluid, so the output capacity can be used completely, the efficiency will be limited only by the switching time.

The quality of service provided by OBS is another concern as there is no criterion to take in consideration when conflict occurs. Some approaches have been already proposed in [13] that give better quality of service to higher priority traffic by assigning a larger offset to the burst. However, in that scheme contention will always remain among bursts belonging to the same class. Furthermore, the flexibility of OBS is diminished as the burst length should respect some constraints.

With segmented bursts, the segment at the end of a burst has a greater probability to survive than the others, as the segments at the beginning of the second burst are dropped. The higher the position in the list of segments the smaller is the probability of being discarded. This important feature can be used to provide different quality of services. Indeed, at the ingress edge, the incoming packets are sorted according to their destination and their class of services (CoS). Information with high priority will occupy the last segments in the burst.

Many classes of services (in term of dropping probability) could be carried by SOBS. The highest priority service makes use of the last segment, whereas the segment at the

head offers a lower priority service. The position of the segment defines its priority; the closer to the tail the higher is the priority.

Compared to the other techniques (basic OBS, buffering, deflection, OBS with conversion), SOBS is a practical way to effectively reduce the loss and enhance the quality of service. Table 3.1 summarizes the difference between these different forwarding methods

Method	**Additional Latency**	**Additional Hardware**	**Support QOS**	**Complexity**
Basic OBS	No	No	Yes	No
OBS with buffer	Yes	Yes	No	Some
OBS with deflection	Yes	No	No	Deflection decision
OBS with conversion	No	Yes	No	Hardware
OBS segment	No	No	Yes	Some

Table 3.1: A comparison between the 5 optical switching methods

3.3.Performance evaluation

In this section, we will focus on the comparison and performance evaluation of efficient OBS protocols that can reduce the burst dropping probability. These techniques are OBS with limited FDL, OBS with deflection and burst segmentation. How segmented bursts are suited to carrying different classes of traffic, especially the dropping probability of the last segment, is also studied.

The OBS techniques are sensitive to the traffic load and network topology; therefore the simulation will take these parameters into account. The traffic load is measured in term of burst arrival frequency and burst length. The network topology is represented by

the number of output links of a given switch and the number of fibers constituting each link.

A system is considered with a single switch with high-speed switching capacity, with Poisson data burst arrival process (no correlation between the bursts coming on different inputs) over each link with a rate λ. It is assumed that the behavior of a single switch will reflect that of the mesh network.

It is also assumed that this switch does not support wavelength conversion, therefore there is no interaction between the wavelengths, and only one wavelength need to be considered in the simulation.

The system generates a list of bursts. Each burst is characterized by its arrival time, its length and the destination, which is one of the outputs:

- the arrival time is generated randomly with Poisson rate λ.
- the burst length is generated with uniform distribution between Lmin and Lmax (Lmin is the minimal length of bursts and Lmax the maximal one).
- the destination is one of the outputs picked uniformly at random.
- the length of bursts is uniformly distributed between 1 ms and 19 ms with an average of 10ms.

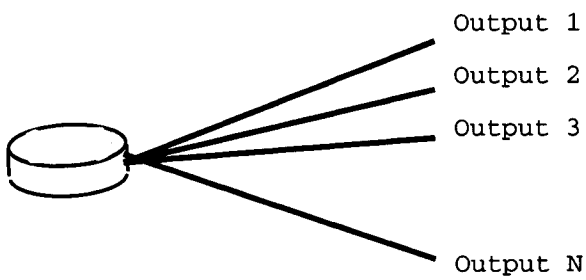


Figure 3.2: Simulated switch

An output link may have one or more fibers. In this simulation, the output link has one or two fibers.

The same list of bursts is used to simulate the forwarding process according to the following switching methods:

1. Basic OBS.
2. OBS with limited buffering: In this simulation the buffer size is 10% of the average size of bursts. An in-depth study of the impact of buffer size on burst loss is given in [24,25].
3. OBS with deflection: It is assumed that when contention occurs, all the other outputs are indifferently considered for an alterative route (if the destination output is busy, one of the free outputs will be selected otherwise the burst will be dropped).
4. SBS. In this simulation, the burst is broken into segments. And all the segments have the same length (0.1 ms).

The performance measure in this simulation is the dropping probability due to collision. The loss is calculated as the ratio of the dropped information and the total information sent. In the case of segmented bursts, the dropping probability of the segment located at the end of the burst (the last segment of the burst) is also measured. This segment carries the information with higher priority. Indeed, when contention occurs, the segments at the beginning of the second burst are removed to forward both the first burst and the remaining part of the second burst. The dropping probability of the last segment is particularly important to know in order to assess how suitable the segmented method is to offering different classes of service.

Figure 3.3 shows the loss ratio of each OBS technique as a function of the load (defined here as the ratio of the offered load to the total capacity). Each output contains one fiber. As expected, the dropping probability increases with the load, except for the deflection method where there is practically no loss due to the fact that the input capacity is equal to the output capacity. However, in this case the number of hops is not known, thereby the delivery time will increase. The burst may even loop indefinitely in the network. One needs some precaution to limit the lifetime of the burst; for instance, one may drop the burst after a certain number of hops.

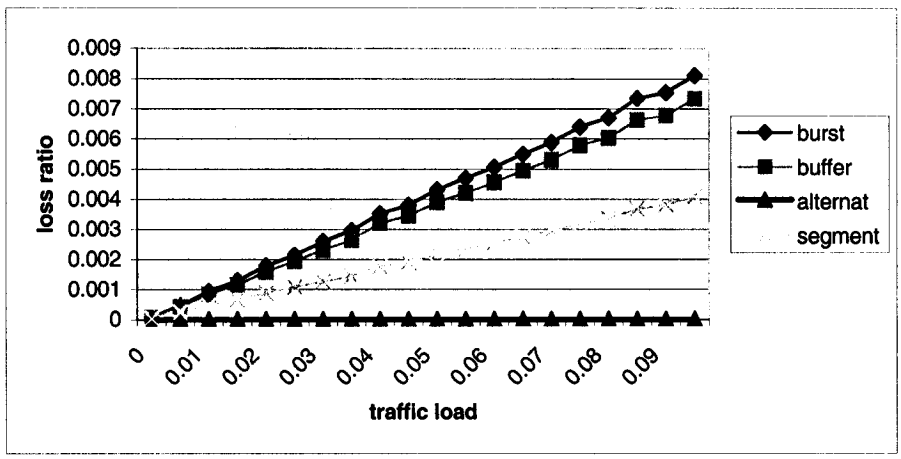


Figure 3.3: Switch with light traffic

The graph also shows that burst segmentation improves the performance, in term of loss ratio, by 50% (the loss ratio of segmented burst switching is always half or less of that of OBS). This can be explained by the fact that when a contention occurs, one or more segments are removed. Let N_s be the number of segments in a burst and P_i the probability to remove i segments. P_i is equal to $1/N_s$. Therefore the average loss is given by the formula $\sum_{i=1}^{Ns} P_i * i * S_s$ (1); where S_s is the average segment size. If $S_s=B_s/N_s$ where B_s is the average burst size, then the average loss is $B_s/2$ according to the formula

(1) This means that the number of dropped segments in the network will decrease by 50%. And hence the OBS will become more stable.

Figure 3.4 shows the dropping probability under heavy traffic. When the traffic is heavy, burst segmentation improves the performance considerably, but for both normal OBS and segmented burst switching, the loss is high. OBS and its variants are very vulnerable to the load, which should be controlled in order to limit the loss.

OBS with limited buffering has the same trend as OBS. Nevertheless it improves the performance with a fixed rate. This rate depends on the length of the buffers.

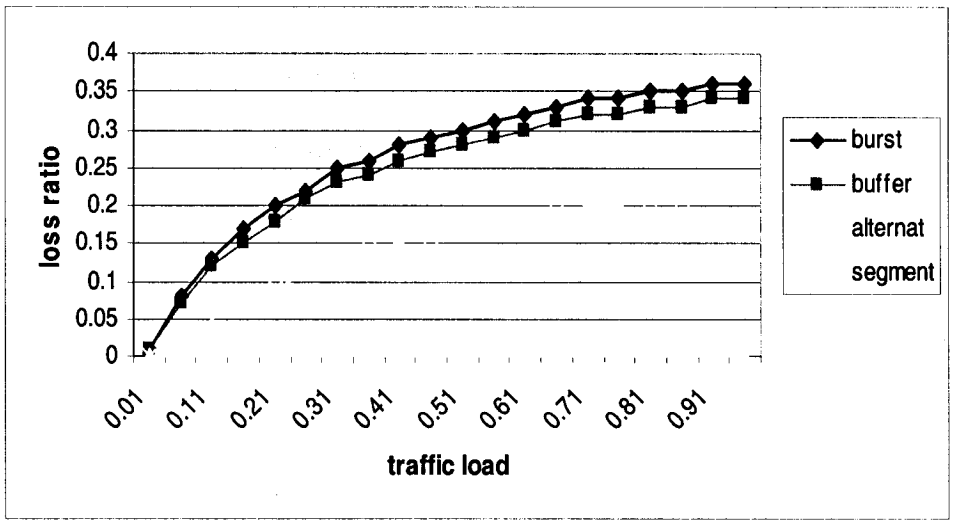


Figure 3.4: Switch with high traffic

When the number of fibers in each output increases, segmented OBS takes advantage of the additional capacity. Indeed, Figure 3.5 shows that when the number of fibers is 2, the loss is significantly lower with segmented OBS compared to OBS. For example, if the output number is 2, the loss ratio falls from 0.004 to 0.002 (for 0.1 load) with segmented OBS whereas the same loss falls only from 0.008 to 0.006 with OBS (Figure 3.3 and 3.5).

This result proves clearly that segmented OBS tends to exploit efficiently the available resources.

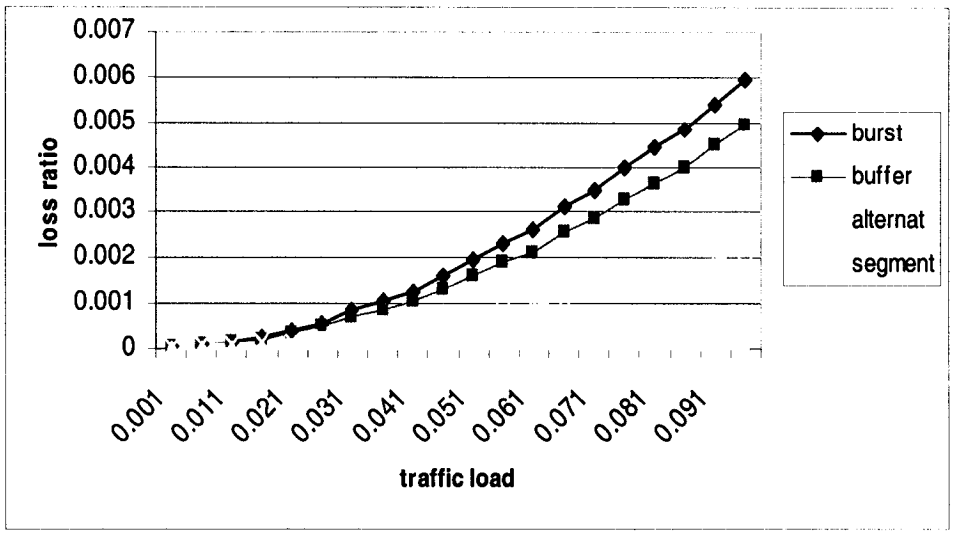


Figure 3.5: Switch with light traffic and 2 fibers in each output

The improvement achieved by segmented OBS is better than those of other techniques used to reduce contention. Indeed, the deflection technique decreases the loss only with light traffic whereas the loss is the same with high traffic [17,18], besides that the average number of hops increases. When buffers are deployed in OBS, the performance can be improved. Nevertheless the results depend on the size of the fiber delay lines (FDL) used and their architecture (simple or multi-stage) [24,25,27]. The loss can be decreased by 80 % under light traffic and with long FDLs. The loss is reduced by only 50 % under heavy traffic. However, this solution requires extra components and a complex architecture in order to decrease the loss. When wavelength converters are deployed with OBS the loss is reduced to 0 % with light traffic. However, with high traffic the results depends on the conversion ratio [27,28,29]. This solution is very expensive.

Load (%)	Burst	Last segment
00	0.0000	0.0000
01	0.0005	0.0000
01	0.0009	0.0000
02	0.0013	0.0000
02	0.0018	0.0000
03	0.0021	0.0000
03	0.0026	0.0000
04	0.0030	0.0000
04	0.0034	0.0000
05	0.0038	0.0000
05	0.0042	0.0000
06	0.0045	0.0000
06	0.0052	0.0000
07	0.0056	0.0000
07	0.0058	0.0000
08	0.0061	0.0000
08	0.0067	0.0000
09	0.0072	0.0000
09	0.0075	0.0000
10	0.0079	0.0000

Table 3.2: Dropping probability of the last segment with low load

The last segment has a greater chance of surviving contention since the dropped segments are those at the beginning of the second burst. Table 3.2 shows the blocking probability, for a switch with 10 outputs, for both the whole burst and the last segment of the burst. Regardless of the number of outputs, a segmented burst always shows better performance. At low loads, the dropping probability is almost zero and the last segment is almost guaranteed to reach its destination. Indeed, the average loss probability is more than 7*10-3 (when the load is larger than 9%) whereas for the last segment it does not exceed 1*10-5. This makes the segmented OBS more suitable for methods offering different classes of service.

Similar results are observed under heavy traffic. The last segment has always the larger probability to reach its destination. Table 3.3 shows the loss probability for the same switch at high load; the loss probability of the whole burst reaches 0.3 but the loss

probability of the last segment is lower than 5*10-3 (with an average of 100 segments in a burst). The last segment has more chance to avoid contention. Indeed, the last segment is dropped only if the two contented bursts arrive at the same time. Such an event occurs with relatively small probability. Furthermore, the loss probability of the last segment can be improved by decreasing the length of the segment. The segment position in the burst determines its priority in term of loss probability and hence the burst can carry different level of classes of services.

Load (%)	Burst	Last segment
05	0.03979	0.0004
10.1	0.07248	0.0007
15.1	0.10092	0.0011
20.1	0.12596	0.0015
25.1	0.14737	0.0016
30.1	0.16660	0.0020
35.1	0.18350	0.0023
40.1	0.19936	0.0025
45.1	0.21262	0.0027
50.1	0.22533	0.0031
55.1	0.23638	0.0033
60.1	0.24752	0.0031
65.1	0.25635	0.0035
70.1	0.26519	0.0036
75.1	0.27331	0.0037
80.1	0.28059	0.0040
85.1	0.28814	0.0039
90.1	0.29412	0.0040
95.1	0.30980	0.0047

Table 3.3: Dropping probability of the whole burst and the last segment with high load

3.4.Summary

DWDM has emerged as a promising technology for the next generation of networks to meet the growing bandwidth demand and to take advantage of the huge bandwidth capacity of fiber. OBS is one of the proposed solutions to be used with DWDM to route information in switched networks. Indeed, OBS has significant potential to exploit the

bandwidth provided by DWDM. In this chapter, the contention problem that occurs when two, or more bursts compete for the same output has been discussed. A new method has been proposed to reduce the loss probability and to bring more flexibility in order to use the whole capacity of the wavelength channel. In this method, bursts consisting of several segments are introduced to avoid the loss of the whole burst in case of contention. The benefits of this method and how one can use it to carry several classes of traffic have been presented; and simulation proves that the last segment has a much smaller probability to be dropped compared with the dropping probability of the burst as a whole.

All the segments have the same length in the proposed scheme. Nevertheless, it is possible to have variable segment sizes [80] in order to carry more easily information coming from different networks (ATM, IP, SONET, etc). However, the header must then carry more information to identify the different segments, which would require more header processing time.

In this work, only one switch with different numbers of output links has been considered. However, to evaluate the edge-to-edge loss probability we need simulation with a whole network using several intermediates nodes must be simulated. Another interesting issue is how to combine burst segmentation with other methods of congestion control in order to avoid the loss and to balance the load over the network. For instance, one can send the segments causing a conflict through another output port to avoid dropping them. Another solution to enhance segmented OBS consists of combining segmentation with wavelength conversion. In case of contention, the whole burst or just the segments causing the conflict could be shifted to another wavelength.

Chapter 4

A hybrid architecture using optical burst switching and routed wavelengths

4.1. Introduction

Optical burst switching and wavelength routing networks are proposed to carry information in optical networks. In this work, a brief description of these two techniques is presented, the advantages and disadvantages of each scheme are discussed, and we furthermore propose a new architecture that uses both methods in order to overcome the limitations imposed by each approach. Network backbones are required to carry different classes of applications with different qualities of services. Some application may require low latency (like multimedia applications) and others may require low loss (like file transfer). Therefore, it is important for network architects to decide what their traffic classes should be and define the expected quality of service.

4.2. Wavelength routing network architecture

With the developments being made now in WDM, which provides several wavelengths in the same fiber and the capability of a cross-connects to switch a wavelength separately, it is possible to set-up a direct light-path between two edge nodes. This pipe of light can be seen as a separate link operating autonomously without interaction without any other links or awareness of any router crossed by the pipe. Such a path, also called a virtual

link, can carry information transparently in one hop from a source to a destination without the need for any conversion of the signal from the optical domain to the electrical domain.

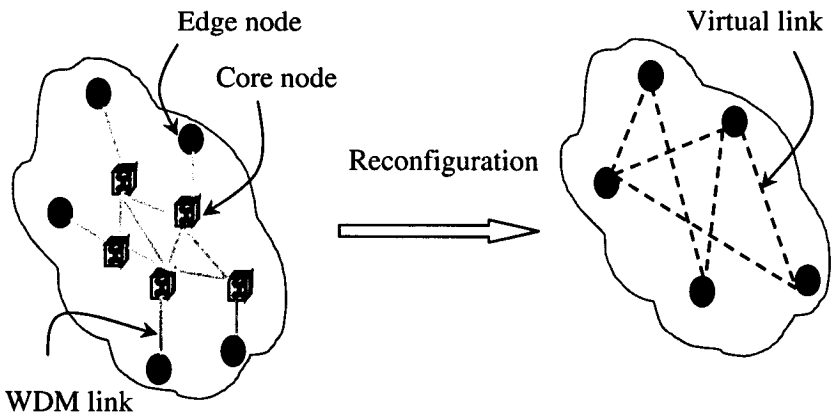


Figure 4.1: Virtual topology

Even better, this technique can be used to build a new topology by setting up many paths inside the network as shown in Figure 4.1. This new architecture aims to build a virtual topology [11,12] completely different from the physical one and reduces the complexity related to optical technology (the virtual topology hides the physical one). The presence of such a layer enhances the transport network and bridges the gap between the conventional and optical networks by providing a flexible high-capacity interconnection service to the electrical transfer modes; indeed the reconfiguration provides a new topology that can be used with the well-known protocols and principles. The virtual topology, that can be completely different from the underlying physical topology is re-configurable, which means that it is able to adapt to variations in traffic load. This idea seems to be very attractive and may be one of the best ways to take advantage of the potential of WDM technology, but there are many challenges and difficulties to overcome, especially those related to the assignment of the different

wavelengths available, also the optimization of the used bandwidth and the propagation delay. The set-up of the new architecture could be done synchronously or asynchronously:

- Asynchronous reconfiguration: Whenever an edge requests a bandwidth to carry information to a given destination, the network manager will analyze the available resources and set up a new path between the source and the destination (if there are enough resources). The path can cross several physical links and several cross-connects to provide a direct light-path from source to destination. This virtual link can use the same wavelength or can switch between different wavelengths if there are some wavelength converters. The main concern in this architecture is how to deliver light-paths in a cost-effective and timely manner. The functionalities needed in this architecture to achieve that goal are:

  - Network state information: to maintain the state of the available resources, the topology and the schedule of the used resources.
  - Light-path computation: basically this is the algorithm and the strategy used to assign wavelengths to a virtual link.
  - Light-path services: the global management of the virtual links, especially the establishment of a new path, the release and the modification of a link.

When a request is received, a perfect manager, will analyze the available resources and establish a direct light-path from the source to the destination. If there are many possibilities, the optimum will be chosen. The optimization criteria could be simply the propagation delay or could be more complicated and takes into account many other parameters such as the effective cost and the restoration constraints. If there is no

possibility of establishing a direct path, the manager will establish two or more light-paths in such a manner that the information can reach the destination with two or more hops. The new path can share already established paths if some bandwidth is left from the previous allocation. The main goal here is to minimize the number of hops and aggregate the traffic to avoid bandwidth waste. This path can be set up for a limited time; this way the manager can schedule the set-up of different links.

- Synchronous reconfiguration: In this approach all the edges' requests are collected and the reconfiguration is performed at the same time. The network seems to shift from a configuration to another to adapt itself to the new load. The reconfiguration could be done periodically or only when the traffic pattern is too different from the previous one. The functionalities needed in this case are:

- Load management: the goal is to collect the requests of the edges, in terms of the bandwidth and eventually the quality of services needed for each class of traffic
- Coordination and synchronization: to keep track of the traffic being carried on the virtual network, and coordination among all the edges to switch to the new topology at the same time
- Light-path computation: depending on the requests by the edges, this function should provide the set of light-paths.
- Topology switcher: to switch to the new topology computed by the light-path computation function, all the nodes must be engaged in this process, this is possible through a protocol that informs each node which configuration to be set and the time it is to be performed.

When the manager decides to switch to another configuration, the edges' requests are analyzed and the best topology is computed. The perfect situation would be to establish a virtual link between each source and destination, but this is impossible since the number of wavelengths is limited and the number of edge nodes may increase. Nonetheless an optimum is possible. The goal is to minimize the hop number, the delay and the throughput. In case there are no direct paths to some destinations, a tandem edge node maybe used. Even if this incurs additional latency, it may be important to aggregate as much traffic as possible in the same wavelength and avoid waste of bandwidth.

The reconfiguration of a virtual topology is one promising way towards an agile network that can be easily adapt to any traffic pattern and hence, balance the load. However, one of the challenges that this technique must face is traffic synchronization, especially during the reconfiguration process. Indeed, due to the propagation delay, the information will still remain inside the network when a virtual topology is being torn down to build another. The easiest way to deal with this problem is to stop sending information for a while, before each reconfiguration, until all the information reaches its destination. However, this can lead to a large waste of resources and may introduce a bottleneck at the edges. The reconfiguration of a virtual topology can be a very elegant solution when it is used with a specific architecture such as the Petaweb [14] where the regularity of physical topology makes it easy to synchronize and coordinate the entire nodes of the network. The virtual topology relies on a wavelength assignment algorithm [15], which can be formulated as an optimization problem with many constraints related to the physical resources and the edges requests. The optimization goal is to maximize the performance in term of throughput, propagation delay, restorability and many other

parameters. To solve such a problem, one needs mathematical models [81] related to constraints satisfaction problems, however, the problem complexity increases rapidly with the size of the network in a manner that an exact solution seems to be unfeasible. So heuristics and simplifications are needed in order to reduce the complexity of the problem [82,83].

4.3.Shortcomings of existing approaches

Used separately, neither OBS nor wavelength routed networks are suitable for all classes of traffic or to guarantee the quality of services needed for all classes of applications. Each method has its own limitations and suffers from many drawbacks. With OBS, the burst loss due to the contention, which is inherent to this technique, decreases the performance in term of throughput and delivery delay, especially with high load. Several methods have been proposed in the literature to decrease the loss ratio. However, the contention still remains. This means that the delay and the delivery are not guaranteed and hence, this type of networks may be useless for certain kinds of applications, especially real-time and loss-sensitive applications. On the other hand, wavelength routed networks have achieved some progress, especially those related to wavelength assignment algorithms. Unfortunately, this technique is still facing many problems, such as:

- The complexity of the wavelength assignment algorithm increases with the network size and the number of wavelengths, which can hinder the future expansion of the network.

- Due to resources limitations, it is sometimes impossible to establish a direct light-path between all the edges. Therefore an intermediate node must be used as a tandem node, which can lead to additional delay and routing complexity.
- Even when a light-path is established between two edges, it is not necessarily the shortest one.
- Edges may not have enough load to fill all the capacity of the established path, and hence a part of the bandwidth may remain unused.
- To establish a new light-path, the manager may need a long time to analyze the resources available.
- Some of wavelengths are left unused on certain links.

4.4.Hybrid architecture

To aggregate different classes of traffic and to overcome the limitations of these methods, used separately, we propose to combine the OBS and wavelength assignment in the same network [84,85]. The basic idea of this architecture is to take advantage of the large number of wavelengths in the fiber by using only a part of the available wavelengths for wavelength routing network and the other part for OBS. Depending on the traffic pattern, the wavelength routing sub-network will be configured to balance the network load while the other part remains free to be used with the OBS technique.

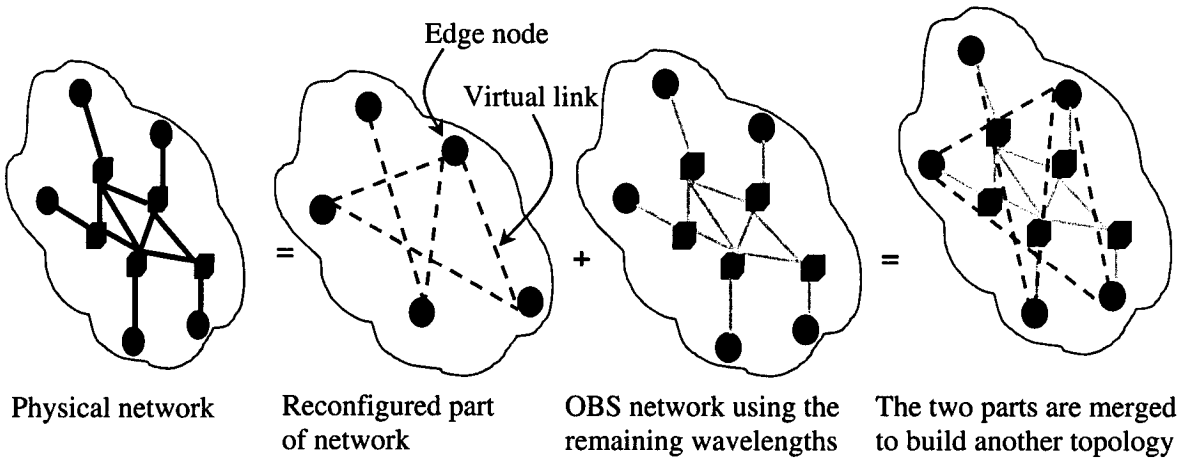


Figure 4.2: Hybrid architecture

Figure 4.2 shows this hybrid architecture; the new network is the result of merging the two sub-networks. The new topology is better than the physical one since we obtain more connectivity. And hence the OBS traffic can be sent over the two sub-networks (OBS and virtual topology). Let CEi be the set of edge nodes directly connected to the edge node Ei via a virtual light-path after the configuration of the wavelength routed part of the network. The traffic can be carried on the new topology whether using the deterministic paths of the wavelength routed network, which can be suitable for many kind of applications such as multimedia application, or using random access with the optical burst switching sub-network, which is also very suitable for other kind of applications. Even better, the new architecture will improve the performance of the OBS technique; indeed, whenever an edge decides to send a burst to a destination, it will first consider a direct light-path. If there are many, the best is taken, otherwise the burst will be carried over the OBS sub-network. Another enhancement can be brought to the OBS technique using this hybrid architecture as can be seen when contention occurs: instead of dropping the burst it can be rerouted to another destination. Let Ej be the original

destination; a destination Ek that belongs to CEj will be chosen. This way, Ek can easily forward the burst to Ej. Also Ek could be the closest node to Ej among CEj. This could decrease considerably the dropping probability and hence decreases the delay and improve the throughput. This architecture does not need any additional hardware and can be easily used with the infrastructure used in OBS. The whole network can be used to carry the traffic between the edges and provides, at the same time, wavelength service by establishing a direct light-path between edges.

The combination of the two techniques may also increase the utilization of all network links; the links left idle in the wavelengths routed part can be used by OBS by adding some wavelength converters in the network. Furthermore, the network restorability may be improved by simply switching the traffic being carried on a failed link to the OBS sub-network during the restoration period. Different classes of traffic can be carried on this network, depending on the destination and the quality of service needed, the source can use either the wavelength routed sub-network or OBS or both of them in some case. Another advantage of the new architecture is related to the dropping probability; indeed the wavelength routed part adds more links and hence the connectivity of the network is increased. The average number of hops to reach the destination is lowered consequently. The next section presents the results of simulations to measure the dropping probability of this hybrid architecture.

4.5.Simulation results

The blocking probability and delay of the hybrid architecture proposed in the previous section is evaluated on a random mesh network with 10 nodes and 6 wavelengths. The experimental set-up for the simulation is based on the following assumptions:

(1) The burst arrivals to the network edges follow a Poisson process with rate λ.

(2) The burst length is exponentially distributed with rate μ.

(3) The bursts are sent only by the edge nodes and the destination is uniformly distributed over all the edge nodes.

(4) The routing table is static, the burst takes the shortest path from source to destination.

(5) The transmission rate is 1 Gbps.

(6) Three wavelengths are used for the wavelength routing sub-network and the other three are used for OBS.

(7) the traffic load is expressed as a ratio of the full load.

In this simulation we focus on two parameters: the loss rate which is the number of dropped bursts divided by the number of bursts sent by the edge node, and the average delay which is the mean time of all the received bursts from source to destination. The simulation compares the performance of the network with two architectures: the first where all the wavelengths are used for OBS and the other where 3 wavelengths are used for OBS and 3 for wavelength routing.

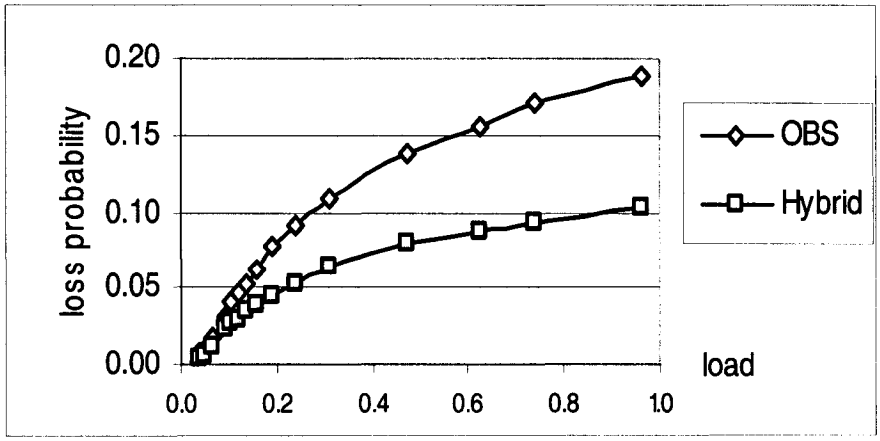


Figure 4.3: Loss probability for OBS and the hybrid architecture

Figure 4.3 shows that the loss probability is always better with the hybrid architecture. The delivery is improved considerably. This enhancement is due to the fact that a part of the traffic is routed over a deterministic sub-network without contention. Clearly, the loss probability depends on the fraction of the traffic that is sent over the wavelength routed sub-network. In this simulation the burst is sent over the network routed sub-network only if there is a virtual path between its source and its destination. Many policies could be applied; however, the policy used may affect the delay due to the processing time at each intermediate node. The new architecture also improves the average of the delivery delay as shown in Figure 4.4. The delay is the average time from source to destination for all received bursts. This average includes the queuing time and the propagation delay. The offset time between the optical header and bursts is negligible (ignored) in this simulation.

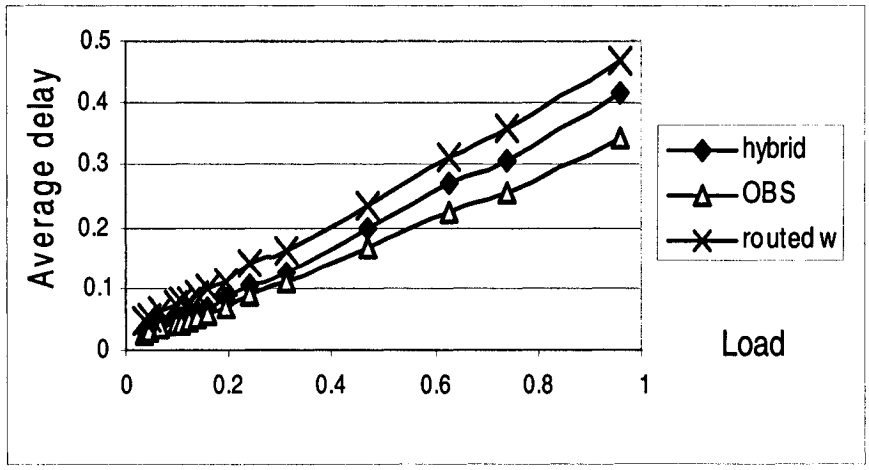


Figure 4.4: Average delay (ms) of OBS, wavelength routing and the hybrid architecture

The delay of the wavelength routed architecture is improved. Nonetheless, depending on the strategy used to dispatch traffic between the two sub-networks (how much traffic is sent over each sub-network), the delay will range between the average delay of OBS network and that of wavelength routing network. Indeed, with wavelength routing the buffering delay maybe higher. This means that there is a trade-off between loss and delay. Therefore one needs to take into consideration both the delay and the loss.

4.6.Summary

In this section, we proposed a hybrid architecture that uses both the OBS technique and wavelength routing technique to take advantage of the large capacity of optical networks. With this hybrid architecture, the network can carry different classes of traffic, using either the deterministic paths provided by wavelength routing or the best efforts provided by OBS. This technique is motivated by the advance in DWDM technology. Indeed, hundreds of wavelengths could be operating at the same time in the same fiber. The

hybrid architecture solution aims to use part of these wavelengths for OBS, the remaining parts for routed wavelength and merge the two sub-networks to provide another virtual topology with more links and more connectivity.

An area for future work is the investigation of the optimum partitioning of the network in order to determine the percentage of the available wavelengths to be used with OBS and with wavelength routing. Another concern is the policy used by the edge nodes to dispatch traffic over OBS and wavelength routing. An example of such dispatching policy could be the one used in our simulation. It consists of using the wavelength routing sub-network if there is a direct virtual path between the source and the destination of the data to send. Another policy could take into account the data being sent (loss-sensitive data will be sent over the wavelength routed sub-network). To reduce the burst loss and decrease the delay, one needs to classify and compare different dispatching policies.

Chapter 5

Contention avoidance using congestion control

5.1.Introduction

The basic differences between an optical burst switched and a packet switched network are the techniques used to forward information at the network nodes as well as the layers involved in the routing process (in OBS the switching is performed in the photonic domain). Indeed, in the packet-switching network, the switches have the capacity to store and process information. In addition, an intermediate node can participate in managing and monitoring the network. Therefore, with this distributed architecture, the network can face difficult situations (in terms of load and congestion) and regulate the network load by using explicit methods to control the traffic flux. However, in optical burst switching all intelligence resides at the edge nodes, which are at the same time the buffer and the processor of the network, whereas the intermediate nodes are used to forward messages according to their destination with no global coordination. Burst paths are determined at the edge nodes according to some information such as the physical topology and the physical features of the switches. This lack of information at the edge nodes (the global state of a network is unknown) may cause the network to drift into an overloaded state where the intermediate nodes are experiencing more contention, leading to a large waste of bandwidth due to an excessive drop of bursts [11,12]. Besides that when a burst is dropped (because of contention), it is simply ignored. The network relies on a higher

layer protocol to recover and retransmit the dropped information (carried by the dropped burst) which may increase the wasted resources and the delivery delay (since a retransmission should be performed from a source to a destination).

In this work, we propose to enhance the performance of optical burst switching and eliminate burst loss completely. As a first line of defense, we propose to reduce contention by controlling the load and avoiding congestion. In the second step, we retransmit the dropped bursts. These two steps are complementary since the retransmission would be useless if the loss rate is very high. Indeed, if the loss rate is very high one could retransmit the same burst many times, which may increase the average number of retransmissions, and hence the delivery delay increases.

In order to control the load and avoid congestion, the intermediate nodes provide the edge nodes with statistic information on the burst loss rate. Using this information one could apply admission control and adjust the traffic at the edge nodes. In this scheme the edge nodes could have an important role, since they can store a burst or postpone its sending whereas intermediate nodes are only reporting losses. This way the edge node could retransmit the dropped burst, which may increase the network robustness and reliability.

5.2. Congestion avoidance in optical burst switching

In order to reduce contention, the load is a key element, since heavy traffic increases the burst loss rate. In addition, each burst dropped means wasted bandwidth, increased delivery delay and decreased throughput. This means that the global efficiency and

performance of the network depends on the loss rate, and hence the performance falls as the load gets higher.

Figure 5.1 shows the performance (in terms of delivery rate) as a function of traffic load. The graph represents only the performance pattern; the shape of the curve may depend on the network connectivity and the physical resources such as the number of channels per link and switches capacity. Each network has its own curve and it is completely characterized by this performance graph.

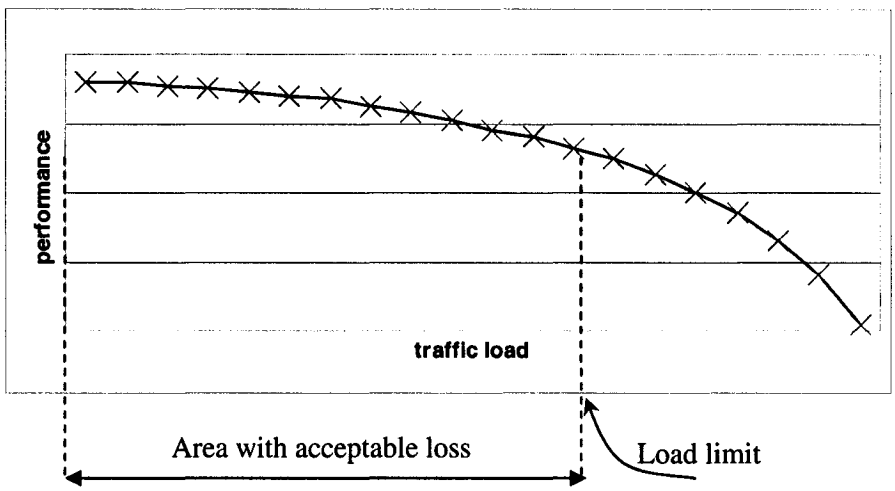


Figure 5.1: Performance as function of traffic load

According to this graph the delivery rate keeps decreasing as the load increases, until it becomes excessively low. We divide the traffic load into two ranges:

- The area where the loss is acceptable. The load limit (LL) is the upper limit of this area (the selection of this value may depend on the class of traffic). The LL itself depends on the maximum acceptable loss rate and the physical topology of the network.
- Contention area where the loss is too high.

In this work we propose [86] an approach to keep the load in the acceptable area and make sure that all the edge nodes contribute fairly to this load. The basic idea of this technique is that the edge nodes receive statistical reports (concerning the loss inside the network) that help to calculate the network performance, and hence determine from the loss-load relationship the current traffic load. Therefore by learning this statistical data, each node increases or reduces its throughput. The statistical reports could be used by the edge nodes to monitor and control the whole network. The edge nodes receive information concerning the loss inside the network as well as the successfully delivered bursts, which allow them to know the congestion state of the network, and hence adjust their traffic accordingly. A statistics distribution protocol could be implemented in the control plane, using the same wavelength used to carry the burst headers.

This approach aims to control the traffic and keep it out of the congestion area. Similar approaches to congestion avoidance, have been considered in the literature for TCP/IP packet switched networks and asynchronous transfer mode (ATM) [87,88,89]. Congestion control is a recovery mechanism that helps a network to recover from a congested state, whereas a congestion avoidance scheme ensures a network operates in a safe area. Many solutions have been proposed in the literature to practically control congestion, the most popular are window-based flow-control and rate control. In the window-based flow-control scheme used by TCP [90], the destination specifies a limit on the number of packet that can be sent by the source. This limit is increased and decreased by the destination dynamically during the whole session to regulate a data flow. In the rate flow-control scheme used by ATM [91,92,93,94] the destination or the network may ask a source to decrease its rate. Besides that, ATM uses other sophisticated mechanisms

to control congestion including traffic shaping and admission control as well as resource reservation. Regardless of the efficiency of these mechanisms, all of them perform congestion control at the packet (or cell) level where some resources are available especially buffers and storage spaces that contribute actively in the control process. The idea of optical congestion control is to push some of these functions to the optical domain where new constraints (buffer-less network) and new challenges rise. Performing congestion avoidance and congestion control in the optical domain increases the performance (in terms of loss rate) of optical burst switching and improves resource utilization.

- Each edge node is guaranteed the amount of service proportional to the whole capacity of the network
- Edge nodes with low traffic do not have to pay for the excessive load generated by other edge nodes. Therefore their dropping probability should be low.
- Each edge node gets a fair share of the excess capacity.

If we assume that Li is the traffic load of edge node Ei, then to keep the loss in the acceptable area, the load Li is constrained by the following formula: $\Sigma Li < LL$, where LL is the load limit. LL is calculated empirically by increasing the load progressively until the loss observed is that specified by the designers. LL corresponds to the load that generates this loss.

According to this formula, a global coordination is needed to meet the optimal conditions. Unfairness (node with light load and suffering heavy loss) may occur with heavy traffic ($\Sigma li > LL$) when some edge nodes send more traffic and overload the network.

The load limit (LLi) of node Ei is defined as the maximum amount of traffic the node can send into the network in case of heavy traffic. The load limit of all the nodes should not exceed the load limit of the network that is $\Sigma LLi < LL$, where LL is the load limit of the whole network. Every edge node Ei is assigned a maximum LLi. LL could be equally shared among the edge nodes; in this case $LLi = LL / N$, where N is the number of edge nodes.

This traffic control scheme could be performed by the edge nodes by the following algorithm:

- Let LR be the loss rate observed in the network, this value is calculated by the edge node using the information received from the intermediate nodes. Indeed, the intermediate nodes report the loss observed and the number of bursts delivered correctly.
- Let LLR be the limit loss rate. This is the loss observed when the network load is at the load limit LL.
- Let LLi be the load limit for each edge node.
- Let B a constant amount of bandwidth. B is introduced here to avoid oscillations.

An edge node Ei will behave as follow:

If the load Li is less than LLi then Ei will not be involved in the adjustment process; and it can increase its load up to LLi.

But if the load Li is more than LLi, the edge Ei must do the following:

- Decreases its load if $LR > LLR + B$ (the node is exceeding its assigned limit and maybe the source of congestion)
- Increases its load if $LR < LLR - B$ (if the edge node has more traffic to send)
- Keeps the same load if $LLR - B < LR < LLR + B$

Table 5.2: Admission control

This algorithm guaranties a minimum bandwidth to each edge node. Nonetheless, when spare bandwidth is available, (if some edge nodes are not using their full quota) the other edge nodes can share it. They will be notified as the loss ratio is below the critical loss, thereby they can increase their load progressively until the loss ratio becomes equal to the critical loss. On the other hand, if some of the edge nodes (with low traffic) increase their load, those with high traffic will give up their advance in terms of used bandwidth and if necessary they will return back to the load limit. The load limit is taken for granted for all the edge nodes.

This algorithm is a simple coordination between the different nodes of the network. Based on the report sent by the intermediate nodes, the edge nodes will measure the network efficiency. For a simple implementation, a single variable is enough to maintain the global state, this variable is updated whenever the edge nodes receive a report, in general all the nodes receive the same information and hence they have the same value of loss rate. But for more details about the network status, the edge nodes could maintain the status of each node; in this case the edge nodes will calculate the traffic load at each node according to the report received from this node and adjust different flows separately.

The information used by this algorithm is sent by the intermediate nodes using a statistic report distribution protocol. In this protocol, all the intermediate nodes will broadcast, to the edge nodes, the number of dropped bursts and some of them (those directly connected to the edge nodes) will broadcast the number of successful forwarded bursts. This accounting information will help the edge nodes to determine in which range the network is running, thereby they can redress and rectify the situation.

The broadcasting may be performed either synchronously or asynchronously

- Synchronously: each station can periodically send its report to all the edge nodes.
- Asynchronously: at specific events (whenever a burst or a given number of bursts are dropped) the intermediate node will send its report to all the edges.

We think that the second technique is more suitable to measure the drop rate. First, there is no need for broadcasting information if there are no drops. Second, with no control information received the edge nodes assume that the network load is in the acceptable loss area.

Statistical reports will be sent by each intermediate node to all the network edges through predefined broadcasting trees established between each intermediate node and the edge nodes. The broadcasting protocol can use the wavelengths reserved for the header to carry the statistics. The protocol could be IP broadcasting or an extension in the MPLS level. However, the information broadcasted should be very light in order to avoid network flooding.

When an edge node decides to reduce its traffic (if the network is in congested state), it can use its buffer to postpone the sending of some data if it has enough buffers or send it over another network if the node is connected to more than one backbone. The basic idea of this protocol is to perform an admission control, at the source node, and avoid overwhelming the network.

5.3.Burst retransmission approach

The congestion avoidance reduces the contention and improves resource utilization. However, bursts may still suffer losses (with a limited loss rate). Loss-sensitive

applications may not tolerate this loss. Therefore strict measures should be taken to eliminate the loss completely especially for loss-sensitive traffic.

In this section we propose [95,96] to retransmit the dropped bursts and make sure that a sent burst is correctly delivered to its destination. In pure OBS, there is no control at the intermediate nodes; the burst is simply ignored in case of contention. The recovery is performed by higher layer protocols. However, in OBS with retransmission, both the intermediate and edge nodes are involved in the process. Indeed, the edge node should keep a copy of a sent burst until its delivery and the intermediate node should notify and send a negative acknowledgement (in case of contention) to the concerned node with the necessary information (burst identification). Some applications may not require retransmission. In this case the retransmission may be an option that could be applied on demand.

The implementation of this retransmission scheme requires additional information: besides the label and other information related to a burst (burst length, arrival time etc) one needs the sequence number of the burst (it could be carried in the burst header).

- The source node sends a burst, keeps a copy and sets a timer (the only delay is the propagation time since a burst is not stored at intermediate nodes during its way to the destination. Therefore the source knows exactly the arrival time of the burst; a timer is set to a round-trip delay from source to destination)
- If the source receives a negative acknowledgement it retransmits the burst and repeats the same process

- If no acknowledgement is received during the lifetime of the timer, the node assumes that the burst has reached its destination and removes the local copy. The timer is the round trip of the burst from the source to the destination.

Some parameters are crucial for the feasibility of such a scheme. One of them is the buffer size of the edge nodes; especially for very large networks where a round trip could be very significant. One may need to store many bursts during one roundtrip period. Another parameter is the delivery delay, which could increase with the number of retransmissions. The network size also affects the delivery delay.

This retransmission scheme is more suitable for relatively small networks (metropolitan or local area networks). In this case, the size of the buffer is acceptable and the propagation delay is short. Retransmissions do not incur long delay. One can afford to store a burst during many round-trips delays. Furthermore, if the delay is large the higher layer (such as TCP) may trigger a retransmission which may increase the network load.

5.4.Analytical model in star network

In order to keep the delivery delay acceptable one should control the average number of retransmissions. By controlling a load and avoiding congestion, the loss rate could be decreased and consequently the number of retransmissions reduced.

In order to evaluate the retransmission scheme we used an optical star architecture. In fact, a star topology is relatively simple and represents an attractive and versatile architecture that could be used to build other complex architectures by combining or overlaying many star networks together.

Figure 5.2 shows the model we are using in this evaluation; the edge nodes send bursts to the core node, which forward them to their destinations (if resources are

available) or drops them. In the latter case a notification is sent to the source node of the burst.

This model has the following assumptions:

- Bursts have fixed length of one normalized time unit
- G is the expected number of transmissions and retransmission attempts (from all edge nodes) per time unit
- S is the number of successful received bursts. It is also the network throughput.
- The Offered (new and retransmitted bursts) load is modeled as a Poisson process with rate G.

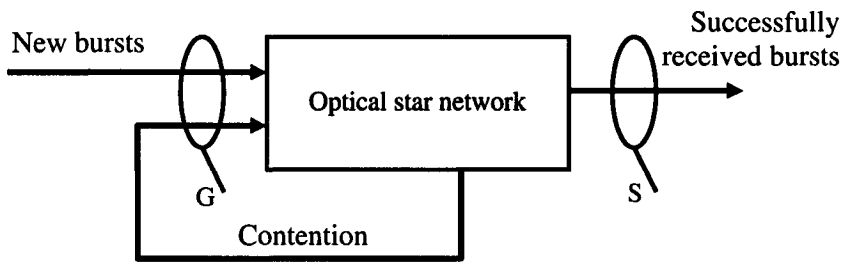


Figure 5.2: A star optical network with retransmission scheme

According to this model the probability to have k bursts generated in t frame times P_{kt} is given by the formula: $P_{kt}= \frac{(G\ t)^k}{K!} .e^{-Gt}$ (1)

No contention means there is only one burst or no burst in a period of time. That is, the probability to have one burst or no burst in one frame time. Let this probability be Pf.

According to (1) $P_f= \frac{(G)^0}{0!} .e^{-G} + \frac{(G\ )^1}{1!} .e^{-G} = e^{-G} + G\ .e^{-G}$

The contention probability is $p=1-(e^{-G} + G\ .e^{-G})$ (2)

The probability to transmit a burst in exactly n transmissions is $p_n = p^{n-1}.(1-p)$.

The approximate average number of transmissions of a burst $\overline{N}_r$ is given by

$$\overline{N}_r = \sum_{n=1}^{\infty} n \cdot p_n = \sum_{n=1}^{\infty} n \cdot p^{n-1}.(1\text{-p}) \qquad \text{that is } \overline{N}_r = \frac{1}{1-p} \qquad (3)$$

It is clear according to the formula (3) (also intuitively) that the number of retransmissions increases with the loss rate.

5.5.Simulation results and analysis

In order to evaluate the performance of the proposed congestion avoidance and retransmission scheme, we performed a number of simulations on a star network. In this simulation we consider a star topology with 8 nodes besides the core node as shown in Figure 5.3. In this model, it is assumed that each fiber carries the same number of wavelengths. All the links are bi-directional, wavelength channels are operating at 2.5 Gbps (one wavelength is used for the control channel). We assume that all the fibers have the same length. The edge nodes can send traffic to all the other edge nodes and receive as well. The core node forwards bursts to their destinations. The switching time and the processing time of a control packet in the core node are set to 5 μs. Also it is assumed that no buffers and no wavelength conversion are used in the core node. The destination of each burst is selected at random from a uniform distribution across the nodes.

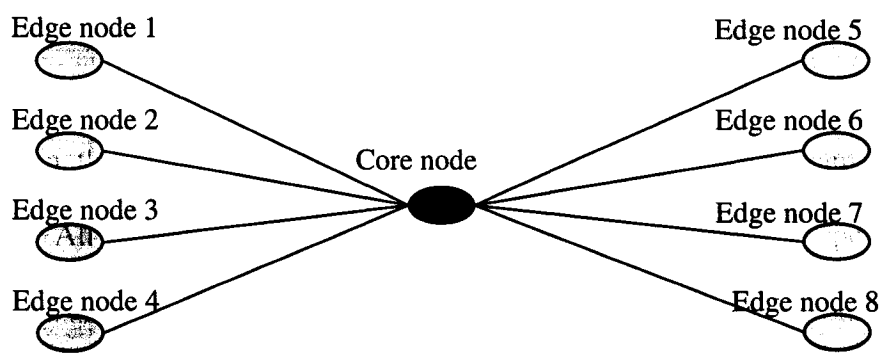


Figure 5.3: A star network with 8 edge nodes

First, in order to determine the load limit for this network, we consider a simulation where each node generates bursts according to a Poisson distribution (burst arrival) where the burst length is 40 μs (100Kb at 2.5 Gbps). Each node is equipped with a burst generator. The inter-arrival time is varied and the loss probability is analyzed for each load. Figure 5.4 shows the loss rate versus the load. As we mentioned before the loss keeps increasing as the load gets higher (this result conforms with the formula (2)). The load limit is a design parameter that determines the loss rate that the network designers are willing to accept. In this simulation, the critical loss considered is 20%.

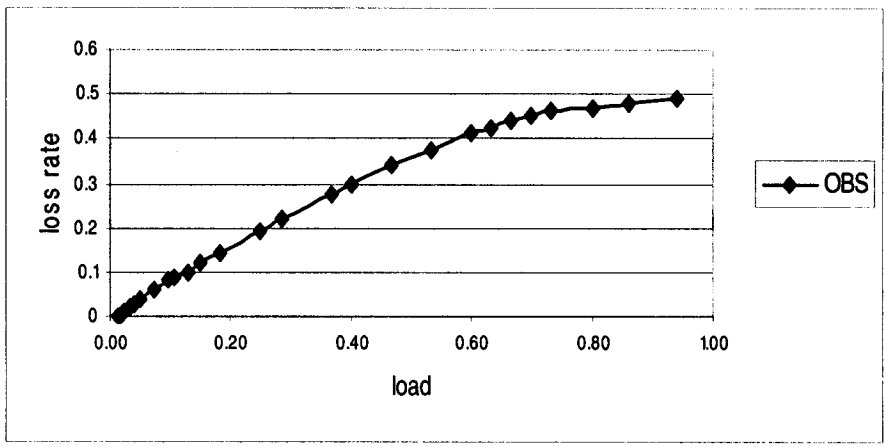


Figure 5.4: Loss rate as function of load

In the second simulation, we test the performance of the proposed congestion control scheme against OBS without congestion control (normal OBS). The performance metric we use for this purpose is burst loss rate. In this model, the edge nodes are receiving traffic (they receive packets and convert them into bursts). The external traffic is feeding the nodes' buffers. These packets, in turn, are aggregated into bursts to be sent to the core network. In the case of OBS without congestion control, the burst are assembled and sent immediately into the network. The average inter-arrival time is increased or decreased to

reduce the buffer length. Whereas in case of OBS with congestion control, the inter-arrival time is adjusted according to the statistics received from the network. All the network nodes receive traffic. Initially, the burst generator in every node is operating with an inter-arrival time corresponding to the load limit (this is for OBS with congestion control).

The burst generation is Poisson distributed with exponential burst length. Initially the average inter-arrival time, for all nodes, is 140 ms. When a node has more traffic and the critical loss is below the critical value, it could decrease the inter-arrival time of its burst generators by 5 ms. Therefore more bursts are sent, which increases the node load. In this simulation the load variation is performed by changing the inter-arrival time, which is decreased by 5 ms is case of the inter-arrival time is larger than 140 ms and the loss rate is higher than the critical value.

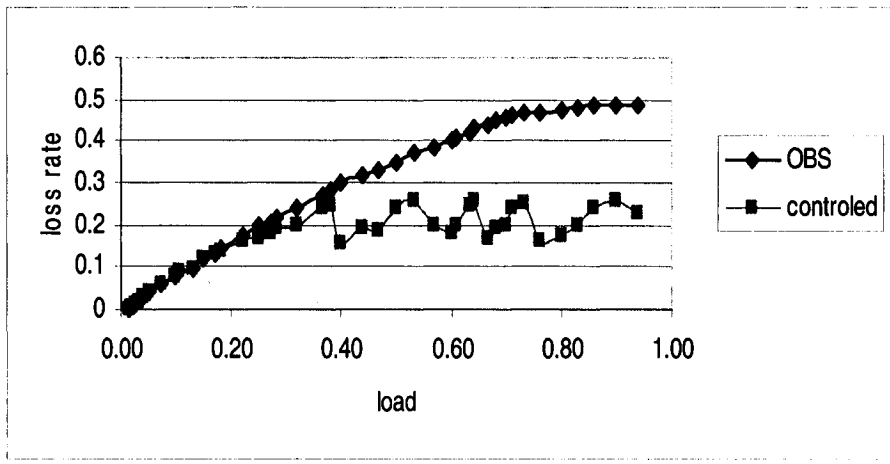


Figure 5.5: Loss rate as function of load with and without congestion control

Figure 5.5 shows the loss rate with and without congestion control with progressive adjustment (when the loss rate is higher than the critical one all the nodes with sending traffic larger than the load limit decrease their load by increasing the inter-arrival time of

their generator by 5 ms). In this case we do not retransmit the lost bursts. The loss of optical burst switching with congestion control keeps the loss lower (around the critical loss). The oscillation observed is due to the fact that the nodes sent their report only after a certain number of bursts are dropped (in this simulation, a notification is sent by a node when 3 bursts have been dropped)

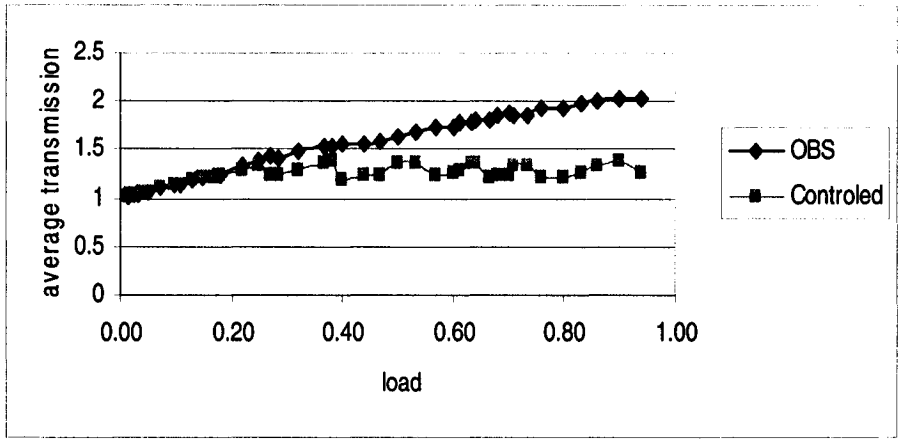


Figure 5.6: Average number of transmissions

We also investigate the average number of retransmissions required to send a burst using the retransmission scheme. Figure 5.6 shows the average number of transmissions with or without congestion control. For OBS without congestion control the number of retransmission increases as the load increases. However, for OBS with congestion control the average number is around a constant value which is below 1.25 retransmissions per burst. These results are conform with the formula (3).

The delay increases linearly with the number of retransmissions. A burst retransmitted n times is delayed by n*T where T is a round trip delay. For a very wide network T maybe very significant. Therefore n should be very small to keep the delivery delay acceptable. However, in local or metropolitan networks the propagation delay is

relatively small. In this context the retransmission scheme is very efficient. The roundtrip and the propagation delays are very small

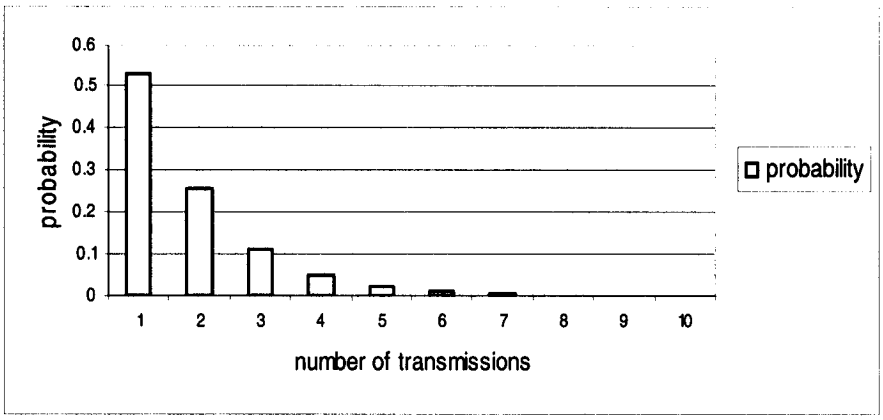


Figure 5.7: Distribution of the number of retransmissions

Figure 5.7 shows the distribution of the number of retransmissions. These results are for high traffic load with controlled load. The probability decreases quickly; only small number of retransmissions is needed to deliver a burst to its destination

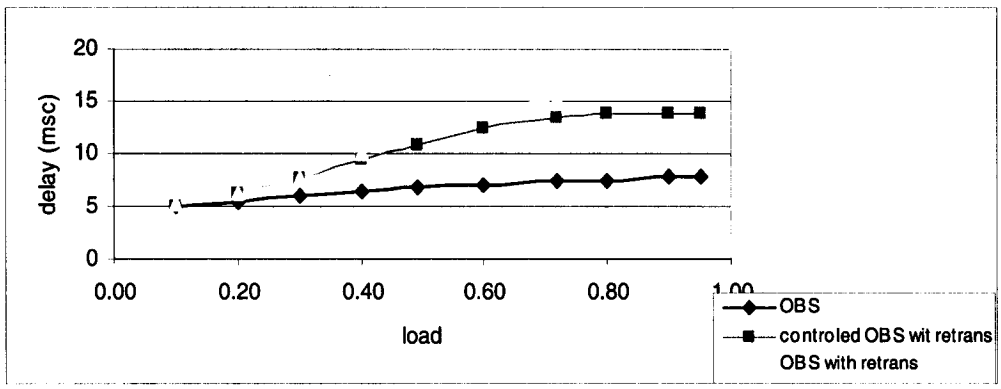


Figure 5.8: Average delay per burst

Figure 5.8 shows the delivery delay for OBS and OBS with retransmission with or without congestion control. This delay takes into account both queuing and propagation

delay. In this simulation all the fiber links have the same length (500 km each, the propagation delay from an edge node to another is 5 ms). The delay of OBS is smaller because there is no retransmission and delay is only for those bursts that reach their destination. In fact, the real delay should take into account the retransmission from a source of a dropped burst which will be longer. The delay of OBS with retransmission and congestion control is better than the one without control. This is because without congestion a burst may be retransmitted several times before it reaches its destination.

5.6.Summary

In this chapter, we proposed a loss-free scheme for optical burst switching. This technique aims to cope completely with the loss and guarantee burst delivery. First we reduced the contention by controlling the load and avoiding congestion. The contention reduction relies on the intermediate nodes to send statistics about the loss inside the network and on the edge nodes to adjust their traffic accordingly. Since the proposed traffic control does not eliminate the losses, we propose another extension that aims to retransmit all the dropped bursts. A source is notified if one of its bursts is dropped and proceeds to its retransmission. The simulation results show that the combination of these two techniques leads to a robust optical burst switching where bursts suffer no loss.

The congestion control, that we propose to reduce the loss, is very similar to the admission control used in other network architectures. The common principle is that the network resources are limited and one cannot go beyond this capacity. Therefore one needs to share these resources efficiently to avoid contention. One way to perform this sharing is by deploying time division multiplexing and use the slot to carry traffic

required by different nodes. The next chapter presents more details about this architecture.

Chapter 6

A Bandwidth allocation scheme in optical TDM network

6.1.Introduction

Wavelength Routing (WR) and Optical Burst Switching (OBS) are two optical network techniques that have received enormous attention over the last decade. However, the two techniques are plagued with many problems. The main concern with WR is the inefficient bandwidth utilization. On the other hand, the problem with OBS is resource contention and burst dropping.

To avoid the contention resulting from optical burst switching and increase the network throughput at the same time, we propose [97] in this chapter a new bandwidth allocation scheme and switch architecture that uses slotted switching with flow reservation, which is a form of TDM. Slotted switching techniques were reported in [98,99]. Slot mapping and assignment schemes without using optical buffering in a TDM network are described in [98]. A switch architecture, using Optical Time-Slot Interchangers (OTSI) to perform timeslot switching, is presented in [99]. The OTSI is made of a set of optical crossbars and a number of variable size delay lines in order to delay data by a number of timeslots. The three basic characteristics that would affect the cost and performance of an OTSI are the size of its internal crossbar, the amount of fibers needed for delay lines to reorder the timeslots, and the number of switching operations needed to be performed on a timeslot within the OTSI. The result of the study reported in

[99] shows that an OTSI, with 4 delay lines of total delays equal to 15 μs, would provide excellent performance under heavy traffic load in a time slotted architecture, where the frame is made of 64 timeslots of 1 μs each. In addition, the average number of switching operations per hop is 3. In our scheme, we propose a similar slot delaying technique, but with a different architecture. We called the employed device an Optical Time Slot Sequencer, or simply a Sequencer.

In our study, we rely on the Labeled Switch Path (LSP) concept to route traffic from source to destination. An LSP corresponds to the reservation of one timeslot per TDM frame. We allow an edge node to reserve a group of LSP to accommodate the transmitted traffic that might need more than one timeslots per frame to a particular destination, and sometimes more than one designated path. An LSP is identified by the path, wavelength on which data travels, and the slot position on each link. Note that, along the path, the timeslot position, assigned at the source, might change due to propagation and switching delays. An LSP group consists of one or more LSPs sharing the same path and wavelength between a particular source destination pair. A Flow is a set of LSP groups riding on different wavelengths and/or different paths. Routing at intermediate nodes is carried out on a time slot basis. Thus, switching from slot to slot must be fast enough to fit the narrow guard time that separates slot boundaries [98]. A further challenge for slotted traffic is the global synchronization. We aim to handle the challenge by assuming the length of fibers corresponds to propagation delays that are multiples of the slot size [98]; and, the clocks should be synchronized to tick in timeslot units. In addition, to account for the variable delay induced by the change in temperature, we need to dynamically align the incoming slot boundaries to the local clock. This can be achieved

using a set of variable size delay lines with an optical switching component to form a Synchronizer [99]. The complexity of the Synchronizer is reduced in the case of star networks.

Besides its capability of avoiding contention and improving bandwidth utilization, like Slotted OBS, the proposed scheme and architecture can carry TDM traffic that is aggregated at the edge nodes. Furthermore, the proposed technique becomes wavelength routed when all the timeslots in a frame are exclusively used for one end-to-end connection between a source-destination pair.

6.2.Resource allocation and switch architecture

In the proposed scheme, the source node decides on the number of timeslots that need to be periodically transmitted to a certain destination. The number of requested timeslots over a given wavelength must be less or equal to N, the number of slots per frame. Afterwards, it engages in a reservation scheme to allocate the network resources that are essential to transport the stream traffic. The resources are reserved to serve the LSPs originating at the source, based on the requested bandwidth and its availability in various paths. The LSPs are organized in groups to form a flow as defined earlier. If the LSP group riding on a wavelength along the shortest path covers only part of the needed bandwidth, alternative LSP groups are checked to accommodate the remaining bandwidth (LSPs are allocated on alternate paths). We propose a simple reservation method that relies solely on the lowest amount of available bandwidth in the physical links of a given path. The link that has the lowest bandwidth among the rest dictates the amount of traffic that can ride on the path, i.e. the size of the LSP group. The reservation method is not

concerned with the timeslot positions since the time slotted traffic will be re-sequenced at intermediate nodes to fill up the empty slots. Thus, the source node is free to choose the slot positions for a reserved LSP group, knowing that the selected position may change along the route based on the various delays and the availability of a corresponding slot. After the reservation phase, presumably quick enough, the transmitted time-slotted traffic on the reserved LSPs follows a schedule along the intermediate nodes. If an incoming time-slot reaches an intermediate node when another slot is using the outgoing link, then the schedule forsees that the timeslot has to be delayed an adequate amount of time to be mapped to an available slot position. For this purpose, a scheduling algorithm is performed when an LSP is established to decide the amount of delays needed at each hop. Once the mapping table is defined by the scheduler, it will be used during the lifetime of the LSP to direct the incoming timeslot to the appropriate entry in the Sequencer, and hence produce the needed delay.

By adopting the proposed scheme, we achieve higher bandwidth utilization (compared to wavelength routing) and avoid contention (observed in OBS). The reservation and scheduling methods are simple and straightforward making the set-up time minimal. Compared with OBS, we discard the need for a data header and rely on the slot position to identify an LSP. In our scheme, a switch forwards the timeslot of an LSP based on its position in the frame. Based on this architecture, multiple classes of services become possible by selecting the number of timeslots to be transmitted from a source to a destination. For instance, high priority traffic can be transmitted in a flow of many LSP groups; and low priority data can travel on a group of a few LSPs. Stating the merits leaves us with one major drawback that need to be resolved, which is the problem of out-

of-order delivery. Since the traffic is slotted and can be sent through multiple LSPs, some slots might reach the destination faster than the others due to the varying delivery delays based on the different propagation and switching delays occurring on different LSPs. Even time slots riding the same LSPG are not guaranteed to reach their destination in the order since time slots on different LSP may suffer different delays at intermediate nodes. Hence, the destination node has to cope with reordering the traffic to match its original pattern. A possible solution is to add a sequence number to the header of each traffic segment, corresponding to a timeslot, before sending it through the optical domain.

6.2.1. Switch architecture

The architecture of an intermediate switch is shown in Figure 6.1. The figure shows only one wavelength per fiber link. In addition, each link carries a dedicated control wavelength that goes to the switch controller, where the control data gets converted into the electrical domain. As shown, every output port is coupled with a Sequencer used to align the slotted data in order to prevent link contention. Each sequencer is connected to the output side of the optical cross-connect via N input lines. The admission control module is responsible for the signaling to decide the amount of bandwidth available on a given path in order to assist the source node to decide on the size and number of LSP groups needed to transport the given traffic. Its main responsibility is to define the lowest amount of available bandwidth on the set of links making a path, and propagate the information back to the source. It can also handle other functions such as forecasting the average delivery delay. The admission control module is used during the reservation phase. On the other hand, the scheduler module is responsible for setting the amount of delays needed per timeslot to map it to an available position in an outgoing frame. Once

the amount of delay is defined for an LSP, the scheduler reserves the appropriate resources in the corresponding Sequencer, and updates the next intermediate node about the new slot position. This exercise is performed when the resource reservation for a new flow is made and the corresponding LSPs are established. During the transmission process, the scheduler periodically instructs the switching element to direct outgoing timeslot to the appropriate FDL entry in order to produce the needed delay.

In order to be scheduled at an intermediate node, a LSP should provide some information to the node controller, such as the input time slot, and the output port. One can use a signaling protocol to carry this information.

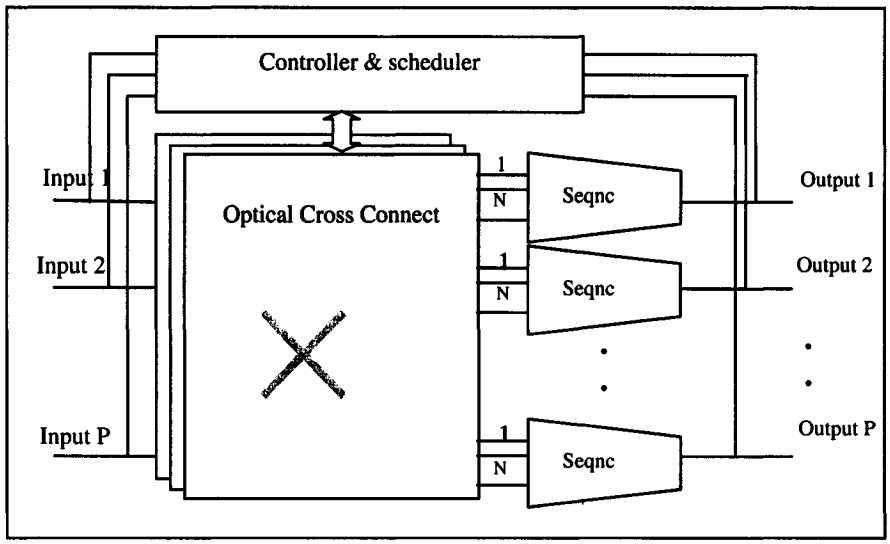


Figure 6.1: Switch architecture

The proposed Sequencer is a multi-input queue of N sequentially connected fiber delay lines (FDLs), each matching a timeslot size; see Figure 6.2. The delay imposed by every FDL is exactly equal to the time slot period. A Sequencer of size N has N inputs connected to a switching component and 1 output. Every input leads to the beginning of one of the sequentially connected FDLs. After a slot enters a designated FDL at the queue

position j, it moves across all the subsequent FDLs in the queue until it reaches the shared output link. In this case, the slot experiences a delay in the sequencer equal to j multiplied by the time slot period T. Note that the first sequencer entry at position 0 goes directly to the output without being delayed. Generally, an FDL position in the sequencer is selected based on the incoming traffic that shares the output link. It is determined by considering all the timeslots that share the same output resource (i.e. link and wavelength). For instance, if K slots are in line to be sent to the output link, the new slot is assigned to the Kth FDL position in the sequencer.

The unique characteristic of the Sequencer is that the number of switchings required for a timeslot per hop is always 1. On the other hand, with the optimized Optical Time-Slot Interchanger proposed in [99], the average number of switching per hop is equal to 3. An OTSI having the same characteristic of the Sequencer is possible. However, the total length of the needed delay lines is equal to $N^2/2$, arranged in N lines of length 1, 2, .. N. Meanwhile, the total length of the single delay line employed in the Sequencer is equal to N.

The switch proposed in [99] has the same number of outputs and inputs compared to that of the switch we are proposing. However, the former deploys an active component to delay a slot, which increases its complexity.

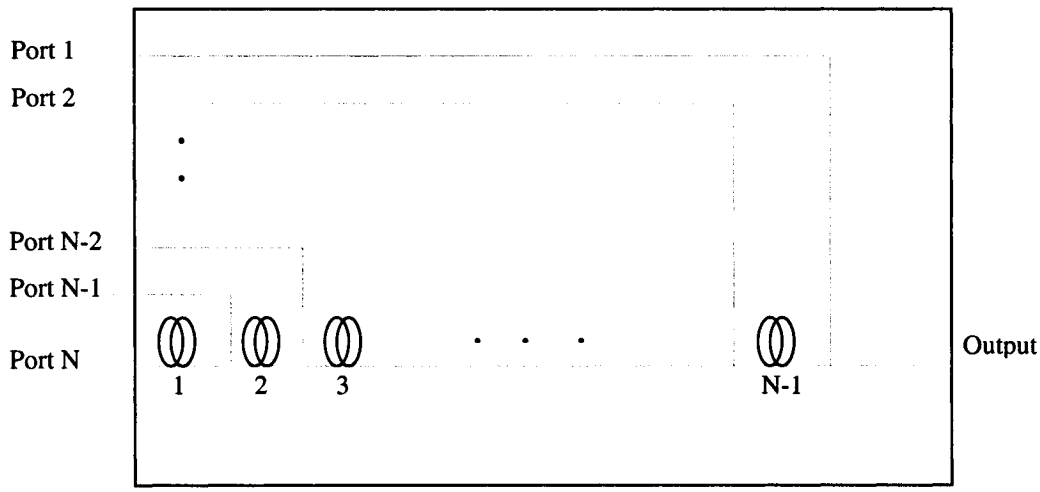


Figure 6.2: Sequencer

To reduce the switch complexity, where an optical cross-connect of $m \times N$ outputs is required (N = number of timeslots per frame, and m = number of output fibers), we propose an alternative approach. We place the Sequencer at the input side of the switch, and connect it to the input fiber by a small $1 \times N$ optical cross-connect, whose role is to direct incoming timeslots to the appropriate entry in the sequencer. See Figure 6.3. In this architecture, the main cross-connect of the switch remains simple. The price of this approach is the introduction of some possibility of blocking. For instance, blocking will occur when 2 consecutive timeslots on an input link are to be switched to the same timeslot position, but in two different outgoing links. Both timeslots get assigned to two consecutive FDL positions. However, the second slot cannot go through its assigned output, since this position would be used by the first after moving one position forward.

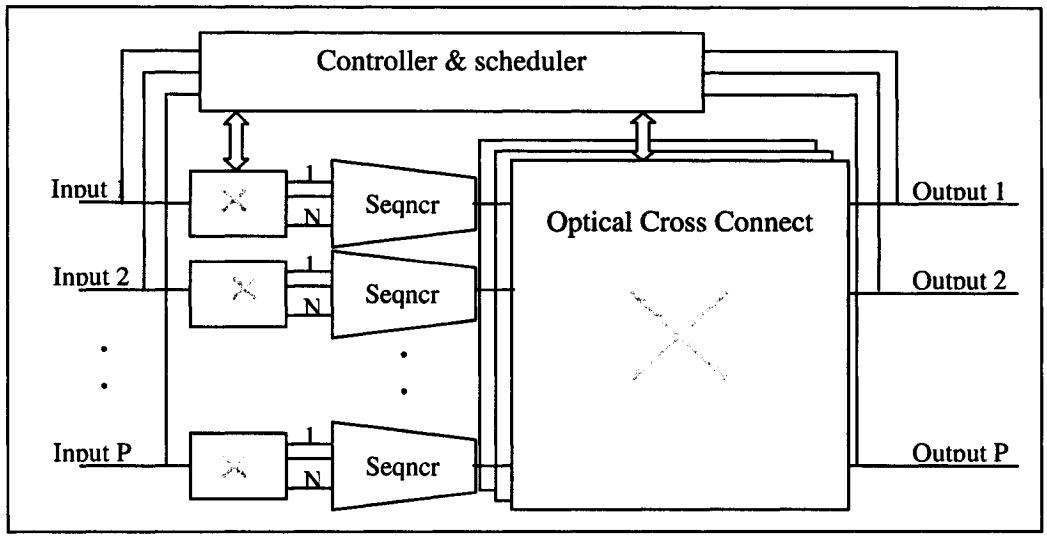


Figure 6.3: Alternative switch architecture

6.2.2. Flow reservation

In this simple reservation scheme, a flow of LSP groups (LSPGs) is established from source to destination before traffic transmission. The size Z of an LSPG, representing the number of its LSPs, is an integer number bounded by the lowest available bandwidth on the set of links forming the path. We use the term "available bandwidth" to describe the number of available timeslots per frame on a given link. A path P, made of a sequence of links $l_1 l_2 ... l_p$, provides a maximum bandwidth Bmax (in terms of timeslots), where $B_{\max} = Min(B_1, B_2, ..., B_p)$ and Bj is the available bandwidth on link lj. The size of a possible LSPG running on P can be between 1 and Bmax ($1 \leq Z \leq$ Bmax). A source node S willing to periodically transmit n timeslots of data to a destination D would require n LSPs ($n \leq N$), organized in a flow of k LSPGs such that $n = \sum_{i=1}^{k} Z_i$ where Zi is the size of the ith LSPG.

To reserve a bandwidth capacity equal to Breq (slots) in the network, a source node starts with the shortest path P1 first, which has a maximum number of available slots equal to B1. If Breq ≤ B1, then the source node reserves a single LSPG1 consisting of a number of LSPs equal to Breq. In addition, the bandwidth Bj on every link lj, forming the path, becomes equal to Bj - Breq (or Bj = Bj – Breq). If Breq > B1, then the source node reserves an LSPG of size B1, and proceeds with considering the second shortest path P2. If Breq – B1 ≤ B2, then an LSPG2 is reserved with a size equal to Breq – B1. Otherwise, the same procedure is repeated for k alternative paths P1, P2, ... , Pk, where k ≤ Breq, until $B_{Req} - \sum_{j=1}^{k-1} B_j \le B_k$. If $k = B_{Req}$ *and* $B_{Req} - \sum_{j=1}^{k-1} B_j > B_k$, then the request can be partially granted to accommodate a number of slots equal to $\sum_{j=1}^{k} B_j$; while the remainder of the requested bandwidth, derived by $B_{Req} - \sum_{j=1}^{k} B_j$, is blocked.

6.2.3. Basic scheduling

Sophisticated scheduling algorithms can be designed to manage the timeslot switching and delaying in the network (i.e. Sequencer delays). In these algorithms, many parameters can be considered such as the number and size of Sequencers, and the maximum tolerated delivery delay. In this section, we propose a basic scheduling algorithm, assuming that every output link is connected to a dedicated sequencer of size N. A logical vector V of size N (or bitmap) represents the states of all the timeslots in the outgoing link. If the ith position in V is set to 1, it means that the corresponding timeslots on the outgoing link is reserved for usage by a cross-connected incoming link. To know

the availability of the i^{th} timeslot position in an outgoing link, we perform the logical OR operation on V and 2^i. If the answer is 0, it indicates that the i^{th} timeslot position on the outgoing link is free. An LSP, riding on timeslot j, must be switched to another timeslot if the j^{th} position is not available on the outgoing link. In this case, once the scheduler identifies the first available position i, starting from the least significant bit in V, it sets the i^{th} position in the corresponding logical vector to 1. Based on this mapping procedure, it assigns the traffic arriving on the j^{th} timeslot to the $(|i-j|+1)^{th}$ FDL in the Sequencer of the outgoing link. Hence, the j^{th} incoming timeslot is delayed by a period equal to ((i - j + N) mod N) x T, where T is the timeslot duration. Note that the position of the incoming timeslot is relative to the local clock at the intermediate switch. Although a node transmits a timeslot at position x, the adjacent node on the path might see the timeslot at position y due to the propagation delay t in the link, [y = (x + t) mod N]. This scheduling process is done only once after the reservation in order to define the mapping tables, which will be used by the controller to forward timeslots to the appropriate positions in the sequencer.

The delivery delay DL imposed by the proposed scheme is equal to the distance propagation delay PD added to the total delays incurred at the intermediate Sequencers. If M is the number of intermediate nodes that an LSP crosses, then the maximum delivery delay is $PD + (M \times (N-1) \times T)$. For instance, if the timeslot duration is 10 μs, the frame is composed of 100 slots, and the number of intermediate nodes is 10, the maximum delay will be 10 ms on top of the propagation delay (5ms/1000km).

As an example, consider an intermediate switch having 3 inputs, 2 outputs, and one wavelength λ available for service (Figure 6.4). We need to setup a new LSPG on input

In3, which passes through the output Out1 and consists of 4 LSPs. At each input link, we show a bitmap representing the timeslots states of the incoming frame. In addition, we use the logical vectors V1 and V2 to reflect the timeslots availability in Out1 and Out2 respectively. Initially, we have some traffic on the input links In1 and In2 that share the output links Out1, and Out2. At the input In1, the first 2 timeslots are mapped to the same timeslot positions in the output Out1, and the 6th in In1 is set for the 6th position in Out2. For this reason, the corresponding bits are set to 1 in V1 and V2. A similar representation is done for the traffic arriving at input In2, where the 5th timeslot is mapped to the same position in Out1, and the 6th is set for the 7th position in Out2. Note that the total number of reserved timeslots at Out1 is 3 as shown by V1. It indicates that the node can accommodate the 4 LSPs of the requested LSPG. This step is done at the reservation phase; however, we mention it for clarification purposes. The arriving frame on In3, carrying the LSPs of the requested LSPG, is represented by the bitmap [10000111] assuming that In3 did not have any traffic previously. Considering the 1st timeslot, we notice that it can go through the 3rd timeslot in Out1. Therefore, the 3rd bit needs to be reserved in the logical vector V1. The amount of Sequencer delay in this case is 2 (i.e. 3-1). Thus, the 1st incoming timeslot will be switched to the 3rd FDL entry in the corresponding Sequencer. We carry out the same procedure with every LSP of the considered LSPG to book its corresponding output timeslot and define its Sequencer delay. To see the mapping of all LSPs from In3 to Out1, consult the table included at the bottom of the figure. In addition, you find the new bit values in the vector V1 included next to the table itself.

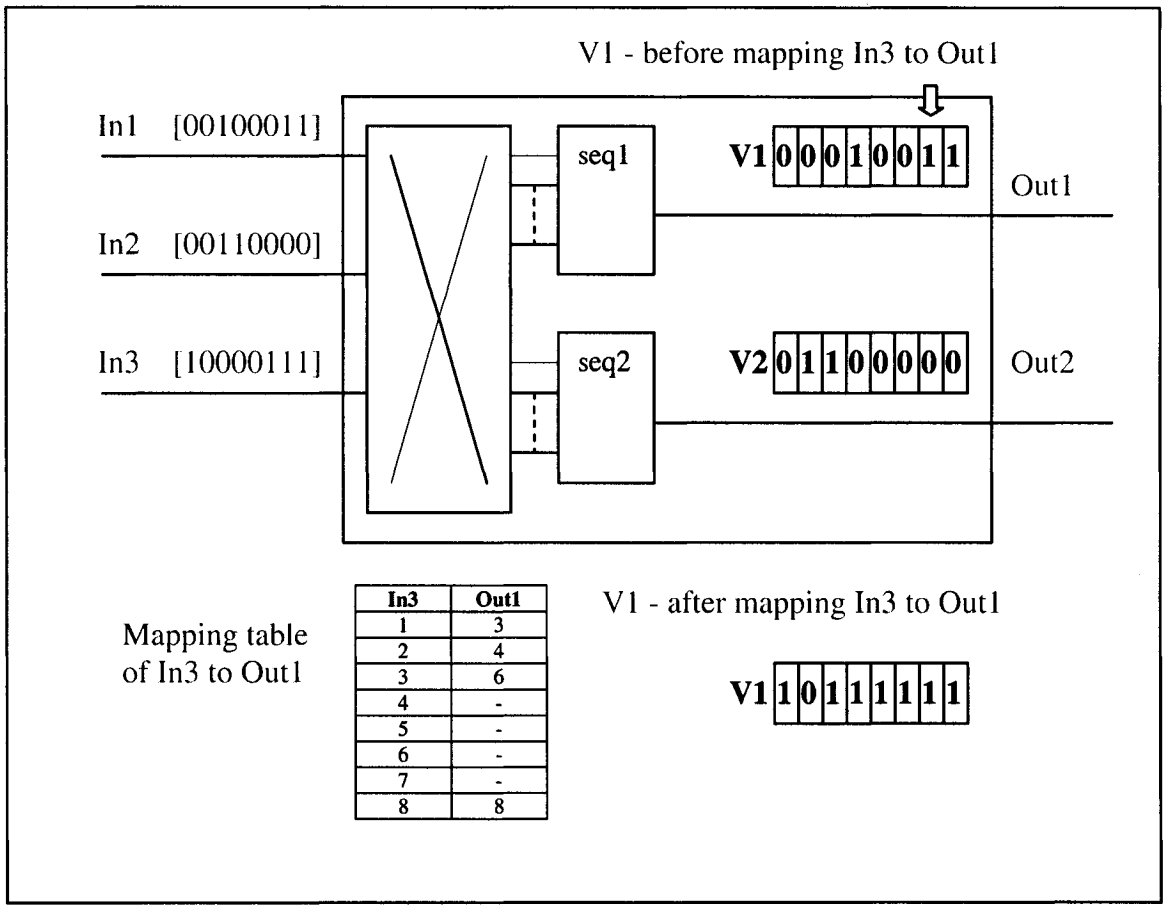


Figure 6.4: Scheduling example

6.3.Optical wavelength conversion and optical slot delaying

We may think of the optical slot sequencer in the same sense we understand the optical wavelength converter. The slot sequencer delays traffic arriving on a certain time slot by a period of n time slots where $1 \leq n \leq d$; d is the number of fiber delay lines that exists in the slot sequencer. On the other hand, the optical wavelength converter shifts traffic riding on a given wavelength λ to another wavelength λ' where λ' belongs to a spectrum of nearby wavelengths [101,102]. The spectrum of nearby wavelengths is a range of r possible frequencies to which λ can be converted. If we consider a network with one

wavelength only, where the bandwidth is shared by dividing it into N timeslots, a sequencer of d fiber delay lines is an analogue to a wavelength converter converting to a range of d nearby frequencies. Instead of converting between nearby channels, the device will be converting between nearby timeslots. Based on the similarity in the behavior of both devices, we expect that the performance analysis of wavelength converters, as reported in the literature [102,103,104], holds true in the case of time slot delayers.

The performance of all-optical networks with wavelength conversion has been studied since the mid 90s. The main factors that hinder the deployment of such devices are their high cost and current immaturity. However, it has been proven in many experimental studies that employing these devices with the appropriate bandwidth allocation scheme can yield a substantial improvement in network performance. "The use of wavelength conversion can considerably reduce the blocking of the network, but there is minimal difference in the wavelength requirements" [101].

As we noted earlier, the results of many years of studies investigating the effect of employing the optical conversion technology in all-optical networks, can be adopted to describe the effect of using the Sequencer in slotted optical networks. Hence, the expected performance improvement induced by using the slot sequencers in TDM optical networks will be similar to the one resulting from employing the wavelength converters in WDM networks. In addition, it has been proven that the number of converters and the conversion range can be drastically reduced to a certain threshold without affecting the network performance [103,104]. Similarly, we expect a reduction in the number of Sequencers and the delay range that maintain the same network performance up to a certain threshold

6.4.Simulation result and analysis

We studied the performance of the proposed architecture by means of network simulations, using the NSFNET topology with 14 nodes as shown in Figure 6.5. We assumed that each single fiber link is bi-directional and has the same number of wavelengths operating at 50 Gbps. The distance of each fiber link is shown in the network graph of Figure 6.5. One of the available wavelengths is initially reserved for signaling and control traffic. Each wavelength is divided into 50 small timeslots (circuits) of 1 Gbps each. The propagation delay between two connected nodes ranges between 1.5 and 14 ms. In the network, a node can route, generate, and receive traffic. A source node is responsible for segmenting the incoming data into timeslots for transmission, and re-assembling the slots upon reception. The traffic is uniformly distributed across the network. In the simulation, we used Dijkstra's algorithm to establish a shortest light-path between source and destination to carry a set of LSPs packed in to one LSPG. Since one path might not provide all the required bandwidth, we may establish more than one alternative path to accommodate all timeslots in an end-to-end flow. We employ a full sequencer (size 50) at each node, where a slot can be delayed for a period of time ranging between 1 and 49 slots times. We do not employ conventional buffers or wavelength converters in the simulated network architecture. Different requests are generated. Each request carries a number of slots ranging from 1 up to the number of slot in the frame (50 in this case). We ran the same simulation under two scenarios; first, we limit the request size to be on average equal to 50% of the frame size. In the second case, we decrease the average to 10%.

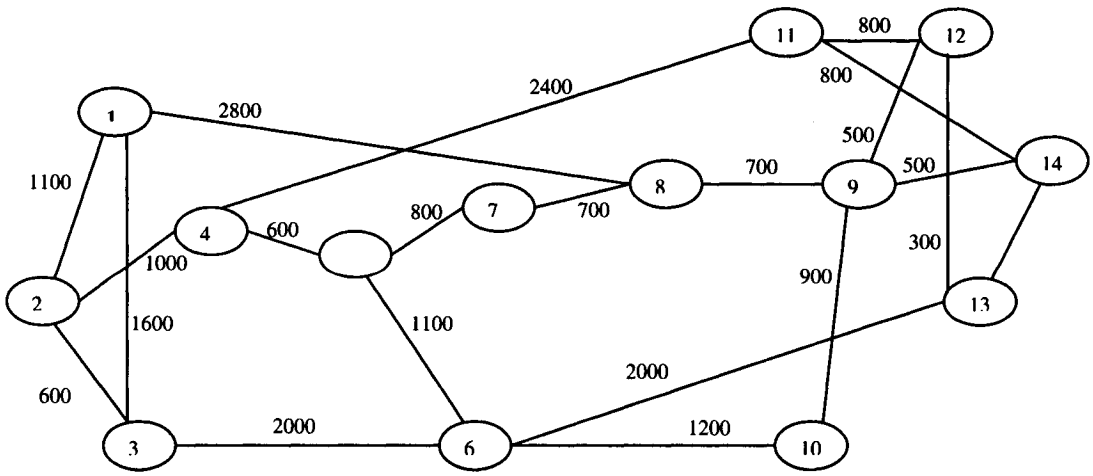
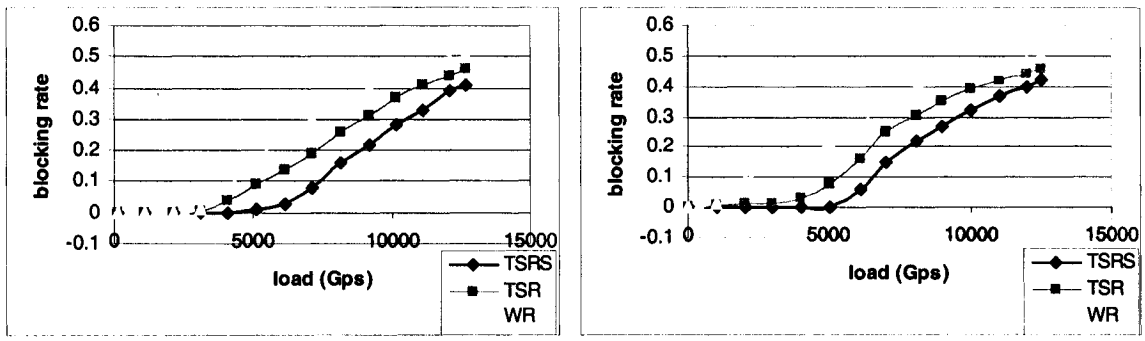


Figure 6.5: NSFNET topology with 14 nodes

The goal of the experiment is to study the performance of our proposed scheme, time slot routing with Sequencer (TSRS), compared to a wavelength routed optical network (WR). In this study, we also investigate another variant of the bandwidth allocation scheme where we remove the Sequencers (TSR).

The performance is measured using various metrics. The first metric is the blocking ratio, which reflects the percentage of traffic that must be discarded due to lack of resources. The charts (see Figure 6.6) show the blocking ratio versus the traffic load in 3 cases: WR, TSR, TSRS. As the traffic load increases, the charts show that TSRS accommodates more traffic than the other methods. In addition, TSR performs better than WR. With a traffic load ranging between 4000 and 6000 Gbps, TSRS maintains a zero blocking ratio; while TSR and WR blocked around 10 and 20% of the traffic, respectively. When the load per request is lighter (Figure 6.6-b), the wavelength routing technique turns down more traffic. This results from the excessive use of resources and inefficient bandwidth allocation with respect to the requested load. However, TSRS and

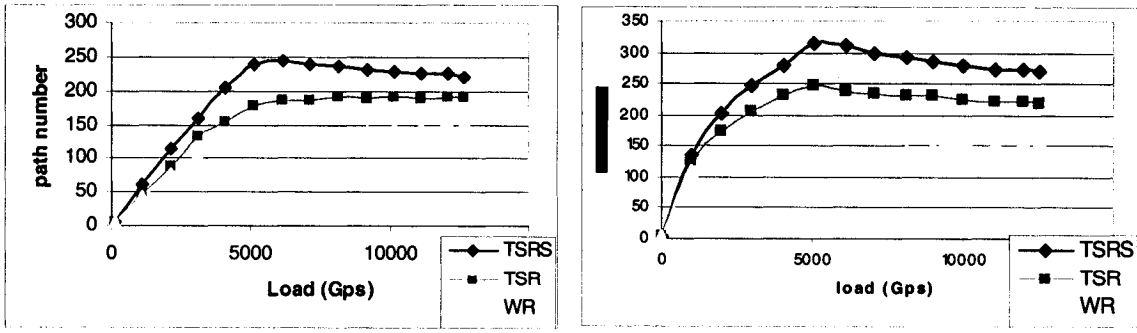
TSR are more stable and their performance is not affected by the granularity of the requested bandwidth. The difference in performance between TSR and WR stems from the efficiency of bandwidth utilization. While WR exclusively utilizes a full channel to transport a load equal to a fraction of its bandwidth, TSR makes better use of the channel by sharing it among multiple low load connections. In addition, when WR cannot accommodate a request, all the requested bandwidth is blocked. However, TSR in this case can accept part of the requested bandwidth that can be accommodated, and block the rest.



a- the average request size 50 % **b-** the average request size 10 %

Figure 6.6: Blocking rate of TSRS, TSR and WR

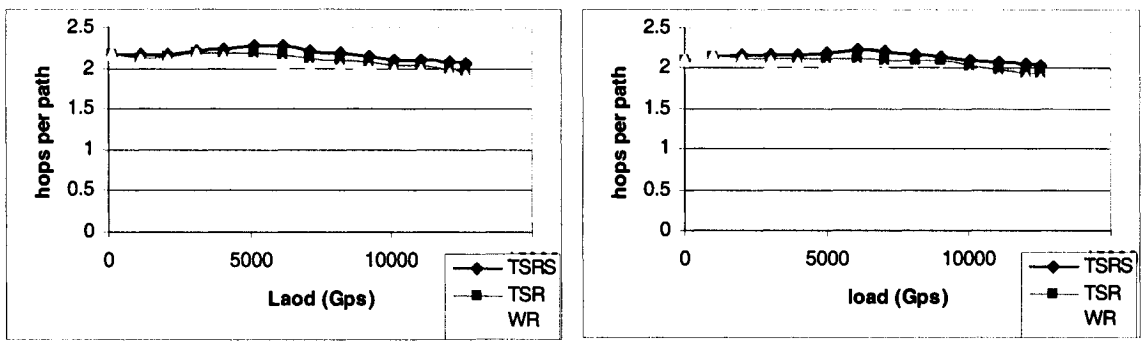
The generated charts, described in Figure 6.7, show the number of established paths Vs the traffic load in the case of TSRS, TSR and WR. It is clear that TSR uses more paths than WR. In fact, while WR uses only one path to ship traffic from a source to a destination, TSR may use more than one LSPG. However, as the traffic load increases, the number of LSPGs slightly decreases since a single LSPG tends to carry more timeslots.



a- the average request size 50 % **b-** the average request size 10 %

Figure 6.7: Average number of paths for TSRS, TSR and WR

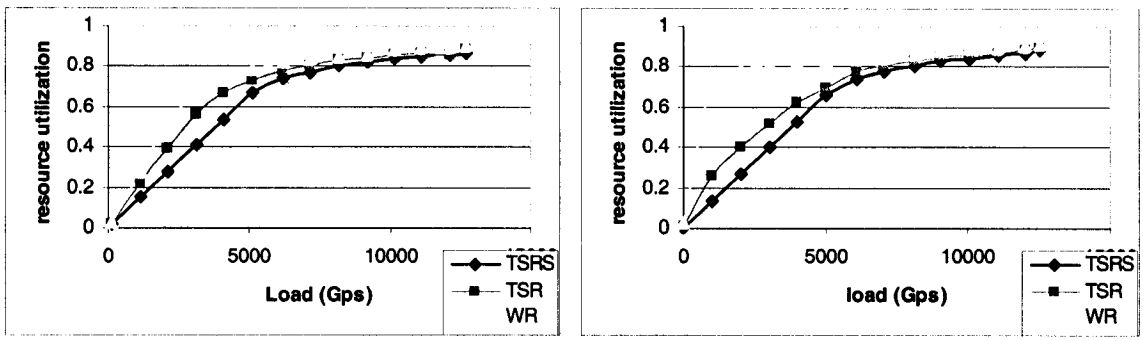
The third metric is the average number of hops crossed by the LSPGs. It reflects the average cost of establishing and tearing down a light-path (or LSPG) between a source-destination pair. The generated charts, described in Figure 6.8, show the average number of hops per path verses the traffic load. As expected, the average number of hops required for TSR is slightly higher than is required for WR. In fact, while WR uses the shortest possible path between a source-destination pair, TSR and TSRS may send traffic through multiple paths. The number of possible paths per a source-destination pair in TSR can be from 1 to 50, with a slightly higher average number of hops than the shortest path. The difference is negligible as shown in the graphs.



a- the average request size 50 % **b-** the average request size 10 %

Figure 6.8: Average number of hop versus the load for TSRS, TSR and WR

The last metric is the resource utilization ratio, which reflects the percentage of network resources used to accommodate the traffic load. The most effective network architecture has the lowest possible resource utilization ratio. The charts, described in Figure 6.9, show the ratio of used resources verses the traffic load in the case of TSRS, TSR and WR. It shows that TSRS and TSR require fewer resources to accommodate the generated traffic load, especially between a range of 2000 and 6000 Gps. The usage of Sequencers yields a slight improvement in the resource utilization ratio as the traffic load increases. The spared resources, resulting from using the Sequencers in TSRS, can be used in balancing the traffic load across the network to accommodate more traffic.



a- the average request size 50 % **b-** the average request size 10 %

Figure 6.9: Resource used by TSRS, TSR and WR

The graph in Figure 6.9-b shows the resource utilization of all the techniques with a relatively small load per request. As stated earlier, WR allocates the resources for a full channel even if it is partially used. Thus, the bandwidth utilization is severely affected by the average amount of traffic per connection (or request size). However, the resource utilization of TSR and TSRS does not depend on the request size. Rather, it grows linearly with the global traffic load. This result is due to sharing channels and resources which maximize their utilizations.

6.5.Summary

In this chapter, we proposed a new bandwidth allocation scheme and switch architecture to share network resources, avoid contention, reduce blocking ratio, and improve bandwidth utilization. The improvement in bandwidth utilization and reduction in blocking ratio were achieved by employing Sequencers, a form of Optical Time-Slot Interchangers, to delay an incoming timeslot for an adequate period of time in order to match a free outgoing timeslot. The proposed Sequencer is an array of FDLs, one feeding into the other, and connected to a switching component. The blocking ratio is improved further by a simple reservation scheme that uses multiple paths to transmit traffic. Every path corresponds to an LSP group consisting of many LSPs; each LSP is identified by its timeslot position. The contention is avoided by employing a basic scheduling algorithm to derive the amount of delay needed at intermediate sequencers, after constructing the mapping tables. In the TSR with Sequencer approach (TSRS), some QoS parameters can be easily managed and guaranteed such as the delivery delay. The delay can be sized by increasing or decreasing the number of LSPGs, modifying the number of timeslots per LSPG, and using sophisticated scheduling. The results of our study show that adopting the TSRS yields to an improved blocking probability over TSR and WR. In addition, it improves bandwidth utilization when the traffic load increases.

The main complexity of this architecture is the switch complexity and the global clock synchronization. Indeed every switch should start at the beginning of the slot.

Further work is needed to account for the traffic engineering parameters. For instance, one could study the usage of better reservation and scheduling schemes that take into account the class of service of the traffic being sent. It would be interesting to measure

the delay with respect to the traffic rate. In addition, some work could be done on optimizing the switch architecture to minimize the number and size of the Sequencers. In this context, a comparison between required optimizations of Sequencers and Wavelength Converters would be interesting.

As noted earlier, the complexity of the switch can be reduced at the expense of some possibilities of blocking. It would be interesting to study the increase of the blocking ratio introduced by simpler switch architectures.

Chapter 7

Shared protection

7.1.Introduction

Optical networks are currently deploying DWDM to meet the tremendous bandwidth requirements caused by the explosive growth in Internet traffic. This rapid development has led to the extensive use of the optical resources available for switching and routing in the second generation of optical network systems. That paved the way for an increase in research and development activities to benefit the most from the extreme speed and bandwidth of all-optical networks. However, this huge capacity makes such a network very vulnerable to any failure in the network components (links or nodes). Indeed, the failure of a network component such as a fiber cut lead to simultaneous failure of many wavelengths crossing this link. Failures may result in a large disruption in the network traffic and data streams. Therefore network survivability enhancement is a necessity. Strong line of defense should be taken and efficient techniques should be deployed for automatic recovery from failures and rerouting the traffic around a failure.

Survivability is the capability of a network to maintain service continuity in the presence of faults within the network. The degree of this survivability is determined by the network's capability to survive single failures, multiple links failures, and equipment failures.

Survivability mechanisms could be implemented either in the physical layer (by protecting physical links and paths), or throughout an entire network by involving many layers in the protection/restoration process [105,106,107]. Nevertheless, in an optical network the opportunity exists to provision and restore the failed traffic in the optical domain [108,109,110]. That adds more resilience to the network and frees the higher layers from the burden of monitoring the network for restoration purposes. Furthermore, restoration in the optical layer provides higher efficiency and higher speed.

Recovery is the sequence of events and the actions taken by a network after the detection of a failure to maintain the required service. These actions depend on the survivability strategy which could be either restoration-based, or protection-based.

In restoration-based networks, once the failure occurs, a computation is made on the fly to restore the affected light-paths from the failure. This process could be done to restore the affected light-paths only (without disrupting the existing unaffected light-paths) or by reconfiguring the whole network and switching to another virtual topology that bypasses the failed part of the network [67,68]. The latter technique, also called the reconfiguration approach, uses the resources more efficiently and has the potential to recover more traffic than the former. However, the major drawback of this technique is the disruption of the ongoing traffic (traffic unaffected by the failure).

A network protection against components failure consists of providing a backup for every flow established. However, this technique is resource consuming and the network may fail in finding a backup for every path because of the shortage of resources. Another alternative is 1:N path protection, which uses a single protection path to protect multiple working paths. More generally M protection paths could be shared among N working

paths [111,112,113]. In this technique, also called M:N protection, up to M paths could be restored at the same time (in general M is less than N). 1:N and M:N path protections make more efficient uses of network resources. This scheme is very useful when the backup paths are a shared risk group disjoint from the primary paths. Consequently the protection cannot use the links used by the primary paths. However, in a network with a huge capacity (and also high bandwidth demand) such a constraint may limit the restoration capacity of the network. Indeed, traffic from a source to a destination may use many flows following different paths and may share many links. Excluding these links may limit the protection. Therefore one needs to use another scheme that uses the whole capacity available. One needs to manage the risk rather than avoid it.

In this section, we propose a traffic protection mechanism for a network allowing for multiple flows that carry the traffic from a given source to a given destination. We propose an algorithm that identifies for every primary traffic (composed of a set of flows) another backup (also composed of a set of flows) for protection. The goal of the algorithm is to achieve the same level of protection provided by a dedicated protection mechanism while using the resources more efficiently. This algorithm relies on the fact that the primary traffic may be composed of many flows following different physical paths. Therefore the protection could be achieved more efficiently and cost effectively (in terms of resources) by provisioning a backup and sharing it among these flows (that compose the primary flow). The protection technique we propose is performed in three steps: computation of the optimal backup capacity, identification of the potential capacity that could be used by the protection mechanism on every link, and backup provisioning. Basically the traffic protection mechanism could be seen as a new request with a capacity

provided by Step 1 on resources provided by Step 2. And hence the protection could be build up by aggregating many flows leading to a flexible model, where one has more possibilities and more chances to find a backup.

We consider the specific case of the protection of an all-optical mesh network (deployment of WDM to interconnect optical switches) using both wavelength routing and time division multiplexing. Flows of slotted bursts are established between end nodes. A flow may be composed of many time slots, which we call a flow unit. Traffic from a source to a destination could be composed of many flows going through different physical paths and each carrying a different number of flow units.

In this work we consider a network as backbone providing bandwidth on demand. The main goal is to enhance the network reliability and robustness and avoid all kinds of losses (caused by contention or by components failure). The reliability should be enhanced and traffic protected. We provide the protection against the most common forms of failure that occur in optical networks including link cuts and node malfunctions. More specifically, we are interested in the protection against a single component failure (node or link) assuming that the event of failure is very rare and independent.

7.2.Network architecture and shared protection scheme

7.2.1. Shared protection

Basically to send traffic from a source to a destination, many flows maybe established (they may have different physical paths). A flow or a channel is identified by its

bandwidth capacity and its physical path. Traffic from a source to a destination may be composed of many channels each with different capacity. A single channel could be:

- A circuit in the case of circuit switching,
- A switched path in the case of packet switching.
- A label switched path in the case of MPLS.
- A virtual circuit (permanent or switched) in case of ATM.
- A wavelength in the case of wavelength routing

In this shared protection scheme [114], the flow capacity is expressed in term of flow units (e.g. if the bandwidth is defined in units of Mbps the flow unit will be 1 Mbs, if the bandwidth is defined by a number of time slots, the flow unit will be 1 time slot).

Two kinds of failures can be observed in a network: the first one is link failure, which may be caused by a fiber cut or a wavelength failure. The other one is node failure caused by some malfunction affecting a whole node. In this work, we consider a single component failure (link or node).

Recovery consists of re-routing the affected traffic by identifying alternative paths and switching the traffic over. If the recovery (back up) paths are computed and allocated in advance, the scheme is protection-oriented whereas if the recovery paths are identified after the failure occurs the scheme is restoration-oriented. Protection offers lower recovery time and guarantees the recovery. However, the redundancy required by this scheme makes it resource consuming. Restoration uses the resources more efficiently but it requires more time for recovery, restoration/protection may be link-based, where the focus is to reroute the traffic around the failed link, or path-based where the focus is to reroute the traffic from source to destination.

Furthermore, protection can be dedicated where every path should have its own backup, or shared where many primary paths can share the same backup. In general, the shared scheme is involved wherever the main concern is the bandwidth saving.

Traffic-based protection is a scheme where the whole information crossing the network from a source to a destination needs to be protected. The traffic involves many flows each carrying a different number of flow units. In the following algorithm, we propose to protect the primary traffic against a single failure using just enough resources. If the flows composing the traffic (from a source to a destination) are a disjoint shared risk group (do not share any link) then a single link failure cannot affect more than one flow. Therefore a backup path with as much capacity as the largest flow of the traffic is enough to protect the traffic. However, practically these flows could be sharing many links making it difficult to protect them separately. The algorithm analyses the shared risk in order to determine the capacity required for the traffic protection. In our protection scheme we allow the backup path to share links with the primary traffic. However, this risk is controlled and the algorithm determines how many flow units could be used by the protection mechanism on each link.

Figure 7.1 illustrates an example of how the proposed shared protection scheme works. Let us assume a network as shown, node number 2 is a source and node 14 is a destination. There are three flows (the three solid bold lines in Figure 7.1) established between the source and the destination carrying a total traffic of 5 flow units as follows:

- Flow L1: (N2, N1, N8, N9, N14) carries 2 flow units.
- Flow L2: (N2, N4, N5, N7, N8, N9, N14) carries one flow units. L1 and L2 are sharing two links $E_{8,9}$ and $E_{9,14}$.

- Flow L3: (N2, N3, N6, N13, N14) carries two flow units.

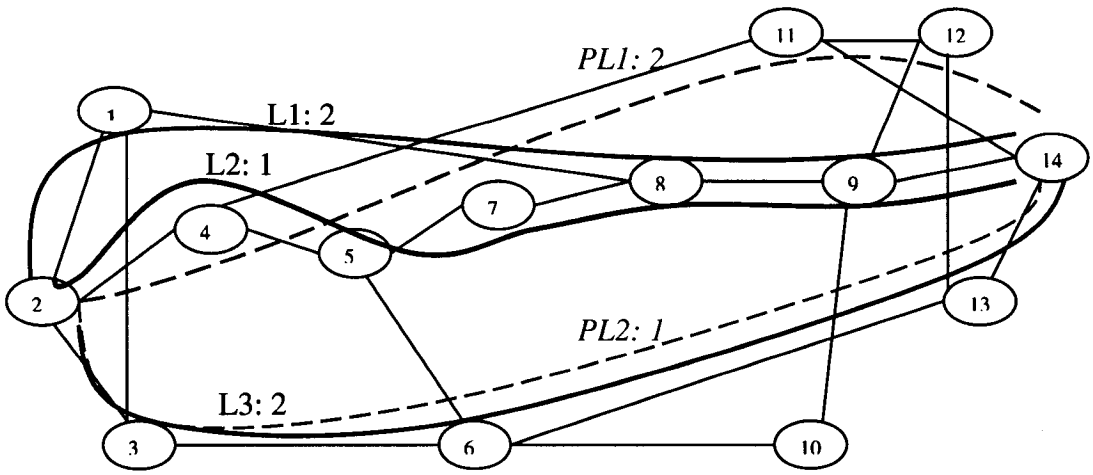


Figure 7.1: A primary traffic (continued lines) with its protection (dashed lines)

In this case a backup path carrying 3 flow units is enough to protect the traffic carried by L1, L2 and L3. This backup is composed of 2 flows as shown in Figure 7.1 (dashed line).

- Flow PL1: (N2, N4, N11, N14). PL1 shares one link $E_{2,4}$ with L2 and carries 2 flow units.

- Flow PL2: (N2, N3, N6, N13, N14). PL2 shares the whole path with L3 and carries one flow unit.

Let us assume that the link E8,9 is down. Both flows L1 and L2 will be affected and three flow units have to be restored. Nevertheless, one can use PL1 and PL2 to carry this traffic and recover from the failure.

In another scenario, if the link $E_{3,6}$ is down both the primary and the protection traffic are affected (L3 and PL2) and 2 flow units should be restored. PL1 is still working (with a capacity of 2 flow units) and could be used to carry the traffic that was carried on L3.

For every single failure one can recover and restore the traffic.

7.2.2. Single link failure protection

The following algorithm describes the steps to achieve the shared protection described in the example. In this algorithm we use the following definitions:

- **N** is the number of nodes in the network.
- $\mathbf{N_i}$ a network node identified by a number i
- $\mathbf{E_{ij}}$ an edge from the node N_i to the node N_j
- $\mathbf{T_{sd}}$ is the traffic being carried from the source N_s to the destination N_d.
- $\mathbf{PT_{sd}}$ the protection (backup) of the traffic $\mathbf{T_{sd}}$
- $\mathbf{F_{sd}}^{(i)}$ is a flow numbered i carrying flow units from the source N_s to the destination N_j. $\mathbf{T_{sd}}$ may be composed of many flows
- **Capacity**($\mathbf{F_{sd}}^{(i)}$) is the number of flow units being carried by a flow $\mathbf{F_{sd}}^{(i)}$
- $\mathbf{PR_{sd,ij}}$ the protection required for the link E_{ij} for the given traffic T_{sd}.
- $\mathbf{PA_{sd,ij}}$ the protection available on link E_{ij} for the given Traffic T_{sd}.
- $\mathbf{NS_{ij}}$ the total number of flow units available in link E_{ij}
- $\mathbf{PNS_{sd,ij}}$ the number of flow units that could be used by the protection scheme on link E_{ij}

At the time of establishing traffic flows from a source to a destination, the protection paths are also identified and resources are reserved for that purpose. This protection scheme is carried out in three phases:

a- protection capacity:

This phase consists of the identification of the protection capacity required for the backup of $\mathbf{T_{sd}}$ from the source s to the destination d. The traffic $\mathbf{T_{sd}}$ is composed of k flows.

1- for $(1 \leq i \leq N)$ and $(1 \leq j \leq N)$, $PR_{sd,ij} = 0$;

2- for all flows that compose the traffic T_{sd} ($F_{sd}^{(h)}$ with $1 \leq h \leq k$)

for $(1 \leq i \leq N)$ and $(1 \leq j \leq N)$ if E_{ij} is crossed by $F_{sd}^{(h)}$ then $PR_{sd,ij} = PR_{sd,ij}$ + capacity ($F_{sd}^{(h)}$).

3- $P_{sd} = \text{Max}_{(1 \leq i \leq N,\ 1 \leq j \leq N)} (PR_{sd,ij})$

For a given traffic T_{sd}, P_{sd} is the minimum number of flow units required to protect the traffic against a single failure. The P_{sd} represents the highest risk; it is also the maximum number of flow units of the traffic, carried on the same physical link.

If PT_{sd} is the backup protection of the working flow F_{sd} then PT_{sd} should carry, at least, P_{sd} flow units in order to provide protection of a working connection with guaranteed recovery of similar grade of service.

b- The identification of the resources that could be used for the protection:

The protection on the primary traffic may share some links. However, one needs to make sure that in case of a link failure, there is enough bandwidth on the other protection flows (those that do not cross the failed link) to restore the traffic affected. The basic idea is that the link carrying more flow units of the primary traffic should carry less of the protection traffic and vice versa. If $PNS_{sd,ij}$ is the number of flow units that could be used by the protection (as stated in the definition) then it should respect the two following constraints:

- The total flow units available: indeed the capacity of a link is limited. The link could also be used by other flows belonging to other traffics. This constraint could be expressed as follows: For $(1 \leq i \leq N)$ and $(1 \leq j \leq N)$ $PNS_{sd,ij} \leq NS_{ij}$

- The shared risk constraint: in order to balance the protection over the links and avoid sending too much traffic on a single link (from the primary and backup traffic), one needs to control the backup flows. This could be expressed as follows.

For $(1 \leq i \leq N)$ and $(1 \leq j \leq N)$, $PNS_{sd,ij} \leq P_{sd} - PR_{sd,ij}$;

$PNS_{sd,ij}$ is the maximum number of flow units that could be used by the backup on link E_{ij}. That is For $(1 \leq i \leq N)$ and $(1 \leq j \leq N)$, $PNS_{sd,ij} = \min (NS_{ij}, P_{sd} - PR_{sd,ij})$.

c- Protection provisioning:

The bandwidth required for the protection as well as the resources available have been identified in Phases 1 and 2. The backup PT_{sd} could be considered as new traffic requiring P_{sd} flow units from the source s to the destination d. The resource allocation module is therefore engaged to reserve the required number of flow units. This task is performed transparently with no difference between the primary and the backup traffic. However, the new network resources are assumed to be those computed in the Phase 2 (the link availability is $PNS_{sd,ij}$).

If the protection provisioning succeeds in finding the resources required then a set of flows (carrying a certain number of flow units each) is provided and the protection capacity is returned. In the example of Figure 7.1, PL1 and PL2 are the result of the protection provisioning in that particular case.

If the protection provisioning does not succeed, P_{sd} is incremented by one flow unit and we return back to the step b, which is the identification of the resources.

Lemma 1:

If the shared protection scheme described above is deployed in a network then it is possible to recover from any single link failure.

Proof

Let us consider a failure in link Eij.

Let T_{sd} be the traffic going from s to d with some flow units riding E_{ij} (T_{sd} has exactly $PR_{sd,ij}$ flow units going on Eij). If T_{sd} is protected with PT_{sd} backup using a shared flow protection scheme then PT_{sd} is carrying P_{sd} flow units.

The protection shares the links with the primary traffic. It may have some flow units carried on E_{ij}. Let A be the number of flow units belonging to the protection flow and going on the link E_{ij}. We know that A + PR_{sdij} is less than or equal to P_{sd}. That means that PF_{sd} must have at least $PR_{sd,ij}$ (which is exactly the number of flow units of T_{sd} going on link E_{ij} that need to be restored) flow units crossing other links. Therefore in case of E_{ij} failure, the $PR_{sd,ij}$ flow units could be switched over the other links resrved for the protection (the protection available on other links could accommodate at least $PR_{sd,ij}$ flow units).

7.2.3. Single node failure protection

The basic idea of this protection technique is that the risk on every link is controlled; the link that provides more resources to the primary traffic will provide less resource to the protection flow and vice versa. This technique could be adapted to support a single node failure. In the single link failure technique, the focus was on the risk on links, whereas in a single node failure one should focus on the risk at every node crossed by the primary traffic. The following is an algorithm for single node failure protection:

Besides the definition used in the first algorithm we will need the following:

- $\mathbf{PR_{sd,i}}$ the protection capacity required for the Node Ni for the given traffic T_{sd}.
- $\mathbf{PA_{sd,i}}$ the protection capacity available on node Ni for the given Traffic T_{sd}.

Like the first algorithm, this protection scheme is carried out in three phases:

a- protection capacity:

This phase consists of the identification of the protection capacity required for the backup of $\mathbf{T_{sd}}$ from the source s to the destination d. the traffic $\mathbf{T_{sd}}$ is composed of k flows.

1- for $(1 \leq i \leq N)$, $PR_{sd,i} = 0$;

2- for all flows that compose the traffic T_{sd} ($F_{sd}^{(h)}$ with $1 \leq h \leq k$)

for $(1 \leq i \leq N)$, if Ni is crossed by $F_{sd}^{(h)}$ then $PR_{sd,i} = PR_{sd,i} + capacity(F_{sd}^{(h)})$.

3- $P_{sd} = Max_{(1 \leq i \leq N)} (PR_{sd,i})$

For a given traffic T_{sd}, $PR_{sd,i}$ is the number of flow units going through node Ni and P_{sd} is the number of flow units required to protect the traffic against a single node failure. The P_{sd} represents the highest risk in regards to nodes failure; it is also the maximum number of flow units of the traffic, crossing the same node.

If PT_{sd} is the backup of the working traffic T_{sd} then PT_{sd} should carry, at least, P_{sd} flow units in order to provide protection of a working connection with guaranteed recovery of a similar grade of service.

b- The identification of the resources that could be used by the protection:

The backup and the primary traffic may go through the same nodes. However, one needs to make sure that in case of a node failure, there is enough bandwidth allocated to the other protection flows (those that do not cross the failed node) to restore the affected traffic. The basic idea is that the node serving more flow units of the primary traffic

should serve less traffic of the protection and vice versa. If Pasd,i is the number of flow units of the protection that could go through node Ni then it should respect the following constraint:

For $(1 \leq i \leq N)$, $PA_{sd,i} = P_{sd} - PR_{sd,i}$

This new constraint combined together with the flow units available on different links defines the new resources that will be used to provision the protection flows. If $PNS_{sd,ij}$ is the number of flow units on link E_{ij} that could be used for the protection flows (as stated in the definition section) then $PNS_{sd,ij}$ should satisfy the two conditions:

- The total flow units available: For $(1 \leq i \leq N)$ and $(1 \leq j \leq N)$ $PNS_{sd,ij} \leq NS_{ij}$ (1)

- The shared risk constraint at node Ni: For $(1 \leq i \leq N)$ $\sum_{j=1}^{N} PNS_{sd,ij} \leq PA_{sd,i}$ (2)

c- Protection provisioning:

The backup PF_{sd} could be accommodated as a new traffic requiring P_{sd} flow units from the source s to the destination d. The task of the resource allocation module is to find and reserve P_{sd} flow units. Constraint (1) and (2) should be observed.

If the protection provisioning succeeds in finding the resources required then a set of flows (carrying a certain number of flow units each) is provided and the protection capacity is returned.

If the protection provisioning does not succeed, P_{sd} is incremented by one flow unit and we return back to the step b, which is the identification of the resources.

Lemma 2:

If the node-based shared protection scheme described above is deployed in a network then it is possible to recover from any single node failure.

Proof

Let us consider a failure of Node Ni.

Let T_{sd} be the traffic going from s to d with some flow units going through Ni (T_{sd} has exactly $PR_{sd,i}$ flow units crossing Ni). If T_{sd} is protected with PT_{sd} backup using a shared flow protection scheme then PT_{sd} is carrying P_{sd} flow units.

The protection paths share the node with the primary traffic. It may have some flow units crossing Ni. Let A the number of flow units belonging to the protection flows and crossing the node Ni. PF_{sd} must have at least (P_{sd} – A) flow units crossing other nodes. We know that A + $PR_{sd,i}$ is less or equal to P_{sd}. That means that PF_{sd} has more than $PR_{sd,i}$ crossing other nodes (which is exactly the number of flow units of T_{sd} crossing Ni and need to be restored). Therefore in case of Ni failure, the $PR_{sd,i}$ flow units could be switched over the protection on the other nodes.

7.2.4. Multiple failures protection

Link and node failures are the most common type of risks networks are facing. Nevertheless, there are some faults that can affect many network components at the same time (the case of earthquake if the components are located in the same area or blackout if the components are powered by the same source etc). The network can be protected against these kinds of failures if the potential faults are carefully investigated and their impact on the different network components is examined.

Let C be a set of components (both links and nodes): $C=\{ c_i \mid i=1,.,n\}$

Let F be a set faults and malfunctions: $F=\{f_j \mid j=1, .,m \}$.

Basically for every fault of F, a sub-set of the component in C is affected. We call this the impact mapping M. If f_j is a fault in F then $M(f_j)$ is a group of components in C affected at the same time by f_j.

If T_{sd} is the traffic from a source s to a destination d then T_{sd} crosses a sub-set of components of C. We call this the route mapping U. U ($F_{sd}^{(i)}$) is the sub-set of those components in C which carry the flow $F_{sd}^{(i)}$:

T_{sd} is affected by a fault fj if $U(T_{sd}) \cap M(f_j) \neq \emptyset$.

The scheme that guarantees a recovery from failure of a single group of components is called a component group-based shared protection. We assume that different groups of components are already identified. This operation is possible by using the impact mapping function for a given fault. One application of a protection against group failure is an inter-domain routing where the information is crossing many autonomous systems. Nevertheless the group is not limited by a geographical area; it could be defined as an aggregation of many resources and components even if they belong to different geographical areas. Figure 7.2 shows an example of a network with 4 groups.

In order to achieve a protection against a single group failure we should focus on a risk that a given traffic is facing when it is crossing every group.

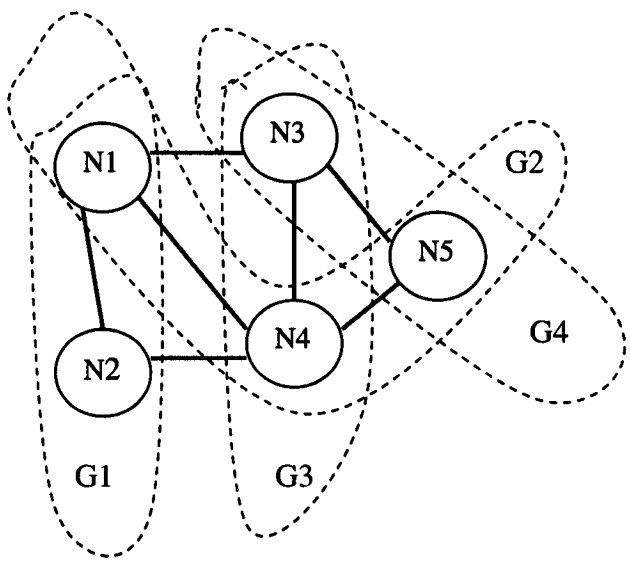


Figure 7.2: A network with 4 groups

In this context we use the following definitions:

- $\mathbf{PR_{sd,i}}$ the protection required for group Gi for the given traffic T_{sd}.
- $\mathbf{PA_{sd,i}}$ the protection available for group Gi fro the given Traffic T_{sd}.

Because of the similarity between the role of node and group of components, one can use the same algorithm to compute the bandwidth required by the protection. The same algorithm could also be used to identify the maximum number of flow units belonging to the protection and crossing every single group.

If the group-based shared protection is deployed, one can guarantee a recovery from any single group failure.

7.2.5. A recovery from a failure

Shared protection is resource efficient. It uses just enough resources for the backup protection. However, this difference (between the capacity of primary and protection traffic) makes the process of recovery in case of failure not as easy as the dedicated protection where the traffic is sent on both primary and backup. In the latter, if a failure occurs the traffic on the backup takes over and becomes active. Whereas in shared protection, in order to activate the protection (in case of a single failure), every node should identify all the affected flows of the working traffic as well as the affected flows of the backup. (The protection may share some links with the primary traffic).

Let F1,F2,..,Fk be the flows affected by the failure and PF1,PF2,..,PFh be the flows of the protection that are not affected by the failure. We know that if a shared protection scheme is used then $\sum_{i=1}^{k} capacity(F_i) \leq \sum_{i=1}^{h} capacity(PF_i)$. And hence, one can fill up the backup flows using flow units of the failed flows.

The signaling required in the shared protection scheme is also more complicated than the one required in dedicated protection. And hence the recovery speed maybe longer than that of dedicated recovery, but definitely faster than the restoration scheme since the recovery flows are already identified and resources are reserved.

The protection algorithm is not coupled with the reservation and routing algorithms. However, it is possible to increase the efficiency by involving more diversity in the primary flow. Indeed, one needs to compose the flow traffic in very small flows and share the minimum links possible.

7.3. Cases studies

7.3.1. Optical network deploying time division multiplexing

The motivation behind deploying time division multiplexing in WDM technology is to achieve high network utilization by partitioning the bandwidth of a wavelength into multiple time-slots. Indeed, in wavelength routing when a source has traffic to send to a destination, a light-path is established (and used during the whole session) between the source and the destination. However, the source may not have enough traffic to fill the whole bandwidth available in a light-path.

Each wavelength is partitioned in the time-domain into fixed-length frames; and each frame consists of a number of fixed-length time-slots with a fixed duration. Every time slot is switched from source to destination following its assigned switched path which consists of a small channel. Therefore multiple sessions could be multiplexed on a single wavelength channel using TDM frames.

In this architecture, the bandwidth required by a session t is specified in time-slots. The network assigns a set of switched paths to carry the required time slots. The routing consists of time slot assignment, which makes sure that the number of time slots over a given wavelength on every link is less or equal to the number of slots per frame. The resources are allocated and reserved during the session.

The constraint of bandwidth granularity, required by the wavelength routed architecture, is relaxed. Therefore the bandwidth is used more efficiently, since the network reserves exactly the number of slots needed. Therefore there is no need for assigning the whole bandwidth of a wavelength. Furthermore, blocking is reduced. Indeed, while the wavelength routed network may fail in establishing a light-path from a source to a destination, the slotted wavelength routing architecture could still find a way to the destination, by aggregating many small available channels.

In order to carry a maximum number of time slots, a session may use many flows following different physical paths. These flows may share some links or some nodes across the network.

7.3.2. Star network

We consider an optical network where the nodes are connected in a star topology, where all the nodes exchange information through a central node. In this architecture, each node can be connected to more than one central node. Therefore, a source node can reach another node through more than one physical path.

To efficiently use the network resources and increase the bandwidth, time division multiplexing is used. The nodes use time slots to send information.

The overlaid star architecture is another extension of the star network where many star networks are connected together, allowing more than one path between two nodes.

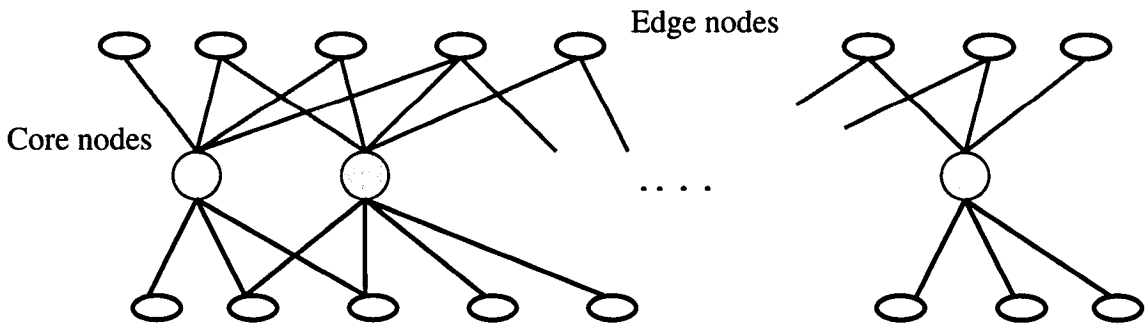


Figure 7.3 : An overlay star network

In this architecture, shown in Figure 7.3, a node could reach another node using different core nodes. Therefore traffic form a source to a destination could be composed of many different flows going trough different stars and each carrying a different number of time slots.

The figure 7.4 shows this situation. Node S and D are connected to n core nodes and can establish n flows. A link Li in Figure 7.4 is composed of two physical links and one core node. Consequently any failure of one of these components will lead to a failure of Li.

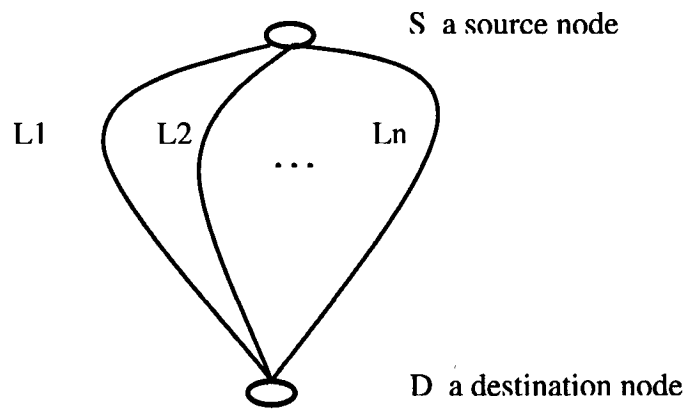


Figure 7.4: A flow of time slots between two nodes

Li is carrying a number of time slots. Let Ci be this number Ci ranges from 0 up to the maximum capacity of the Link. The capacity of the primary traffic going from S to D is C. $C = \sum_{i=1}^{n} Ci$.

To protect this traffic one needs to find an optimal protection capacity Pc and n different flows going from S to D. Let Pci the number of slots going on link Li and belonging to the protection as shown in Figure 7.5.

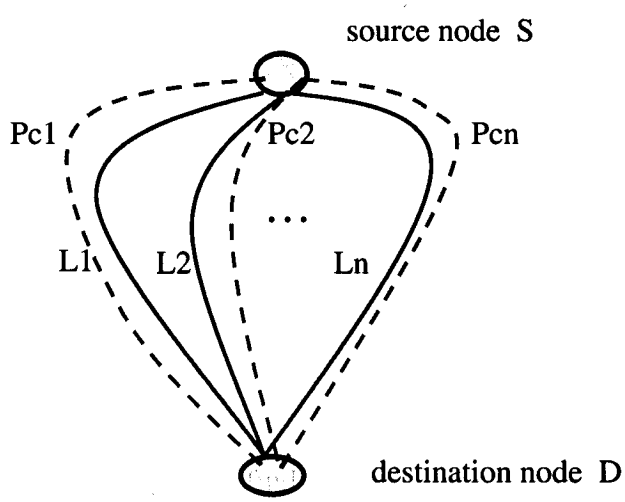


Figure 7.5: A flow of time slots between two nodes with a protection

The shared protection aims to minimize the capacity Pc and find a set of flows, from S to D, to carry Pc time slots. The protections flows should satisfy the following constraints:

- $Pc = \sum_{i=1}^{n} Pci$. (1)
- Max $(1 \leq i \leq)$ Ci $\leq$ Pc (2)
- Pci $\leq$ Pc – Ci (3)

- In case there is a link Lj not used by the primary traffic (Cj=0): we assume that there are available resources on Link Lj. The solution could be the following:

$Pc = Max\ (1 \leq i \leq)\ Ci$.

$Pci = 0$ for all $i \neq j$

$Pcj = Pc$.

This is a1:N protection where all the flows are sharing the same protection.

- The primary traffic is sending the same number of time slots on the different link: in this case for all the link Ci = b. we assume that there are available resources on all the links. The protection could send the same number of time slots on the different link. Let P this number:

$Pc = n * P$ from the equation (1)

$b \leq Pc$ from the equation (2)

$P \leq Pc - b$ from the equation (3)

(3) is satisfied if $P = Pc - b$ which means according to (1) that $p = b / (n-1)$. With this value of p the constraint (2) is also satisfied.

The proposed algorithm is a practical way to find a solution for these equations. The idea is to increase the capacity of the protection progressively until the optimal value is found. The first iteration of the algorithm determines the minimum capacity of the protection required. Sometimes it is not possible to protect the whole primary traffic because of the shortage of resources. In this case the algorithm finds the capacity that protects the maximum traffic.

7.4.Simulation result and analysis

We studied the performance of the proposed shared flow protection scheme by means of simulations, considering the NSFNET topology with 14 nodes as shown in Figure 7.6. We assumed that each single fiber link is bi-directional, and has the same number of wavelengths (15 in this case) operating each at 50 Gbps. The distance of each fiber link is shown in the network graph of Figure 7.6. Each wavelength is divided into 50 small timeslots (circuits) of 1 Gbps each. In the network, a node can route, generate, and receive traffic. A source node is responsible for segmenting the data into flow units (timeslot) for transmission, and re-assembling the slots upon reception. The traffic is uniformly distributed across the network. Traffic is generated as a number of connection requests. The number of connections is varied (increased uniformly) to study the impact of the load on our scheme. The source and the destination are also chosen randomly and uniformly among the network nodes. The bandwidth required by the connection is uniformly distributed between 1 and 20 flow units. In the simulation, more than one flow may be used to carry the flow units, we used the k-Dijkstra algorithm to identify the k shortest paths (we use 4 paths in this simulation) between a source and a destination, since one path might not provide all the required bandwidth. We do not employ conventional buffers or wavelength converters in the network switches.

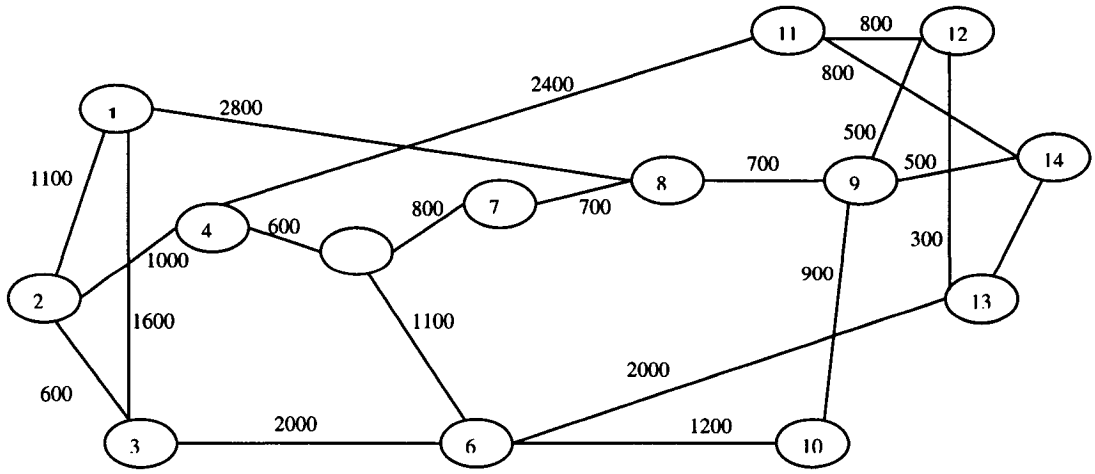


Figure 7.6: NSFNET topology with 14 nodes

The goal of the simulation experiment is to study the performance of our proposed protection scheme used in time slotted optical network (TSON) as compared to the same protection in wavelength routed optical network (WRON).

The performance is measured based on various metrics. We are first interested in investigating the protection efficiency, which is the ratio of the required back-up capacity over the capacity of the primary traffic; this reflects how much capacity is required to protect the primary flow. For a dedicated protection this metric is 1. However, shared protection aims to reduce this value to achieve bandwidth savings.

In order to analyze the impact of the routing strategy on the performance of shared protection, we use two schemes to distribute the flows units over the k shortest flows; in the first one, which we call shortest-paths-first scheme (SPFS), we start by filling up the shortest flows first. In the second one, which we call diversity-first scheme (DFS), we distribute the flow units over all the k shortest paths in order to have the maximum diversity possible.

The chart (see Figure 7.7) shows the protection efficiency versus the traffic load for SPFS routing strategy. In this simulation the number of requests is varied (we increase the number of requests) for both shared flow protection and wavelength routed network (WRO). As the traffic load increases, the charts show that shared protection uses less resource than WRO. Nevertheless when the traffic is light the backup traffic uses almost the same amount of bandwidth. The request is accommodated in only one flow which needs to be protected by another similar flow. As the traffic gets higher the requests carry more flow units and need more than one flow creating more opportunities for shared protection to save bandwidth. For wavelength routed protection the requests use the whole wavelength to carry the traffic and therefore the whole wavelength should be protected. Thus the protection requires almost 100% of the primary traffic. When the traffic gets higher (more than 10000 Gps) there is more chance to fill more than one wavelength (which could share a backup in this case) and consequently some bandwidth saving can be achieved.

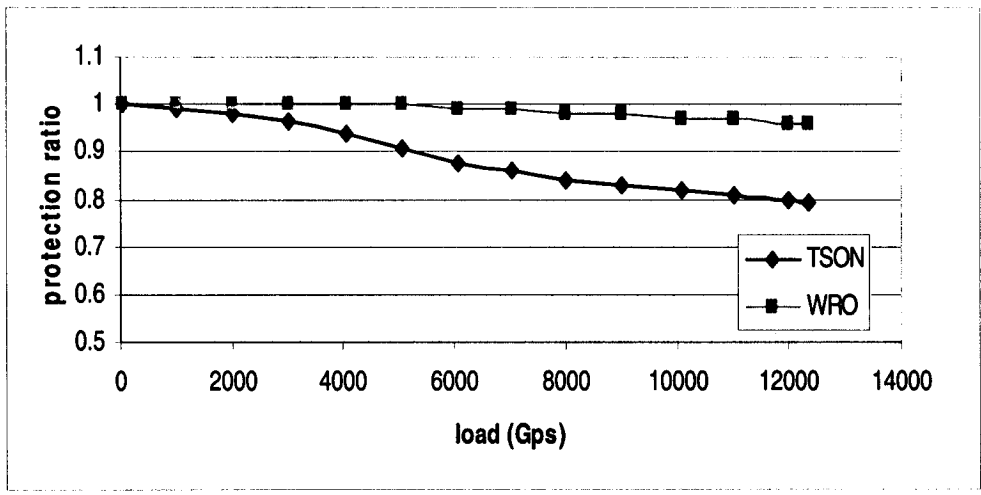


Figure 7.7: Protection efficiency for SPFS

Figure 7.8 illustrates the impact of a routing strategy on the protection efficiency. Indeed when the DFS routing strategy is used, the traffic of a request is sent over many different flows. This gives more chance to our proposed scheme to optimize the shared backup. When the traffic volume is light the efficiency is very high. However, when the traffic increases the number of paths between a source and a destination becomes limited and the diversity decreases. When the traffic volume is large we get almost the same result as with SPFS strategy. Indeed, all the flows are filled and no choice is left (regardless of the routing strategy). As for the wavelength routed network, the whole traffic is sent over one wavelength when the traffic is not very high. When the traffic is high and the source has enough load to fill more than one wavelength the possibilities, from a source to a destination, become more limited and only a small bandwidth saving is achieved.

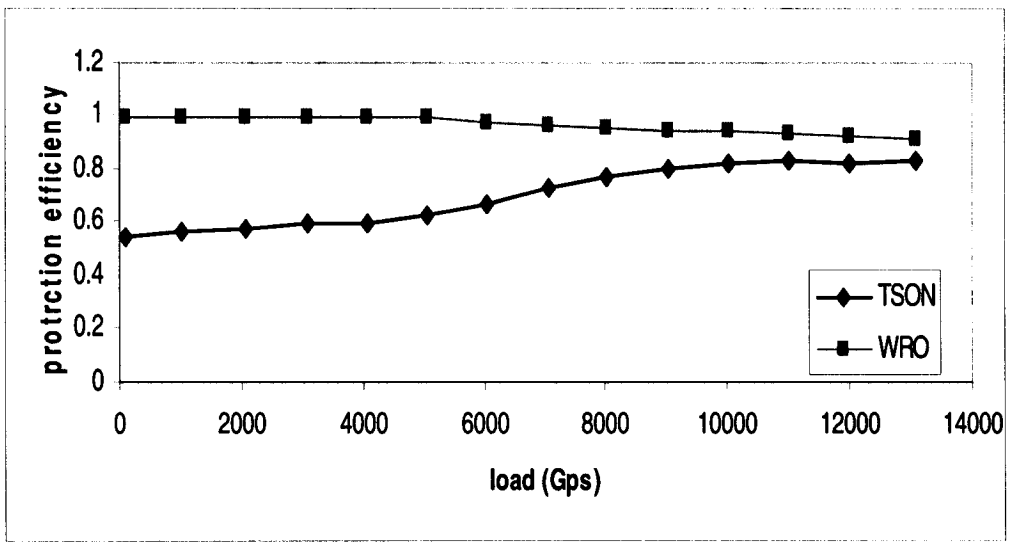


Figure 7.8: Protection efficiency for DFS

In the previous simulation, we investigated only the bandwidth needed for the backup without any reservation. However, if some resources are effectively used for the backup then both future primary traffic and their protection may suffer shortage in resources leading to some blocking for some requests or their backups. One of the metrics we investigate in the second simulation is the blocking ratio, which reflects the percentage of traffic that must be discarded due to shortage in resources

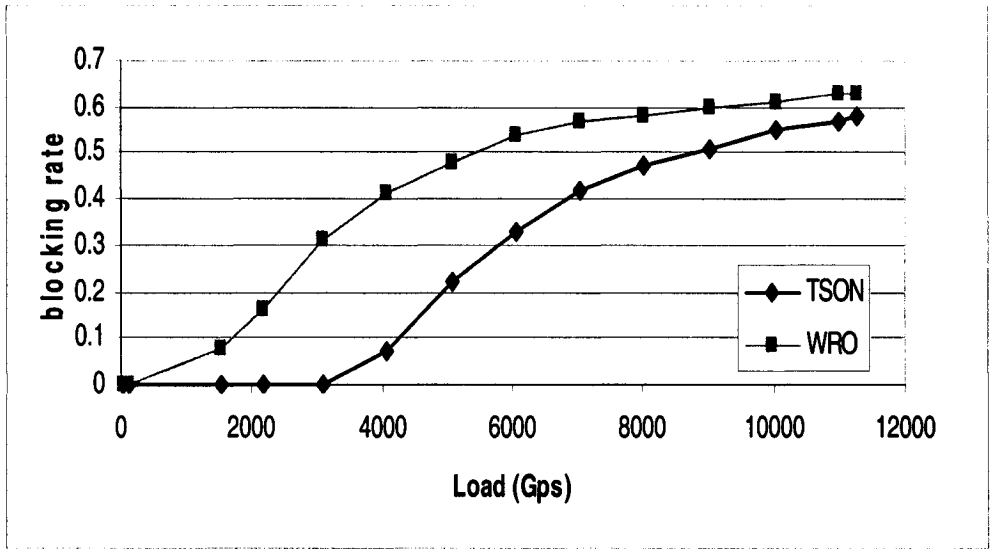


Figure 7.9: Blocking rate for DFS

The charts (see Figure 7.9) show the blocking ratio versus the traffic load for both a wavelength routed network and optical network deploying division multiplexing. As the traffic load increases, the charts show that TSON accommodates more traffic than the WRO technique. In addition, TSON maintains a zero blocking ratio; while WRO blocked more than 10% of the traffic. When the load is highter, the blocking rate is very high (more than 60%); TSON is performing slightly better than wavelength routing technique. This is resulting from the excessive use of resources by the primary and backup.

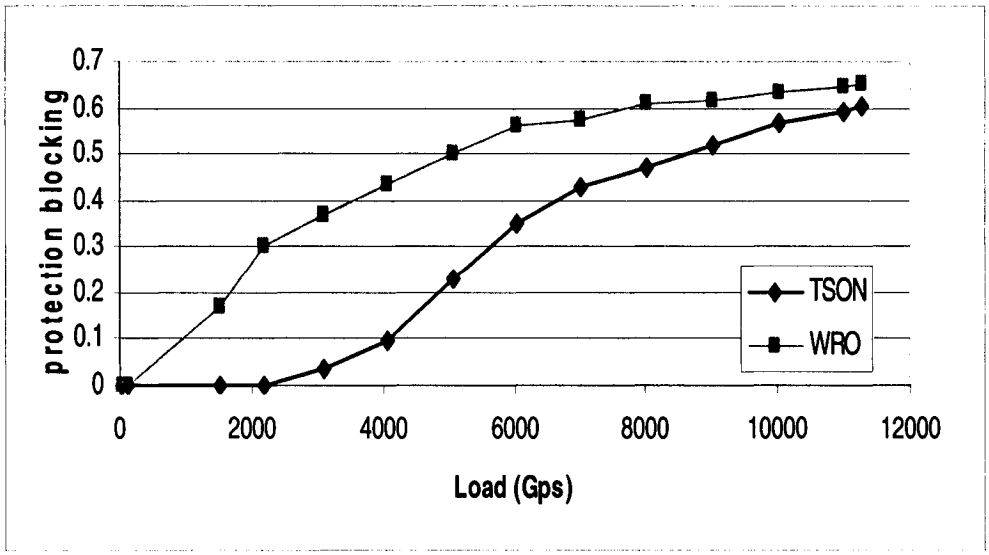


Figure 7.10: Protection blocking rate for DFS

Besides the primary traffic, resources also should be allocated and reserved for the backup. However, it is not possible to find enough resources for all the protection and some blocking maybe observed especially under heavy load. The chart in Figure 7.10 shows the blocking rate for the protection. This reflects the percentage of protection traffic that must be discarded due to shortage of resources. The trend of the curves is very similar to the blocking rate for the primary traffic. Indeed for the resource allocation module the primary and backup are considered equally. And hence the backup suffers the same blocking ratio as the primary traffic. Nevertheless the protection blocking is a slightly higher than that of the primary traffic because the protection is observing more constraints than the primary traffic.

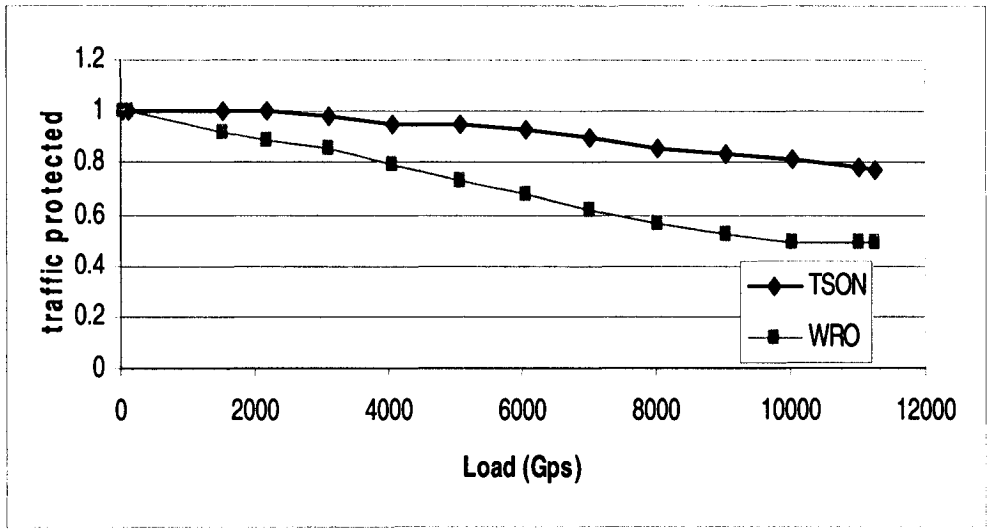


Figure 7.11: Ratio of the traffic protected for DFS

When the protection algorithm fails to find enough resources for a backup, a part of the primary traffic is left without protection. Figure 7.11 shows the fraction of traffic protected. For light traffic the TSON protect 100% of the traffic whereas WRO drops some protections early. This is because with TSON the blocking rate is very low (it is almost zero at light loads). Besides that the protection requires only a small capacity for the backup. The WRO requires 100% of the primary traffic. Therefore both primary and backup traffic will suffer some blocking according to Figure 7.9 and 7.10. When the traffic is heavier, both WRO and TSON fail to protect the whole primary connections. However, as the traffic volume gets higher the gap between the two techniques becomes larger. This is due to the granularity; when WRO fails in provisioning a light-path for a backup the whole traffic is left without protection whereas in the case of TSON if the whole protection is not possible, a fraction of the bandwidth required by the backup (a number of flow units) could be allocated. Consequently, a part of the primary traffic can be protected.

7.5. Summary

In this work, we proposed a new shared protection scheme that aims to use the available bandwidth more efficiently while providing the same level of protection as a dedicated protection scheme. For a given traffic from a source to a destination, the different flows are analyzed and the optimal shared protection is identified. This shared protection could be performed to protect a network against a single link or a single node failure or even against a single group of components failure if a network has been divided into regions.

In a time slotted optical network, one may need to establish more than one flow between a source and a destination. Therefore a failure of a single node or a single link may affect only a part of the traffic going through these different flows. Our proposed shared protection analyses the primary traffic in order to identify the maximum bandwidth required by the protection.

The simulation proves that this shared protection scheme combined with slotted optical network is resource efficient and provides the same level of protection as dedicated protection while saving more bandwidth. The proposed scheme is also better than dedicated protection in the context of a wavelength routed network, especially under heavy load.

Further work is needed to deal with the multiple failures where more than one link or more than one node is down. Another issue is related to the protection provisioning; indeed we assume that once the protection bandwidth is computed, the backup is considered as a new request.

Chapter 8

Conclusion

8.1.Summary and conclusive remarks

Advances in optical networks are driven by both developments made in wavelength division multiplexing and new applications requiring more bandwidth such as multi-media services.

Optical networks can be used as a backbone to carry information between the different edge nodes in an efficient way. The robustness of such a backbone could be affected by contention and components failure:

- Contention happens in the context of optical burst switching. Whenever two or more bursts compete for the same output at the same time, only the first one will be sent and the others are dropped.
- The component failure or malfunction affects the reliability of a network and can lead to some loss of data.

In this work, the main goal is to enhance the network robustness and provide techniques that help to build intelligent networks able to carry data at high rate (between different nodes) with no loss. The network should be able to face any kind of failure and assure its survivability against the major problems.

In the first step and in order to reduce the contention, we proposed another variant of OBS, which use a segmented burst instead of the burst. With the segmented burst, the

ingress routers will group incoming packets into a set of segments with constant length. In case of contention, instead of dropping the whole burst, only the contending segments, belonging to the second burst will be discarded whereas the rest of segments can continue their way. Furthermore, this technique could be used to carry different classes of traffic. Indeed, the segments at the burst tail have more chance to survive contention than those at the head. This technique has the merit to save some data that would have been discarded by the regular OBS.

In order to aggregate and accommodate different classes of traffic and avoid loss of data, we propose to combine OBS and wavelength assignment in the same network. The basic idea of this architecture is to take advantage of the large number of wavelengths in a fiber by using only a part of the available wavelengths for wavelength routed network and the other part for OBS. Depending on the traffic pattern the wavelength routed sub network will be configured to balance the network load while the other part remains free to be used with the OBS technique.

Another issue is related to the OBS mechanism. Indeed, the edge nodes keep sending bursts regardless of the network load and without any global coordination which may overwhelm the network leading to a situation where the contention is very high and the loss is unacceptable. Furthermore, in case of contention an intermediate node simply drops a burst and ignores it. The higher layers are therefore in charge of detecting and recovering the loss. This may increase the burden of higher layers and increase the recovery time, and hence resource wasting (since the recovery is performed farther from the source). In order to avoid this problem, we think that intermediate and edge nodes must be engaged in a global process to keep the loss in an acceptable level and recover

from any eventual loss. Such a process aims to enhance the performance of optical burst switching and eliminate burst loss completely. We propose to reduce contention by controlling the load and avoiding congestion. Basically, the intermediate nodes provide the edge nodes with statistic information on burst loss rate. Which in turn use this information to adjust their traffic and balance the load over the different wavelengths. This technique is augmented by a retransmission scheme where intermediate nodes will notify and report the loss. This way the edge node can retransmit the dropped burst and hence increase the network robustness and reliability.

The basic idea of congestion control is that the network resources are limited. Wise utilization and rational sharing is therefore a necessity. To achieve this goal, we proposed a new bandwidth allocation scheme and switch architecture to share network resources, avoid contention, reduce blocking ratio and improve bandwidth utilization. In this scheme, the wavelength bandwidth is broken up into many time slots which are used to carry traffic from sources to destinations. The improvement in bandwidth utilization and reduction in blocking ratio were achieved by employing Sequencers, a form of Optical Time-Slot Interchangers, to delay an incoming timeslot for an adequate period of time in order to match a free outgoing timeslot. The blocking ratio is improved further by a simple reservation scheme that uses multiple paths to transmit traffic. Every path corresponds to an LSP group consisting of many LSPs; each LSP is labeled by its timeslot position. The contention is avoided by employing a basic scheduling algorithm to derive the amount of delay needed at intermediate sequencers, after constructing the mapping tables.

In the second step, and in order to enhance the network reliability, we proposed a new shared protection scheme that aims to use the bandwidth more efficiently while providing the same level of protection as a dedicated protection.

The capacity of the protection as well as the resources available on each component of the network is determined by our proposed algorithm. This shared protection scheme could be used to protect a network against a single link failure, a single node failure or even against a single group of components failure (if a network has been carefully divided into groups). Like dedicated protection, in case of failure the shared protection is able to restore the whole traffic and achieve , at the same time, better bandwidth saving.

8.2.Future research

The proposed techniques could be used to improve and increase the robustness of the optical network. Further elaboration of each technique can provide more benefits and can offer more alternatives for a design of the next generation of optical backbones.

Segmented burst switching has attracted more interest in the literature. It could be extended to a slotted segmented OBS where each segment can fit exactly one time slot. The size of one segment and the dropping policy (which segments should be dropped in case of contention) are some of the parameters that should be investigated. Data bursts can also carry segment of different sizes to aggregate different type of data in the same burst (IP packet, ATM cells, etc). Thus the header must carry enough information to identify the different segments as well as the dropped elements. One needs to analyze the size of such information in the header and its impact on the processing time and queuing model.

In the hybrid architecture, the combination of OBS and wavelength routed technique brings some new challenges especially those related to the portion of the wavelength channel that should be used by OBS and wavelength routed. The repartition of wavelengths could be static (if wavelengths are assigned to the two techniques) or dynamic (if the assigned wavelengths vary with the traffic pattern). However, no matter what the repartition strategy is, we think that network topology and class of traffic should be taken into account.

For the congestion avoidance that aims to control the traffic at ingress nodes. It acts as a traffic shaping and admission control in the photonic domain. Nevertheless this scheme could be extended to control the congestion at every node. This will require collecting statistics from different nodes. This kind of measurements will allow the source nodes to adjust traffic flows separately taking into account the load of crossed nodes. It will also allow source nodes to redistribute their traffic over other paths, and hence converge to a global load balancing.

The retransmission scheme relies on buffers at the edge nodes that can hold the sent burst until it reaches its destination (no negative acknowledgement received during a period of time). The size of such buffer depends on the network size and links capacity. The retransmission may incur additional delay to a burst (if one or more transmissions are needed). However, this delay may not be acceptable to some classes of traffic.

For the time slotted technique, one needs to optimize a switch (that performs the time slot switching in the optical domain). This optimization goes hand in hand with the resource allocation and slot assignment in order to avoid blocking at a switch and increase the network utilization.

To increase network survivability, we proposed an algorithm that identifies the required resources to achieve protection efficiently. This scheme assures the protection against a single failure (link, node or a group). However, the risk to have more than one failure is still there and one needs to extend the proposed algorithm to handle such situations. The protection provisioning is completely decoupled from the primary traffic routing strategy. However, using some specific strategies can create the opportunity to achieve more bandwidth saving.

References

[1] Papadimitriou, G.I.; Papazoglou, C.; Pomportsis, A.S "Optical switching: switch fabrics, techniques, and architectures".; Journal of Lightwave Technology, , Volume: 21 , Issue: 2 , Feb. 2003 Pages:384 - 405

[2] P. Ferreira, B. Cossa "Testing the Scalability of DWDM networks," 4th International Conference on Technology Policy and Innovation, Brazil, Aug. 2000

[3] M. Listanti, V.Eramo, R. Sabella, "Architectural and Technological Issues for Future Optical Internet Networks", IEEE Communications Volume: 38, Issue: 9, Sept. 2000 Pages: 82 – 92

[4] Gipser, T.; Ming-Seng Kao, "An all-optical network architecture," Journal of Lightwave Technology, Volume: 14, Issue: 5, May. 1996 Pages: 693 – 702

[5] Fei Xue; Ben Yoo, S.J.;"High-capacity multiservice optical label switching for the next-generation Internet" IEEE Communications Magazine, , Volume: 42 , Issue: 5 , May 2004 Pages:S16 - S22

[6] J. Turner, "Terabit Burst Switching, " Journal of High Speed Networks, Volume: 8, Issue: 1, 1999, Pages: 3-16.

[7] Sanjeev Verma, Hemant Chaskar, and Rayadurgam Ravikanth, "Optical burst switching: a viable solution for terabit IP backbone", IEEE Network, Volume: 14, Issue: 6, Nov.-Dec. 2000 Pages: 48 – 53

[8] Yang Chen; Chunming Qiao; Xiang Yu "Optical burst switching: a new area in optical networking research"; Network, IEEE , Volume: 18 , Issue: 3 , May-June 2004 Pages:16 – 23

[9] Battestilli, T.; Perros, H, "An introduction to optical burst switching," IEEE Communications Magazine, , Volume: 41, Issue: 8, Aug. 2003, Pages: S10 - S15

[10] Wanjiun Liao; Chi-Hong Loi] "Providing service differentiation for optical-burst-switched networks"; Journal of Lightwave Technology, , Volume: 22 , Issue: 7 , July 2004 Pages:1651 - 1660

[11] A.Grosso, E.Leonardi, M.Mellia, A.Nucci, "Logical Topologies Design over WDM Wavelength Routed Networks Robust to Traffic Uncertainties", IEEE Communications Letters, Volume: 5, Issue: 4, Apr. 2001, Pages: 172–174,

[12] Hartline, J.R.K.; Libeskind-Hadas, R.; Dresner, K.M.; Drucker, E.W.; Ray, K.J "Optimal virtual topologies for one-to-many communication in WDM paths and rings".,

IEEE/ACM Transactions on , Networking Volume: 12 , Issue: 2 , April 2004 Pages:375 – 383

[13] M.Sridharan, R.Srinivasan, A.K.Somani "Dynamic Routing with Partial Information in Mesh-Restorable Optical Networks " 6 IFIP Optical Network Design and Modelling Italy feb 2002, Pages: 327-343

[14] R. Krishnan, James P.G. Sterbenz, and L. Chapin, "Routing Issues in interconnecting IP networks with the PetaWeb", Networks 2000 Conference, Toronto, Ontario, Canada, Sep 2000

[15] R.Dutta and G.N. Rouskas, A Survey of virtual topology design algorithms for wavelength routed optical networks", Optical Network Magazine, Volume: 1, Issue: 1, Jan 2000, Pages: 73-89.

[16] X. Wang, H. Morikawa, and T. Aoyama, "A Deflection Routing Protocol for Optical Bursts in WDM Networks", Proc. Fifth Opto-electronics and Communications Conference (OECC 2000), Japan, Jul. 2000, Pages: 94-95,

[17] X. Wang, H. Morikawa, and T. Aoyama, "Burst optical deflection routing protocol for wavelength routing WDM networks", proceeding. SPIE/IEEE OPTICOMM 2000, Dallas, USA, Oct. 2000, Pages: 257-266

[18] SuKyoung Lee; Sriram, K.; HyunSook Kim; JooSeok Song "Contention-based limited deflection routing in OBS networks"; Global Telecommunications Conference, 2003. GLOBECOM '03. IEEE , Volume: 5 , 1-5 Dec. 2003 Pages:2633 - 2637

[19] Ching-Fang Hsu, Te-Lung Liu, and Nen-Fu Huang, "Performance Analysis of Deflection Routing in Optical Burst-Switched Networks", INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies., Volume: 1 , June 2002, Pages: 66 – 73

[20] Maach and G.V. Bochmann, "Segmented Burst Switching: Enhancement of Optical Burst Switching to Decrease Loss Rate and Support Quality of Service", Proceedings of Optical Network Design and Modeling (ONDM 2002), Italy, Feb. 2002, Pages: 69-84.

[21] Vinod Vokkarane, Jason Jue, Sriranjani Sitaraman, "Burst Segmentation: an Approach for Reducing Packet Loss in Optical Burst Switched Networks", Proceedings IEEE, International Conference on Conference (ICC) 2002, New York, NY, April-May 2002. Volume: 5, 28 April-2 May 2002, Pages :2673 – 2677.

[22] Zukerman, M.; Wong, E.W.M.; Rosberg, Z.; Gyu Myoung Lee; Hai Le Vu "On teletraffic applications to OBS"; IEEE Communications Letters, , Volume: 8 , Issue: 2 , Feb. 2004 Pages:116 - 118

[23] Rosberg, Z.; Vu, H.L.; Zukerman, M "Burst segmentation benefit in optical switching".;IEEE Communications Letters, , Volume: 7 , Issue: 3 , March 2003 Pages:127 - 129

[24] C. Gauger, "Dimensioning of FDL Buffers for Optical Burst Switching Nodes", Proceedings, Optical Network Design and Modeling (ONDM 2002), Italy, Feb. 2002, Pages: 117-132

[25] I. Chlamtac, A. Fumagalli, and C.J.Suh, "Multi-Buffer Delay Line Architectures for Efficient Contention Resolution in Optical Switching Nodes", IEEE Transactions on Communications, Volume: 48 , Issue: 12 , Dec. 2000. Pages: 2089 - 2098

[26] Huhnkuk Lim; Chang-Soo Park "An optical packet switch with hybrid buffer structure for contention resolution of asynchronous variable length packets"; Workshop on High Performance Switching and Routing, 2004. HPSR. 2004, 2004 Pages:162 – 166

[27] C. Gauger, "Performance of converter pools for contention resolution in optical burst switching", Proceedings, Optical Networking and Communication Conference (OptiComm) 2002, Boston, MA, July-Aug 2002. Pages: 109-117

[28] Batagelj, B, "The need for wavelength converters in all-optical transport network," Electrotechnical Conference, 2000. MELECON 2000, Volume: 1, May 2000, Pages: 277 - 280

[29] John Strand, Robert Doverspike, Guangzhi Li, "Importance Of Wavelength Conversion In An Optical Network," Optical Networks Magazine, Volume: 2, Issue: 3, May/June 2001, Pages: 33-44.

[30] Gaoxi Xiao; Kejie Lu; Chlamtac, I "An evaluation of distributed wavelength provisioning in WDM optical networks with sparse wavelength conversion".;Lightwave Technology, Journal of , Volume: 22 , Issue: 7 , July 2004 Pages:1668 - 1678

[31] Manchester, J.; Bonenfant, P.; Newton, C, "The evolution of transport network survivability," IEEE Communications Magazine, Volume: 37, Issue: 8, Aug. 1999, Pages: 44 – 51

[32] Kartalopoulos, S.V, "Surviving a disaster [optical communications]," IEEE Communications Magazine, Volume: 40, Issue: 7, July 2002, Pages: 124 – 126

[33] Chalasani, S.; Rajaravivarma, V.;" Survivability in optical networks". Proceedings of the 35th Southeastern Symposium on System Theory, 2003, 16-18 March 2003, Pages: 6 – 10

[34] Guizani, M.; Memon, A "Survivability performance evaluation of an optical switch," IEEE Global Telecommunications Conference, GLOBECOM '00, Volume: 2, 27 Nov.-1 Dec. 2000, Pages: 1192 – 1195.

[35] Suwala, G.; Swallow, G.; "SONET/SDH-like resilience for IP networks: a survey of traffic protection mechanisms" Network, IEEE , Volume: 18 , Issue: 2 , Mar-Apr 2004 Pages:20 - 25

[36] Nir, J.; Elhanany, I.; Sadot, D."Tbit/s switching scheme for ATM/WDM networks" Electronics Letters, Volume: 35, Issue: 1, Jan. 1999, Pages :30 – 31

[37] Pao, D "On-demand packet discard scheme for TCP over ATM-UBR service.;" IEE Proceedings- Communications, , Volume: 151 , Issue: 3 , June 2004 Pages:190 - 196

[38] Cavendish, D. "Evolution of optical transport technologies: from SONET/SDH to WDM"
IEEE Communications Magazine, Volume: 38, Issue: 6, June 2000 Pages: 164 – 172

[39] Berry, R.; Modiano, E "Grooming dynamic traffic in unidirectional SONET ring networks," Optical Fiber Communication Conference, 1999, and the International Conference on Integrated Optics and Optical Fiber Communication. OFC/IOOC '99. Technical Digest ,Volume: 1 , Feb. 1999 Pages:71 - 73

[40] Bernstein, G.; Mannie, E.; Sharma, V,."Framework for MPLS-based control of optical SDH/SONET networks," IEEE Network, Volume: 15, Issue: 4, July-Aug. 2001, Pages:20 – 26

[41] Bocci, M.; Guillet, J, "ATM in MPLS-based converged core data networks," IEEE Communications Magazine, Volume: 41, Issue: 1, Jan. 2003, Pages: 139 – 145

[42] Rojanarowan, J.; Koehler, B.G Owen, H. "MPLS Based Best Effort Traffic Engineering".; IEEE Southeastcon 2004, 26-29 Mar 2004 Pages:239 - 245

[43] Jae-Hyun Park; "Validation of the detailed design of the label distribution protocol for the multiprotocol label switching system," IEEE, Global Telecommunications Conference, 2001. GLOBECOM '01. Volume: 1, Nov. 2001, Pages: 17 – 24

[44] Dong Zhou; Ten-Hwang Lai "Efficient resource allocation in self-healing multiprotocol label switching mesh networks"; IEEE, Global Telecommunications Conference, 2001. Volume 4, 25-29 Nov. 2001 Pages:2671 - 2675

[45] Banerjee, A.; Drake, L.; Lang, L.; Turner, B.; Awduche, D.; Berger, L.; Kompella, K.; Rekhter, Y.; "Generalized multiprotocol label switching: an overview of signaling enhancements and recovery techniques," IEEE Communications Magazine, Volume: 39, Issue: 7, July 2001, Pages:144 – 151

[46] Jong-Moon Chung; Khan, H.K.; Hooi Miin Soo; Reyes, J.S.; Cho, G.Y.; "Analysis of GMPLS architectures, topologies and algorithms" The 2002 45th Midwest Symposium

on Circuits and Systems, 2002. MWSCAS-2002. Volume: 3, 4-7 Aug. 2002, Pages: III-284 - III-287

[47] Ding Zhemin; Hamdi, M.; Lee, J.Y.B "Integrated routing and grooming in GMPLS-based optical networks ". IEEE International Conference on Communications, 2004, Volume: 3, 20-24 June 2004, Pages: 1584 - 1588

[48] Myungsik Yoo; Chunming Qiao;"Just-Enough-Time (JET): a high speed protocol for bursty traffic in optical networks," Digest of the IEEE/LEOS Summer Topical Meetings, Aug. 1997, Pages: 26 – 27

[49] Byung-Chul Kim; You-Ze Cho; Jong-Hyup Lee; Young-Soo Choi; Montgomery, D.; "Performance of optical burst switching techniques in multi-hop networks," IEEE Global Telecommunications Conference, 2002. GLOBECOM '02, Volume: 3, Nov. 2002, Pages: 2772 – 2776

[50] Rosberg, Z.; Ha Le Vu; Zukerman, M.; White, J "Performance analyses of optical burst-switching networks," IEEE Journal on Selected Areas in Communications, Volume: 21, Issue: 7, Sept. 2003, Pages: 1187 – 1197

[51] M. Yoo and C. Qiao, "A New OBS Protocol for Supporting QoS", in SPIE Proc. of Conf. All-optical Networking, Volume: 3531, Nov. 1998, Pages: 396-405

[52] W.H. So and Y.C. Kim, "Offset Time Decision for Supporting Service Differentiation in Optical Burst Switching Networks", Proceedings of COIN-PS 2002, Cheju Island, Korea, July, 2002.

[53] Hai Le Vu; Zukerman, M.; "Blocking probability for priority classes in optical burst switching networks," IEEE Communications Letters, Volume: 6, Issue: 5, May 2002, Pages: 214 – 216

[54] Barakat, N.; Sargent, E.H "An accurate model for evaluating blocking probabilities in multi-class OBS systems" .; IEEE Communications Letters, , Volume: 8 , Issue: 2 , Feb. 2004 Pages:119 – 121

[55] Sungchang Kim; An Seek Choi; Minho Kang "Performance analysis for prioritized multi-classes in optical burst switching networks" The 6th International Conference on Advanced Communication Technology, 2004., Volume: 1, Feb. 9-11, 2004 Pages:72 – 74

[56] E.Kozlovski, P.Bayvel "QoS Performance of WR-OBS Network Architecture with Request Scheduling " Optical networks design and modeling, Turin, Italy feb, 2002, Pages: 101-116

[57] M. Dueser and P. Bayvel "Analysis of a Dynamically Wavelength-Routed Optical

Burst Switched Network Architecture" Journal of Lightwave Technology, Volume: 20, Issue: 4, April 2002. Pages: 574-586.

[58] Duser, M.; Bayvel, P.;"Performance of a dynamically wavelength-routed optical burst switched network," IEEE , Photonics Technology Letters, Volume: 14 , Issue: 2 , Feb. 2002, Pages:239 – 241

[59] Hauser, O.; Kodialam, M.; Lakshman, T.V.;"Capacity design of fast path restorable optical networks," INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies, Volume: 2, June 2002, Pages: 817 - 826

[60] Assi, C.; Ye, Y.; Shami, A.; Dixit, S.; Ali, M.;"efficient path selection and fast restoration algorithms for shared restorable optical networks" IEEE International Conference on Communications, 2003. ICC '03., Volume: 2, May 2003, Pages:1412-1416

[61] Ali, M "Shareability in optical networks: beyond bandwidth" optimization". IEEE Communications Magazine, , Volume: 42 , Issue: 2 , Feb 2004 pages:S11 - S15

[62] B. Meagher, G.K. Chang, G. Ellinas, Y.M. Lin, W. Xin, T.F. Chen, X. Yang, A. Chowdhury, J. Young, S.J. Yoo, C. Lee, M.Z. Iqbal, T. Robe, H. Dai, Y.J. Chen, and W.I. Way, "Design and implementation of ultra-low latency optical label switching for packet-switched WDM networks", Journal of Lightwave Technology, Volume: 18, Issue 12, , Dec 2000, Pages: 1978-1987

[63] Assi, C.; Yinghua Ye; Shami, A.; Dixit, S.; Habib, I.; Ali, M.A.; "On the merit of IP/MPLS protection/restoration in IP over WDM networks," IEEE Global Telecommunications Conference, 2001. GLOBECOM '01, Volume: 1, Nov. 2001, Pages: 65 – 69

[64] Qin Zheng; Mohan, G.; "Protection approaches for dynamic traffic in IP/MPLS-over-WDM networks," IEEE Communications Magazine, Volume: 41 , Issue: 5 , May 2003, Pages:S24 - S29

[65] V. Sharma and F. Hellstrand (Editors), "A framework for MPLS-based recovery," RFC 3469, February 2003

[66] K. Owens, V. Sharma "Network Survivability Considerations for Traffic Engineered IP Networks", Internet draft (March 2000)

[67] Ziying Chen "The LSP Protection/Restoration Mechanism in GMPLS" A Master Project , University of Ottawa, 2002

[68] T. Eilam, S. Moran, and S. Zaks. "Approximation algorithms for survivable optical networks", In The 14th international Symposium on Distributed Computing (DISC), 2000, pp. 104-118.

[69] E. Modiano and A. Narula-Tam, "Survivable routing of logical topologies in WDM networks," Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies., Volume: 1, April 2001, Pages: 348 - 357

[70] Pin-Han Ho and H. T. Mouftah, "A Framework of a Survivable Optical Internet using Short Leap Shared Protection (SLSP)," IEEE Workshop on High Performance Switching and Routing, 2001, May 2001 Pages:21 - 25

[71] Pin-Han Ho; Tapolcai, J.; Mouftah, H.T.; Chi-Hsiang Yeh "Linear formulation for path shared protection" IEEE International Conference on Communications, 2004 Volume: 3, 20-24 June 2004 Pages:1622 – 1627

[72] Dahai Xu; Chunming Qiao; Yizhi Xiong "An ultra-fast shared path protection scheme - distributed partial information management, part II"; 10th IEEE International Conference on Network Protocols, 2002. Proceedings. 12- 5 Nov. 2002, Pages:344 - 353

[73] Hungjen Wang; Modiano, E.; Medard, M.;Computers "Partial path protection for WDM networks: end-to-end recovery using local failure information" Seventh International Symposium on Communications, 2002. Proceedings. ISCC 2002., 1-4 July 2002 Pages:719 - 725

[74] Ajay Todimala and Byrav Ramamurthy "A Dynamic Partitioning Sub-Path Protection Routing Technique in WDM Mesh Networks" Proceedings of the 15th international conference on Computer communication, 2002, Pages: 327 - 340

[75] Tapolcai, J.; Pin-Han Ho "A deeper study on segment shared protection"; 7th international Symposium on Parallel Architectures, Algorithms and Networks, 2004., 10-12 May 2004 Pages:319 - 325

[76] W. D. Grover, J. Doucette, "Topological design of survivable mesh-based transport networks," Annals of Operations Research, special issue on Topological Network Design in Telecommunication Systems, in press, Volume: 106, Sept 2001, Pages: 79-125

[77] G. Sai Kiran Reddy, C. Siva Ram Murthy, "Reconfiguration Based Failure Restoration in Wavelength-Routed WDM Networks " International Conference on Dependable Systems and Networks, 2000. June 2000 Pages: 543 – 552

[78] E. Kozlovski and P. Bayvel, "Link Failure Restoration in Wavelength-Routed Optical Burst Switched (WR-OBS) Networks". Optical Fiber Communications Conference, 2003. OFC 2003, March 2003 Pages:222 – 223

[79] Neuts, M.; Rosberg, Z.; Hai Le Vu; White, J.; Zukerman, M "Performance enhancement of optical burst switching using burst segmentation".; IEEE International Conference on Communications, 2003. ICC '03., Volume: 3, May 2003 Pages:1828-1832.

[80] M. Jin and O. W. W. Yang "An IP Support Technique in the Optical Internet", the 22nd Biennial Symposium on Communications, Queen's University, May 31- June 3, 2004

[81] R. Ramaswami and K. Sivarajan, "Routing and wavelength assignment in all-optical networks, " IEEE/ACM Transactions on Networking, Volume, 3, Issue 5, 1995, Pages: 489-500.

[82] Z. Zhang, A. Acampora. "A Heuristic Wavelength Assignment Algorithm for Multihop WDM Networks with Wavelength Routing and Wavelength Re-Use," ACM/IEEE Transactions on Networking, Volume: 3, Issue: .3, June 1995 Pages: 281-288

[83] R.M.Krishnaswamy, K.N.Sivarajan, "Design of Topologies: a Linear Formulation for Wavelegth Routed Optical Networks with No Wavelegth Changers", IEEE Infocom'98, San Francisco, Ca, USA, Mar 1998 Pages: 919-927.

[84] A. Maach and G.V. Bochmann " A Hybrid Architecture using Both Optical Burst Switching and Routed Wavelength," Communications and Computer Networks, Cambridge, USA, Nov. 2002 Pages: 263-268

[85] Chunsheng Xin; Chunming Qiao; Yinghua Ye; Sudhir Dixit "A hybrid optical switching approach" Global Telecommunications Conference, 2003. GLOBECOM '03. IEEE , Volume: 7 , 1-5 Dec. 2003 Pages:3808 - 3812

[86] A Maach, G v Bochmann, H Mouftah "Contention avoidance in optical burst switching" 3rd International Conference on Networking ICN'04 February – March, 2004

[87] R. Jain, K. K. Ramakrishnan, and D.-M. Chiu. (1988). "Congestion avoidance in computer networks with a connectionless network layer," Technical Report. DEC-TR-506, Digital Equipment Corperation.

[88] S. Floyd and K. Fall. "Promoting the use of en-to-end congestion control in the internet," IEEE/ACM Transactions on Networking, Volume: 7, Issue 4, 1999 Pages: 458 – 472.

[89] Floyd, S. and V. Jacobson.; "Random Early Detection Gateways for Congestion Avoidance," IEEE/ACM Transactions on Networking, Volume: 1, Issue: 4, 1993, Pages: 397 - 413.

[90] Shudong Jin, Liang Guo, Ibrahim Matta, and Azer Bestavros; "A spectrum of TCP-friendly window-based congestion control algorithms," IEEE/ACM Transactions on Networking (TON), Volume: 11, Issue: 3, Jun. 2003, Pages: 341 – 355

[91] K. Laberteaux, C. Rohrs, and P. Antsaklis. "A Practical Controller for Explicit Rate Congestion Control," IEEE Transactions on Automatic control, Volume 47, 2002, Pages: 960-978.

[92] Kenneth P. Laberteaux, Charles E. Rohrs, and Panos J. Antsaklis.). "An Adaptive Inverse Controller for Explicit Rate Congestion Control with Guaranteed Stability and Fairness" International Journal of Control, Volume: 76, Issue 1, 2003, Pages: 24-47.

[93] J. Padhye, J. Kurose, D. Towsley, and R. Koodli. "A model based TCP-friendly rate control protocol," in Proceedings of International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV), 1999

[94] C. Su, G. de Veciana, and J. Walrand; "Explicit rate flow control for ABR services in ATM networks," IEEE/ACM Transactions on Networking, Volume 8, Issue 3, 2000, Pages: 350 – 361.

[95] A Maach, G v Bochmann, H Mouftah "Robust optical burst switching," 11th international telecommunications network strategy and planning symposium Jun. 2004, Pages: 447-452

[96] A Maach, G v Bochmann, H Mouftah "Congestion Control and Contention Elimination in Optical Burst Switching," Telecommunication Systems Journal 2004

[97] A Maach, H Zeineddine, G v Bochmann "bandwidth allocation scheme in optical TDM," 7th IEEE international conference, HSNMC 2004, Jun. 2004, Pages: 801-812

[98] Liew, S.Y.; Chao, H.J.; "On slotted WDM switching in bufferless all-optical networks," 11th Symposium on High Performance Interconnects, Aug. 2003, Pages: 96-101

[99] Ramamirtham, J.; Turner, J.; "Time sliced optical burst switching," INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies, Volume, 3, Issue 30, March - April 2003, Pages: 2030 - 2038

[101] J. Yates, J. Lacey, D. Everitt, M. Summerfield, "Limited-range wavelength translation in all-optical networks", in Proceedings of INFOCOM, 1996, Pages: 954-961.

[102] R. Ramaswami, G. H. Saski, "Multiwavelength optical networks with limited wavelength conversion," in Proceedings of INFOCOM 97, 1997, Pages: 490-499.

[103] H. Zeineddine, A. Jakel, S. Bandyopadhayay, a. Sengupta "Efficacy of Wavelength Translation in all-Optical Networks," Proceedings, International Conference on Computing and Information (ICCI), June 1998, Pages: 43-50.

[104] Gaoxi Xiao; Kejie Lu; Chlamtac, I "An evaluation of distributed wavelength provisioning in WDM optical networks with sparse wavelength conversion"; Lightwave Technology, Journal of , Volume: 22 , Issue: 7 , July 2004 Pages:1668 – 1678

[105] Hongwei Zhang, Arjan Durresi, "Differentiated Multi-Layer Survivability in IP/WDM Networks", 8th IEEE-IFIP Network Operations and Management Symposium (NOMS 2002), Florence, Italy, April 15-19, 2002 Pages: 681-694

[106] P. Demeester, M. Gryseels, A. Autenrieth, C. Brianza, L. Castagna, G. Signorelli, R. Clemente, M. Ravera, A. Jajszczyk, D. Janukowicz, K.V. Doorselaere, Y. Harada, "Resilience in Multilayer Networks", IEEE Communications Magazine, August 1999, Pages: 70-76,

[107] Georgios Ellinas, Eric Bouillet, Ramu Ramamurthy, Jean-Franois Labourdette ; "Restoration in Layered Architectures with a WDM Mesh Optical Layer", (Invited Paper) IEC Annual Review of Communications, June 2002.

[108] Gerstel, "Opportunities for Optical Protection and Restorations", Proc., OFC '98, San Jose, CA, Volume: 2, February 1998 Pages: 269-270,

[109] Ramamurthy and B. Mukherjee, "Survivable WDM mesh networks, part I -- Protection, " Proceedings of IEEE INFOCOMM '99, March 1999, Pages: 744--751,

[110] Fumagalli, A.; Valcarenghi, L " IP restoration vs. WDM protection: is there an optimal choice?"IEEE Network, Volume: 14, Issue: 6, Nov.-Dec. 2000, Pages: 34 – 41

[111] Pin-Han Ho; Mouftah, H.T.;"Shared protection in mesh WDM networks" Communications Magazine, IEEE , Volume: 42 , Issue: 1 , Jan 2004 Pages:70 - 76

[112] Pin-Han Ho; Tapolcai, J.; Mouftah, H.T "On achieving optimal survivable routing for shared protection in survivable next-generation Internet" Reliability, IEEE Transactions on , Volume: 53 , Issue: 2 , June 2004 Pages:216 – 225

[113] Pin-Han Ho "Segment shared protection in WDM networks with partial wavelength conversion" Communications Letters, IEEE , Volume: 8 , Issue: 6 , June 2004 Pages:394 – 396

[114] A Maach, G v Bochmann, H Mouftah "Shared protection for time slotted optical networks" 3rd IEEE International Symposium on Network Computing and Applications NCA'04 Cambridge, USA Sept 2004 Pages:333-338