UPC E. T. S. d'Enginyeria de Telecomunicació
Barcelona

PROJECTE FI DE CARRERA

Eindhoven University of Technology
Faculty of Electrical Engineering
Division of Telecommunication Technology and Electromagnetics

Architecture and Management Of Optical Burst Switching Nodes

Gustavo Gil Lopez
Graduation Thesis

Performed at:	Electro-Optical Communication Group
Period:	October 2001-June 2002
Graduate Professor:	Prof. ir. A. M. J. Koonen
Supervisors:	Dr. ir. I. Tafur Monroy
	Dr. ir. H. de Waardt

Escola Tècnica Superior d'Enginyeria de Telecomunicació de Barcelona
UNIVERSITAT POLITÈCNICA DE CATALUNYA

FULL NÚMERO 1
CURS: ....

a) A complimentar per l'estudiant:

Dades de l'estudiant:

COGNOMS: GIL LÓPEZ
NOM: GUSTAVO
TITULACIÓ: INGENIERÍA SUPERIOR DE TELECOMUNICACIONES
ESPECIALITAT (si escau): ....
PLA D'ESTUDIS: 92

Dades del Projecte de Fi de Carrera:

TÍTOL: ARCHITECTURE AND MANAGEMENT OF OPTICAL BURST SWITCHING NODES

PARAULES CLAU: IST, STOLAS, GMPLS, MANAGEMENT SYSTEM, OPTICAL BURST SWITCHING, RSVP TE, CR-LDP

RESUM: DENTRO DEL PROYECTO IST-STOLAS QUE INTRODUCE UN NOVEDOSO ESQUEMA DE MODULACIÓN ORTOGONAL PARA LA ASIGNACIÓN DE "LABELS" ESTE PROYECTO HACE UN ESTUDIO DE LAS POSIBILIDADES DE USO DE LA TÉCNICA OBS Y DE SU ADAPTACIÓN AL PROYECTO. ARQUITECTURA DE NODO NECESARIA PARA CUMPLIR ESPECIFICACIONES Y ANÁLISIS DEL MANAGEMENT SYSTEM

DIRECTOR/A/S/ES: IDELFONSO TAFUR MONROY, TON KOONEN (TU/e)

PONENT (si escau): ....

TU/e technische universiteit eindhoven

July 8, 2002

TO WHOM IT MAY CONCERN

It is a pleasure to me to introduce to you Gustavo Gil Lopez. Gustavo Gil Lopez performed his graduation project work, under my supervision, at the Electro-Optical Communications Group of the Eindhoven University of Technology, Eindhoven, The Netherlands, while being an Erasmus exchange student from the Polytechnic University of Catalonia, Barcelona, Spain. He performed his graduation work during the period of October 01, 2001 to June 27, 2002. The title of his graduation thesis is ***"Architecture and Management of Optical Burst Switching Nodes"***.

Gustavo Gil Lopez is an extremely high motivated, intelligent, enthusiastic and resourceful person. During his graduation work Gustavo Gil Lopez showed the capability to perform research work independently and to be able to communicate his result to others in the group. I am sure that Gustavo Gil Lopez is able to deliver an excellent contribution the professional activities he decides to pursue. I am therefore reachable for further enquiries or references on Gustavo Gil Lopez, would you consider it necessary.

Sincerely yours,

Dr.ir Idelfonso Tafur Monroy
Assistant professor

Eindhoven University of Technology
Telecommunications Technology
and Electromagnetics
P.O Box 513, EH 11:08
5600 MB Eindhoven
The Netherlands

Tel: + 31-40-2473219
Fax: +31 -40 -2455197

TU/e technische universiteit eindhoven

July 5, 2002

TO WHOM IT MAY CONCERN

This is to certify that **Mr. Gustavo Gil Lopez** has performed his graduation project work at the Electro-Optical Communications Group of the Eindhoven University of Technology, Eindhoven, The Netherlands, during the period of October 01, 2001 to June 27, 2002. The title of his graduation thesis is ***"Architecture and Management of Optical Burst Switching Nodes"*** under supervision of Professor A.J.M Koonen and dr.ir. Idelfonso Tafur Monroy. His project work has been graded with a mark of 7 in a scale of 1-10 of the Dutch University evaluation system. This grade (7) is a very good mark as for instance, the mark 10 is rarely given to students at our group and 9 is granted only to excellent work.

Dr.ir Idelfonso Tafur Monroy

TECHNISCHE UNIVERSITEIT EINDHOVEN Faculteit ELEKTROTECHNIEK

Eindhoven University of Technology
Telecommunications Technology
and Electromagnetics
Electro-Optical Communications Group
Tel: +31-40-2473219
Fax:+31-40-2455197

Escola Tècnica Superior d'Enginyeria
de Telecomunicació de Barcelona
UNIVERSITAT POLITÈCNICA DE CATALUNYA

FULL NÚMERO 1
CURS: ....

a) A complimentar per l'estudiant:

Dades de l'estudiant:

COGNOMS: GIL LOPEZ
NOM: GUSTAVO
TITULACIÓ: INGENIERIA SUPERIOR DE TELECOMUNICACIONES
ESPECIALITAT (si escau): ....
PLA D'ESTUDIS: 92

Dades del Projecte de Fi de Carrera:

TÍTOL: ARCHITECTURE AND MANAGEMENT OF OPTICAL BURST SWITCHING NODES

PARAULES CLAU: IST, STOLAS, GMPLS, MANAGEMENT SYSTEM, OPTICAL BURST SWITCHING, RSVP-TE, CR-LDP

RESUM: DENTRO DEL PROYECTO IST-STOLAS QUE INTRODUCE UN NOVEDOSO ESQUEMA DE MODULACIÓN ORTOGONAL PARA LA ASIGNACIÓN DE "LABELS" ESTE PROYECTO HACE UN ESTUDIO DE LAS POSIBILIDADES DE USO DE LA TÉCNICA OBS Y DE SU ADAPTACIÓN AL PROYECTO. ARQUITECTURA DE NODO NECESARIA PARA CUMPLIR ESPECIFICACIONES Y ANALISIS DEL MANAGEMENT SYSTEM.

DIRECTOR/A/S/ES: IDELFONSO TAFUR MONROY, TON KOONEN (TU/e)

PONENT (si escau): ....

FULL NÚMERO 2

b) A complimentar pel tribunal:

NOTA DEL PROJECTE DE FI DE CARRERA

EL/LA VOCAL	EL/LA PRESIDENT	EL/LA SECRETARI/A
Sr/a:	Sr/a:	Sr/a: ....

Barcelona, ...

Eindhoven University of Technology
Faculty of Electrical Engineering
Division of Telecommunication Technology and Electromagnetics

Architecture and Management Of Optical Burst Switching Nodes

Gustavo Gil Lopez
Graduation Thesis

Performed at:	Electro-Optical Communication Group
Period:	October 2001-June 2002
Graduate Professor:	Prof. ir. A. M. J. Koonen
Supervisors:	Dr. ir. I. Tafur Monroy
	Dr. ir. H. de Waardt

Abstract

El crecimiento exponencial del tráfico en Internet requiere una adaptación de los protocolos. Para conseguir éste reto está emergiendo rápidamente el Multiprotocol Label Switching (MPLS) como la tecnología que permite mejorar la velocidad y escalabilidad, así como la provisión de servicio para Internet. Por otro lado, los nuevos desarrollos de la tecnología WDM han permitido un incremento increíble en la capacidad de las redes ópticas. Como el procesado electrónico se ha vuelto el punto débil, desde el punto de vista de velocidad, en las redes, se está investigando en la manera de introducir más inteligencia en el plano de control de los sistemas de transporte óptico, que les permitirá ser más flexibles, controlables y abiertos a las técnicas de ingeniería de tráfico. Los esfuerzos se están centrando en la integración eficiente de las capas de IP y WDM. La integración directa y el transporte está sopoprtado por Multiprotocol Lambda Switching (MPλS), el cual está basado en MPLS, y generaliza el uso del concepto "label-switching" para incorporar "optical label-switching".

Una de las técnicas prometedoras para soportar redes de IP sobre WDM es la llamada Optical Burst Switching (OBS). OBS, explotando los conceptos de wavelength division multiplexing (WDM) encaja perfectamente con el concepto de MPλS, permitiendo al mismo tiempo escalar hasta altas velocidades y eliminar los potenciales cuellos de botella del procesado de las cabeceras electrónico, como en el típico Optical Packet Switching. Por ejemplo, en el reciente proyecto IST-STOLAS, los nodos cross-connect pueden hacer uso de los mensajes MPLS para controlar el "burst-switching", administración del ancho de banda, reservas y conversión de longitudes de onda.

En éste proyecto primeramente, se introduce una visión general del concepto de OBS. Para describir la arquitectura del nodo de STOLAS se ha incluido una analogía entre el nodo OBS y el de STOLAS. Después de ésta descripción, hemos hecho hincapié en los principales bloques que forman el "burst-switching node" de STOLAS, array waveguide grating (AWG), buffer mediante FDL, fuentes láser y los controladores, para encontrar los parámetros óptimos para su funcionamiento de acuerdo con las especificaciones adoptadas en el proyecto STOLAS. Esto significa, considerar las posibilidades de la arquitectura del nodo, específicamente consideraciones respecto a la tecnología utilizada, y los parámetros relacionados con las características del tráfico. Para poder llevar a cabo éste estudio se ha realizado una profunda investigación en las técnicas de OBS y en las características específicas de STOLAS. Este estudio nos ha mostrado que no es posible el uso del típico esquema de OBS para el proyecto STOLAS, ya que en el proyecto la combinación de time-offset y modulación ortogonal no son compatibles. Sin embargo, debido al novedoso esquema de modulación ortogonal para asignación de "labels" en STOLAS, se deben de adaptar una serie de parámetros y bloques.

A continuación, una serie de protocolos de señalización MPLS han sido propuestos. Después de una breve introducción a los protocolos de señalización, describimos cuál podría ser la mejor solución para el problema de señalización dentro de la red. La propuesta está basada en una exhaustiva comparación entre los protocolos CR-LDP y RSVP-TE. Además, una parte de éste informe está dedicada a MPλS y la interconexión entre los planos de datos y control. Debido a que STOLAS no hace

reserva de recursos se deben de hacer una serie de modificaciones en los protocolos para poder adaptarlos.

Como tercer punto, se ha llevado a cabo un completo estudio de las caractristicas de Management System para STOLAS. Como el desarrollo de un nuevo sistema no es el objetivo del proyecto se ha adaptado el esquema del proyecto IST-WINMAN como Management System para STOLAS. Una lista de nuevos elementos (específicos para STOLAS) ha sido definida. Estos nuevos elementos tienen que ser controlados dentro de los nodos y a lo largo de la red, basados en un sistema de bases de datos que debe ser implementado dentro del proyecto.

Por último, este informe presenta una lista de conclusiones y recomendaciones para el trabajo futuro. Unas guías de cómo mejorar el funcionamiento del nodo OBS y del sistema de management están al igual incluidas.

Abstract

With the exponential growth in Internet traffic, scaling of Internet Protocol (IP) data networks far beyond their present performances is required. In order to meet this challenge, Multiprotocol Label Switching (MPLS) is rapidly emerging as the technology that allows for enhanced speed and scalability as well as service providing capabilities in the Internet. On the other hand, recent developments in WDM technology have dramatically increased the traffic capacities of optical networks. As the electronic processing is becoming the weak point for the bit-rate (speed) of the networks, research is ongoing to introduce more intelligence in the control plane of the optical transport systems, which will make them more survivable, flexible, controllable and open to traffic engineering techniques. The research efforts are focusing on the efficient interworking and integration of higher layers IP and WDM transport layer. Direct integration and transport is supported by Multiprotocol Lambda Switching (MPλS), which is based on MPLS, and generalizes the use of label-switching concept to incorporate optical label switching.

One promising technique to support IP over WDM networks is Optical Burst Switching (OBS). OBS, exploiting wavelength division multiplexing (WDM) fits well into the MPλS concept, while offering flexibility, enabling scaling to very high bit rates and eliminating potential bottlenecks in the electronic header processors like in Optical Packet Switching. In optical burst switched networks, MPLS protocol-related signaling can be carried using control channels. For instance, in the recent IST-STOLAS project, cross-connect nodes can make use of MPLS messages to control burst switching, bandwidth administration, delays reservations and wavelength conversion.

In this report, firstly, a general view of the OBS concept is introduced. In order to describe the STOLAS optical burst switching node architecture, an analogy between the OBS node and the STOLAS node has been included. After this description, we have focused on the main building blocks of the STOLAS burst witching node, array waveguide grating (AWG), fiber delay lines (FDL), tunable laser sources and controllers, to find the operation parameters for its performance according to the specification adopted within the STOLAS project. This in it turns, means to consider the node architecture, specifically issues related to the technology being use, and the parameters related to the data traffic characteristics. In order to realize such a description a deep study on the OBS node techniques and the STOLAS specific characteristics has been carried out. This study has shown that it is not possible to use the typical OBS scheme for the STOLAS project, since within the project time offset and orthogonal modulation are not compatible. However, due to the specific novel orthogonal modulation scheme to label the signals in STOLAS, a number of adaptations should be done.

Secondly, MPLS routing and signaling protocols have been proposed for STOLAS. After a brief introduction to the signaling protocols in MPLS, we described what could be best solution for the signaling problem inside the network. The proposed solution is based on an exhaustive comparison between RSVP-TE and CR-LDP protocols. Moreover, a part of this report is also dedicated to MPλS and the control and

data plane internetworking. A conclusion drawn form this study is that because STOLAS is not using resource reservation an adaptation of signaling protocols has to be carried out.

Thirdly, a complete study of the characteristics of the Management System for STOLAS was accomplished. As the development of a network management system is not the focus of the project we adapted IST-WINMAN Management System approach for STOLAS. A list of new Elements (specific to STOLAS) has been made. These new elements have to be controlled inside the nodes and along the network based on a system of data bases that should be implemented for the STOLAS implementation.

Finally, this report presents a list of conclusions and recommendations for further work has been summarized. Guidelines on how to improve the performance of the burst switching node and management system are also included.

My own contribution within the project was, after assisting to many workshops and discussions about network and node architecture, to make an accurate inventory of the node architecture, also proposing some configurations. I have done a deep study on MPLS protocols and a first step on the adaptation for the Management System in STOLAS.

Index

0. Introduction

The main overall objective of the STOLAS project is to improve the throughput of packet switched networks, by providing novel optical routing techniques. Additional overall objectives are to develop key optical functionalities for optical packet routing, and to validate the technical feasibility in a limited-scale testbed.

The amount of packet-based data traffic in today's telecommunication networks is increasing steeply, and in many networks has already overtaken circuit-switched traffic. In order to answer the exploding demand for network capacity, wavelength division multiplexing (WDM) is being introduced consecutively in core networks and in metropolitan area networks, and will penetrate in access networks in the near future.

A novel approach termed lambda labeling or Multi-Protocol Lambda Switching (MPλS) has been presented for direct IP-over-WDM transport. The scheme has been derived from the IP-MPLS protocol and the label-switching concept to provision wavelength-switched channels; the label-switched paths (LSPs) are associated with optical channels. Optical packets (or bursts of packets) are switched to other wavelengths (label swapping or switching) at the transmitters in the edge nodes; wavelength routers in the core nodes guide the packets along the appropriate network paths. The label swapping or switching can be done by wavelength converters of which the pump signal is generated by a wavelength-tunable laser.

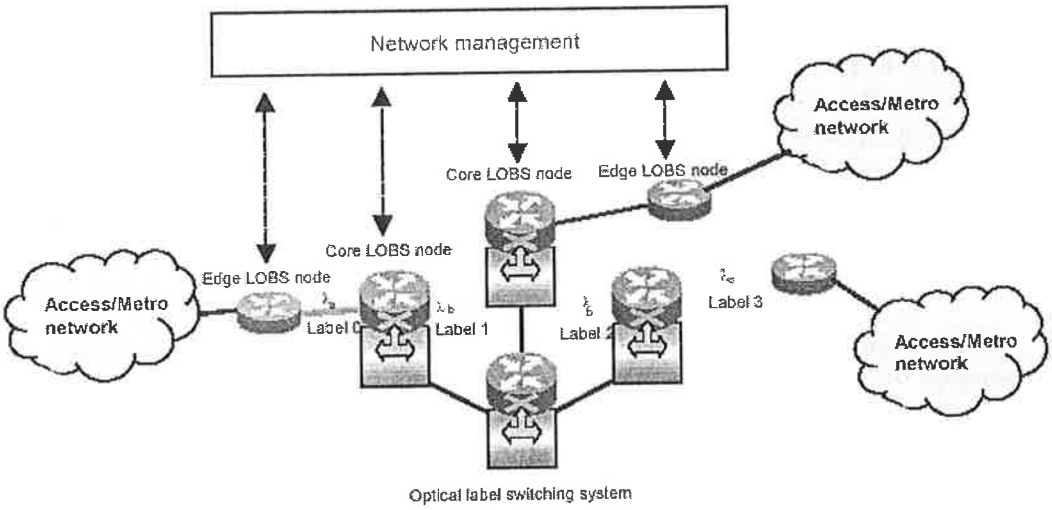


Fig.1. Burst IP packets over WDM network supported the GMPLS and OBS

In addition to optically labeling a packet (or burst of packets) by assigning a particular wavelength to it, a next-level label can be attached by using a particular modulation scheme. The latter can be done by modulating the label information orthogonally to the packet data. Having two optical labels (wavelength and DPSK/FSK) is in agreement with the label-stacking as foreseen in MPLS (i.e., 2 or more labels corresponding with a packet stream). The orthogonally modulated label also facilitates the implementation of multi-casting the streams. Furthermore, the orthogonal channel can be used as an embedded network management and control channel, e. g., for supervising the routing (crossconnect and add/drop) functions.

The STOLAS project will focus on some key issues for optical packet/burst routed networks:

- Tunable lasers for wavelength labeling which can be wavelength-switched very fast, including the control electronics.
- Wavelength converters for optical label swapping on two levels (wavelength, and an orthogonal modulation domain such as phase or frequency), including the label processing electronics.
- Optical-label-controlled crossconnects and add/drop nodes.
- 2R multi-wavelength regenerator
- Management and control of the optical labeling and routing functions, and the validation of the optical routing techniques in a small-scale testbed containing a few edge routers and core routers.

GMPLS includes wavelength switching next to label switching and fiber (space) switching. It provides quick and efficient forwarding of IP packets because it uses only a single forwarding algorithm based on label swapping (removal and reinsertion). The label is a short, fixed length value that is contained in each IP packet or forwarding equivalence class (FEC) and used to forward the packet through the network. Therefore, IP/GMPLS is a low latency, low overhead routing technique that simplifies packet forwarding and enables scaling to terabit rates.

GMPLS-based LOBS [16] provides an effective IP over WDM networking due to the advantages of optical burst switching (OBS). OBS, which represents a balance between circuit and packet switching, has advantages in comparison with optical packet switching (OPS), namely, a label and its corresponding burst data in LOBS can be much more loosely coupled in both space (by using separate label and data wavelength) and time (by using a nonzero offset time) than a label and its payload are in OPS, the requirements for processing label and synchronizing between burst (as well as between burst data and its label) in OBS can be much less stringent than those for processing label and synchronizing between payload (as well as between payload and its label) in OPS. In addition, OBS can provide differentiated services (priority scheme) by using an extra offset time. OBS can avoid optical buffer, as long the offset time is longer than the total delay from router followed by burst IP packet. OBS is able to offer delayed reservation through the offset time.

In the GMPLS-based LOBS, burst data can be composed by assembling several IP packets (in ingress LOBS node). It may contain several megabytes of data. However, a small size of bust data is still possible to occur, therefore, for efficiency of network throughput, the GMPLS-based LOBS with segmented burst data is preferred.

Two main approaches have been investigated for labeling of optical signals: bit serial label and optical sub-carrier multiplexed label. The former multiplexes a fixed bit rate label in front of the IP packet while the latter multiplexes the label information that is modulated a RF sub-carrier in the IP packets. Both approaches put the label at the same wavelength of the IP packet [17]. We propose a new approach, based on orthogonal modulation of the label information with respect to the modulation format of the IP payload.

CHAPTER 1:

Architecture for an Optical Burst Switch Node

1. NODE ARCHITECTURE FOR AN OPTICAL BURST SWITCH ROUTER

1.1. Generic Node Architecture

In a generic optical burst router three different parts can be distinguished: the input interface, the switching element and the output interface. Figure (1) shows the modular structure of a typical optical burst switch router.

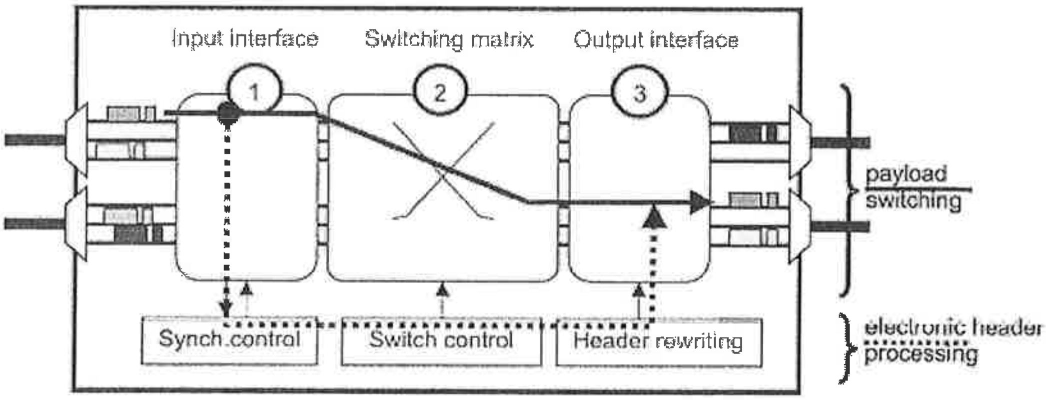


Fig (1). Generic Optical Burst router

1. The input interface consists of: the demultiplexer or channel mapper and the input FDLs, Later on, this chapter goes deeper is some of the elements.

2. The switching element consists: the switching matrix, the input wavelength converters, the SCU, the scheduler, the drop/multicast element and the output wavelength converters. The focus of this chapter is on this part of the router: the switching element.

3. The output interface that consists of: the header rewriter and the multiplexer.

Although the control element is placed in the switching element, it should be said that this component is connected to all the components.

Through this chapter we will make an analogy between the typical optical burst switch and the STOLAS core router, trying to identify the elements present in both and the consequences of their use.

The architecture of an N x M optical core router is shown in Fig. (2), which mainly consists of input FDLs (fiber delay lines), an optical switching matrix, a switch control unit (SCU), also called switch matrix controller, and routing and signaling processors. Data channels are connected to the optical switching matrix and control channels are terminated at the SCU. Channel mapping logically decouples the channels from physical fibers wavelengths. The input FDLs are used to delay the arriving data bursts and synchronize them, thus allowing the SCU to have enough time to process the associated BHPs. Data bursts still remain in the form of optical signals in the core routers. The optical buffers of FDLs are used to resolve data burst contentions on outgoing DCGs (data channel groups). The use of electronic buffers instead of FDL optical buffers was considered in [1]. Note that there are J incoming DCGs and J' outgoing DCGs in Fig.(2). A typical example of the general architecture is a symmetric router with equal number of input and output fibers is given in fig (2), where each fiber has one DCG of n wavelength channels and one CCG (control channel group) of *k* channels.

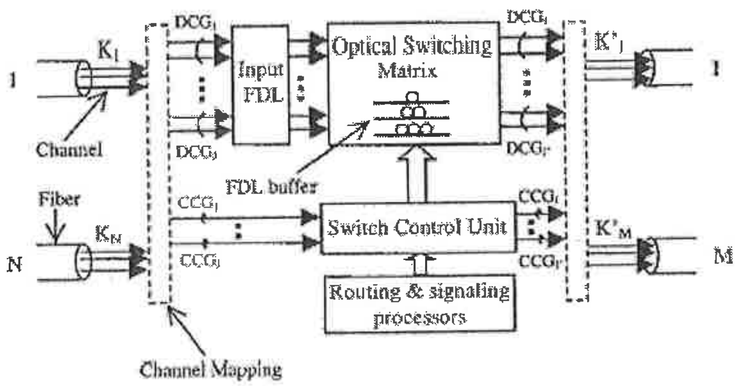


Fig (2). Optical Core Router architecture (general)
Addapted from figure references [3].

A block diagram of an N x N non blocking optical switching matrix is given in Fig.(3) where the spatial switch is able to switch a data burst from any incoming wavelength/channel to any FDL as long as it does not overlap with other data bursts. Each optical buffer has *B* WDM FDLs with i^{th} FDL being able to delay *Qi* time slots,1<=*I*<=*B* , and it is assumed that *Q1*<*Q2*<....<*Qb*. Note that an FDL in Fig. (2) has *(K – k)* wavelengths. By default there is always an FDL with zero delay time, denoted by 0 with *Q0* = *0*.

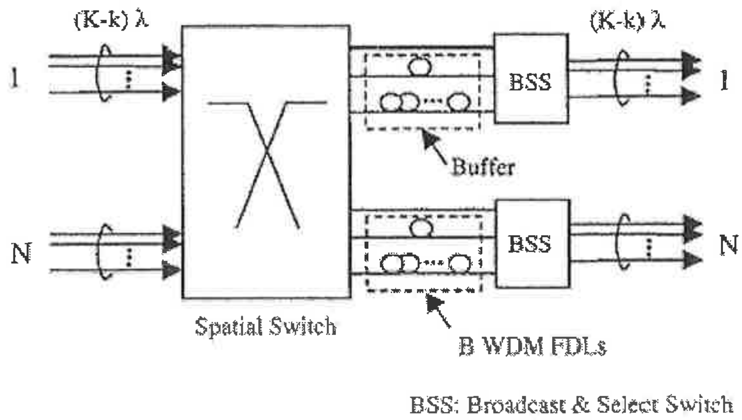


Fig. (3). Block diagram of a non-blocking (symmetric) optical switching matrix. Addapted from figure rferences [3].

The function of the SCU in Fig.(2) is similar to that of a conventional electronic router. The routing processor runs routing and control protocols for the whole OBS network. It creates and maintains a routing table and computes the forwarding table for the SCU. Forwarding can be connectionless or connection- oriented (prior path establishment through signaling). Apart from forwarding table lookup, the SCU decides on which outgoing DCG and CCG to forward each arriving data burst and its BHP. If free data and control channels are available from these groups, either when the data burst arrives at the optical switching matrix or after some delay in an FDL buffer, the SCU will select the FDL and configure the optical switching matrix to let the data burst pass through. Otherwise, the data burst is dropped. While arranging the transfer of a data burst and its corresponding BHP in the optical switching matrix and SCU, respectively, the SCU tries to resynchronize the data burst and the BHP by keeping the offset time τ as close as possible to $\tau 0$. If a data burst enters the optical switching matrix before its BHP has been processed (this phenomenon is called *early burst arrivals*), the burst is simply "dropped." This is because data bursts are optical analog signals. If no path is set up when a data burst enters the optical switching matrix, it is lost. Since a BHP and its data burst are switched in the SCU and the optical switching matrix, respectively, the delay Δ introduced by the input FDL should be properly engineered such that under the normal traffic condition data bursts are rarely dropped due to early arrivals.

The block diagram of a SCU is given in Fig.(4). Depending on the optical switching matrix, the SCU can have either a centralized configuration as shown in Fig. (4) or a distributed configuration.

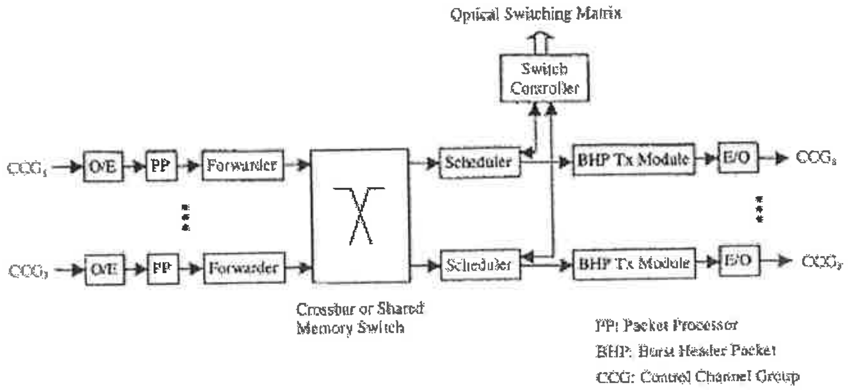


Fig. (4). block diagram of the switch control unit (centralized control). Addapted from figure references [3]

In a distributed configuration, each scheduler has its own switch controller. Distributed configuration could be applied to the broadcast-and-select type switches [2]. Here we focus on the description of a centralized configuration. The functionality of each building block in Fig. (4) is detailed below.

The forwarder performs the forwarding table lookup to decide on which outgoing CCG to forward the BHP. The associated data burst will be forwarded to the corresponding DCG. The mapping of logical channels to physical wavelengths and fibers is done in the forwarder. The forwarder then simply forward BHPs across the switch in a certain order (e.g., FIFO). To reduce the switch delay, it is preferred to use a switch with output-queueing discipline. To support multicast traffic, the switch requires native multicast capability. Otherwise, copies of a multicast BHP are made in the forwarder.

The scheduler in Fig.(4) is responsible for both the scheduling of the switch of the data burst on an outgoing data channel and the scheduling of the transmission of its BHP on an outgoing control channel. The scheduler is optical switching matrix specific. For the nonblocking optical switching matrix in Fig. (3), there is one scheduler for each DCG and CCG pair, and each scheduler only needs to keep track of the busy/idle periods of a single outgoing DCG and an outgoing CCG.

The scheduler works as follows. It first reads the time-stamp and the data burst duration information from a BHP to determine when the corresponding data burst will enter the optical switching matrix and how long the data burst will last. It then searches for an idle outgoing data channel time slot to carry the data burst, making potential use of the FDLs to delay the data burst. Once the idle outgoing data channel is found and the FDL to be used (if necessary) is determined, the scheduler knows the departure time of the data burst from the optical switching matrix. Subsequently, it schedules the time to send out the BHP on the outgoing CCG, trying to resynchronize the BHP with the data burst. After successfully scheduling the transfer of the data burst and its BHP, the scheduler will send the configuration information to the switch controller which will in turn configure, just-in-time, the optical switching matrix accordingly to let the data burst pass through. The configuration information includes incoming data channel identifier, outgoing data channel identifier, time to switch the data burst, duration of the data burst, and the FDL buffer identifier.

The scheduler is bi-directionally connected to the switch controller. After processing the configuration information sent by the scheduler, the switch controller sends back an acknowledgment to the scheduler. The scheduler then updates the state information of the DCG and CCG, modifies the BHP (e.g., the offset time and the data channel identifier) and passes it along with the time-to-send BHP information to the BHP transmission (Tx) module.

One of the reasons we need an acknowledgment from the switch controller is to limit the number of configuration requests from the schedulers. Hence, the maximum time for the switch controller to process the request can be estimated.

1.2. STOLAS node architecture

A possible architecture for such a node is made combining Array Waveguide Gratings (AWG) and all-optical tunable wavelength converters (TWC). We will see that an inherent problem of this switch is its internal blocking. This drawback can be greatly overcome by using an intelligent and efficient wavelength algorithm within the node.

The core element of the switching node is the switching matrix, and is also the most critical for switching speed. The time between the controller having calculated a new configuration for the switching matrix and the moment when the switch is stable in the new state must be short. Crucial for this element is also its transparency. The switch should just guide the light containing the payload information from input to output, without using any information on bitrate, data coding technique....

Firstly, we present the principle of operation of AWG, secondly we present several node architectures for the STOLAS node using AWG's plus tunable laser sources.

1.2.1. The Array Waveguide Grating with Tunable Wavelength Converter

An Array Waveguide Grating (AWG) is a diffractive element, and as such it doesn't contain any active devices, making it a very reliable component. It is sometimes also referred to as a PHASed ARray (phasar). We will shortly describe the functionality of an AWG, without giving any physical details (these can be found in e.g. [3]). The dashed box in Fig. (5) shows a black box model of an 8x8 AWG. Depending on the wavelength at input port i, the light comes out at output port j. This allows us to write out a simple matrix M, where the input ports are written vertically (index i) and the output ports horizontally (index j). An element M_{ji} in this matrix is the wavelength, which will come out on port j, if it entered the device at input port i. All these wavelengths are equally spaced by $\Delta\lambda$. An AWG has a cyclic nature: if $M_{ji}=\lambda_a$, there exists a λ_{cyc} for which $\lambda_b=\lambda\alpha+k.\lambda_{cyc}$ ($k \in \mathbf{Z}$, λ_{cyc} is called the period) shows the same behaviour as λ_a. In other words, if λ_b enters on port i, it will also exit on port j, just as it was the case for λ_a. It is possible to design the AWG in such a way that for an NxN AWG, the period is $N^*\Delta\lambda$. Putting all of this together, we come to a table like the one shown in Fig. 6. In this figure another property of the AWG can be seen. Let $\lambda_i=\lambda_0+i.\Delta\lambda$. If from a certain input port λ_i goes to output port k, then λ_{i+1} will go to output port (k+1)mod N. This is a physical constraint that has an impact on the available configuration options, since it limits the degree of freedom to a large extent.

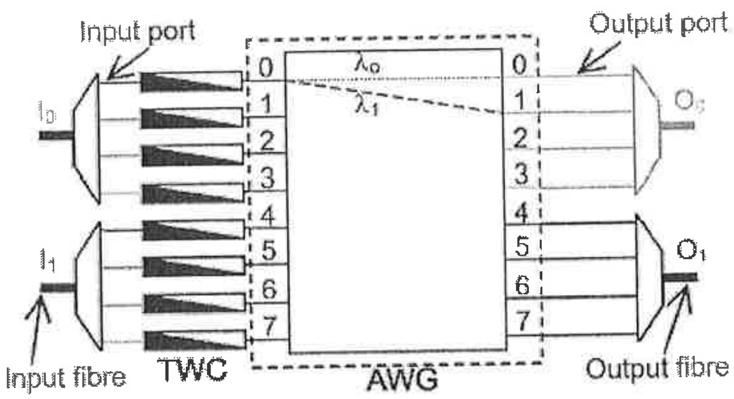


Fig. (5) STOLAS node: classic configuration.
Adapted from figure references [1].

Output ports \ Input ports	0	1	2	3	4	5	6	7
0	λ_0	λ_7	λ_6	λ_5	λ_4	λ_3	λ_2	λ_1
1	λ_1	λ_0	λ_7	λ_6	λ_5	λ_4	λ_3	λ_2
2	λ_2	λ_1	λ_0	λ_7	λ_6	λ_5	λ_4	λ_3
3	λ_3	λ_2	λ_1	λ_0	λ_7	λ_6	λ_5	λ_4
4	λ_4	λ_3	λ_2	λ_1	λ_0	λ_7	λ_6	λ_5
5	λ_5	λ_4	λ_3	λ_2	λ_1	λ_0	λ_7	λ_6
6	λ_6	λ_5	λ_4	λ_3	λ_2	λ_1	λ_0	λ_7
7	λ_7	λ_6	λ_5	λ_4	λ_3	λ_2	λ_1	λ_0

Fig. 3 An AWG table representation

As an example two configurations of the AWG are shown on both Fig.(5) and Fig.(6), in order to clarify this notation further. The signal at input port 4 can be sent to output port 4 if it were on λ0, (dotted line) while it would end up at output port 5 if the wavelength is λ1(full line). I will continue working with this example of an 8x8 AWG, and a switch based upon it.

The table learns that only converting these input ports to either *l1* or *l3* can do this. So we have a problem, since three ports are contending for only two wavelengths, resulting in the loss of one of the input ports, since this loss is not due to contention (problems between wavelengths and packets destined in a same output fiber) is due to internal blocking of this configuration of the switch, conversion to some wavelengths can degrade the function of the node. So it is clear that the way the output ports are combined into an out put fiber is very important and has a severe impact on the performance of the node.

Now that the concept of the node is clear, let's move on to the operation of the node. At a certain moment in time a header will arrive, indicating what to do with the associated payload. First the header is processed to determine the output fiber. Using the table for that specific output fiber, we need to find out to which wavelength has the signal to be converted by the TWC in order to switching to that correct output fiber. Thus, a wavelength assignment algorithm is needed, considering slotted operation, we know which input ports are active at the start time of every slot. It has to be said that the wavelength assignment algorithm should be done for each output fiber independently from each other. This assumes that each packet/burst has a unique output fiber, determined by a lookup table (case of GMPLS), and excludes the possibility of a routing mechanism.

Different approaches can be taken for this design, heuristic, random or maximum matching, for example:

- LUW-LFP (heuristic)
- Maximum Matching
- RORA (random)

In figure (7), a comparison of the blocking probability performance is illustrated, for a Poisson traffic and load of 0.8. Clearly both the heuristic and the maximum matching have about the same performance, if we consider the time of simulation and performance, heuristic is a lot slower than maximum matching, so we will consider Maximum Matching as our wavelength assignment algorithm.

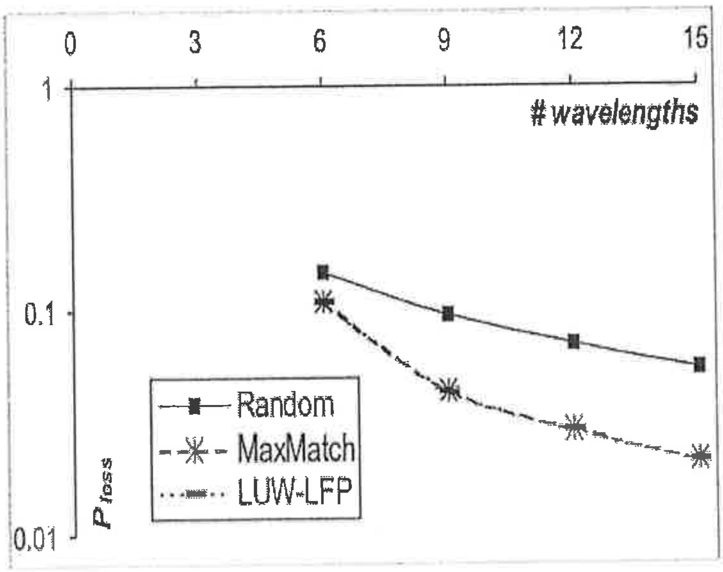


Fig. (7). Performance comparison of thed ifferent wavelength algorithm, depending on then umber of wavelengths per fiber. (IMEC simulations)

But in our "classic" STOLAS node, and because of the internal structure of the TWC, it is not available the conversion to the original wavelength, that means that the we have some unfairness performance of the node because some ports have less blocking probability than others. At low loads it provokes a severe deterioration of the node performance, even for higher loads. See figure (9).

	in							
out	0	1	2	3	4	5	6	7
0						λ_3		λ_1
1	λ_1	λ_0					λ_3	λ_2
2	λ_2		λ_0					
3	λ_3	λ_2	λ_1	λ_0				
4		λ_3		λ_1				
5			λ_3	λ_2	λ_1	λ_0		
6					λ_2		λ_0	
7					λ_3	λ_2	λ_1	λ_0

Fig (8).Table when conversion to original wavelength is not possible

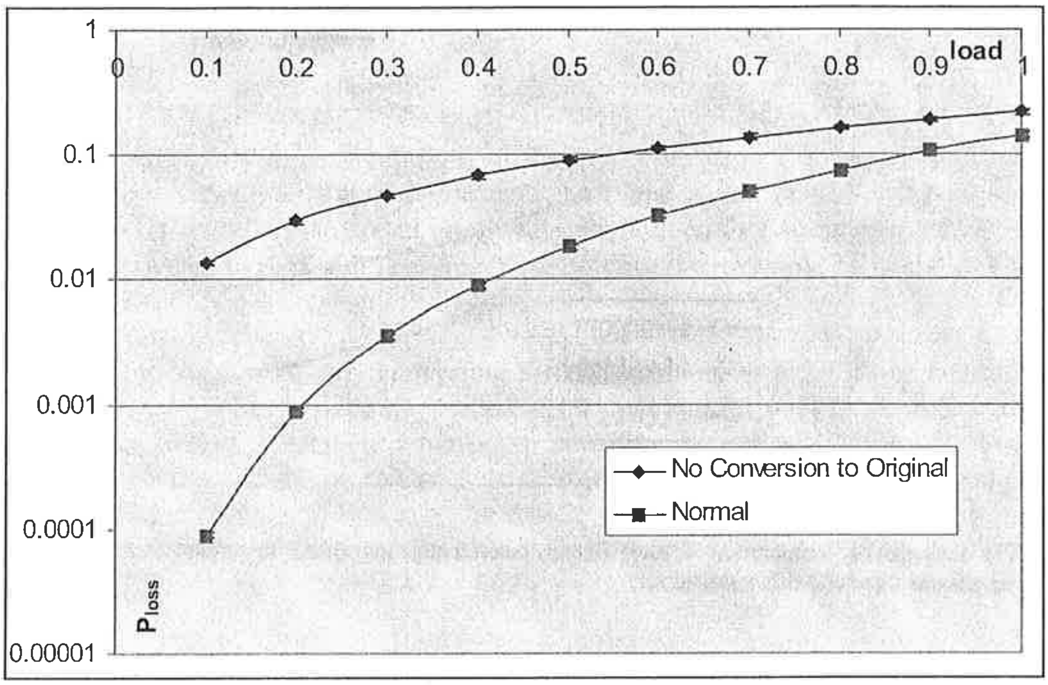


Fig. (9). Performance of a 2-fibre, 4-wavelengths STOLAS Node with and without Original Conversion. (IMEC simulations).

How can we solve the problems? let's look for different configurations of the node. There are other possibilities.

1.An alternative for this could be the over-dimensioning of the internal wavelengths in the node.

Instead of using only the W wavelengths, which are present on a fiber, use FW wavelengths (the dimension of the AWG). In this way the table of the AWG in the STOLAS node becomes:

		output ports							
		0	1	2	3	4	5	6	7
input ports	0	0	1	2	3	4	5	6	7
	1	7	0	1	2	3	4	5	6
	2	6	7	0	1	2	3	4	5
	3	5	6	7	0	1	2	3	4
	4	4	5	6	7	0	1	2	3
	5	3	4	5	6	7	0	1	2
	6	2	3	4	5	6	7	0	1
	7	1	2	3	4	5	6	7	0

The dark and light shadow denote a certain output fiber, note that any combination is now good in contrary to the "classic" case. We see that we can direct any input signal (less or equal than 4 of course) to different output ports. Only problem is that we no longer have wavelengths 0 to 3 at the output ports. So an extra wavelength conversion stage would be needed. And if they would be completely flexible, the dual labeling process would again become feasible. The node would then look like :

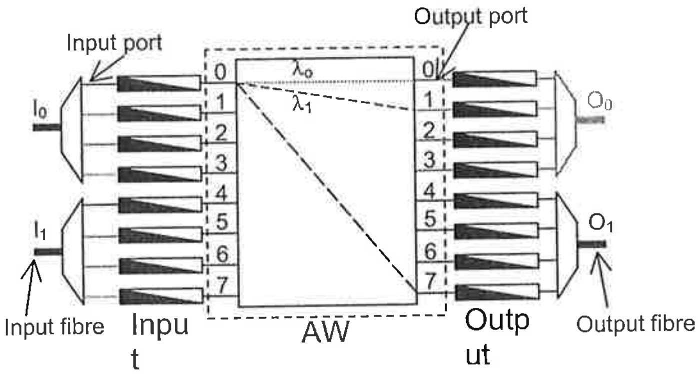


Fig. (10). Over-dimensioning of the internal wavelengths. Adapted from figure references[1].

The downside of this architecture is probably its scalability, an FWxFW AWG would be needed (can this behavior be emulated by some connection of smaller AWGs??). Notice however that although the number of TWCs needed is doubled, it stays a linear function!

Upgradability is also poor: if more wavelengths are to be used, the range of the TWC has to be increased and a larger AWG is needed. However this also seems the case for the classic STOLAS node architecture

2. Anoth er configuration that could avoid the scalability problems and technological limits is the multistage architecture. A 2-stage non- blocking solution, based on a three stage Clos switch is presented in figure (11).

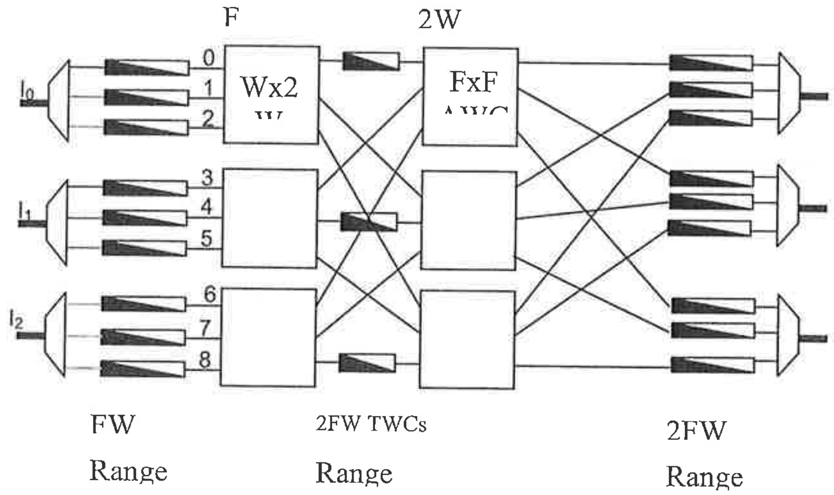


Fig. (11). 2-stage STOLAS architecture

This architecture is non-blocking, in the first stage an input port can be routed to any second stage switch, in these second stage, the output port of the AWG immediately determines the output fiber. Note that in the Clos three stage switch there would also be an AWG in the third stage, however switching signals into the same output fiber is done by the third stage TWCs.

1.3. Contention avoiding

For the contention problem in STOLAS, the proposal is to use a combination of accurate scheduling and retransmission (loop-back configuration) with FDL. We have to be sure that the edge nodes send data only at "good" moments in time, and refrain from sending at "bad" moments. That means that for each slot time every node has a schedule of how many wavelengths go to a certain output.

The scheduler is the key component in the SCU. The design of this block poses a new challenge, as it has to work at extremely high speed (100-200 ns per BHP) and the gaps introduced by the FDL will greatly complicate the design.

The core part of this scheduler is the data channel scheduling algorithm. For instance, LAUC-VF (*latest available unused channel with void filling*) algorithm can be used for this purpose. How it works? The gap between two data bursts in a channel is unused channel capacity, this algorithm try to fill the voids with new arriving data bursts.

The implementation complexity will depend on the DCG size, the BHP delay in the SCU, the maximum data burst length, the switched FDL design, and the data burst characteristics. To make the scheduling more strict many rules could be imposed on it in order to avoid interclass contention, like avoid first class traffic over x %. To do so, the scheduler keeps track of the utilization of the control and data channels.

To adapt the switch FDL module for back-loop configuration to the node the AWG should be over dimensioned with one or several inputs and outputs ports. In that way, data bursts affected by contention could be routed to those outputs to be looped-back through the AWG. The delay of the switched FDL will be set to accommodate the burst, depending on the length, by the SCU and the scheduler since all the necessary data is contained on the BHP. After the FDL module there should be another TWC capable to convert to all the different wavelengths in the node. The final scheme is shown in figure (12).

Another issue to take into account is that the scheduler has to be updated periodically managed by the management system, that means that the delay related to this operation could be a big concern.

One of the main parameters in OBS is the time offset between the header and the payload, that offset tries to avoid burst losses. In the STOLAS project there is a big handicap because our novelty is to send the header information in the signal phase orthogonal to the signal amplitude. But then the problem arises what to do when "no light" (no data) is preceding the burst, meaning that we have no amplitude to perform FSK on. This rises new issues.

One proposal solution is to use no-time offset and to use an optical buffer to store the payload while the header is being processed. So this buffer should be carefully dimensioned to support the different CoS and the variable burst length.

Within the STOLAS project further work should be done in this aspect to decide what kind of scheduling and control is the best solution for our system and node architecture.

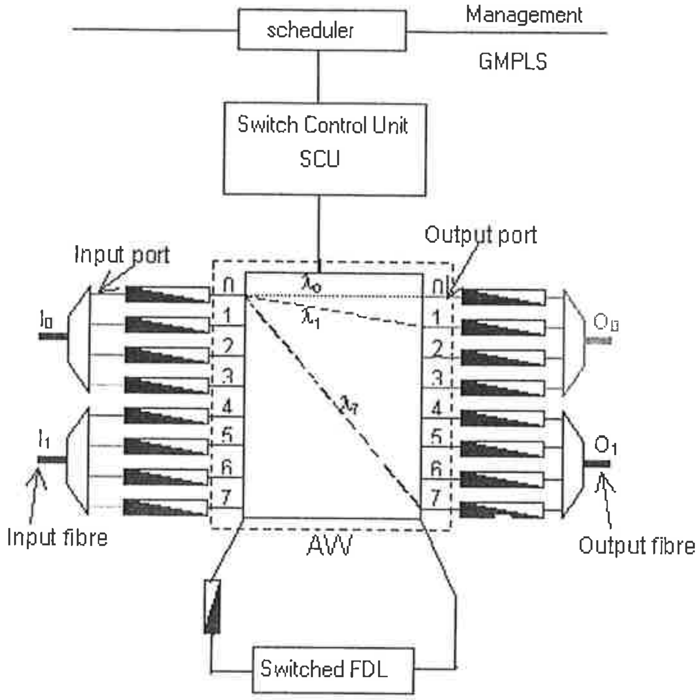


Fig. (12). Loop-back configuration for STOLAS node.

1.4. Burst length, timings and dimensioning

Except for the separate transmission of headers and payloads and being switching in different domains, there is no fundamental difference between packet switching and OBS.

The link utilization of the OBS network will largely depend on the number of channels dedicated to transmitting BHP (as well as other control packets) and the guards in each data burst.

Consider a WDM link having K channels with k control channels, and $K - k$ data channels, $1 \le k \le K$. Suppose the data channel rate is R Gb/s and the control channel rate is r Gb/s. The maximum link utilization will be η, defined as:

$$\eta = \frac{(K-k)R}{(K-k)R + kr}$$

As a data burst can be sent out on a data channel only if its BHP can be sent out on a control channel, there is a minimum requirement for the average data burst length in order to prevent congestion on control channels [4], so the performance of the control channel is strongly related to the average burst length. Since we will often deal with time domain issues in OBS, it is convenient to use time duration instead of bytes to represent lengths. The basic time unit is μs.

Let L_b be the average duration of a data burst (or $R^* L_b$ *kbits* in length) and L_h the average duration of a BHP. Consider that both control and data channel are fully loaded. Under this situation, the maximum average BHP transmission rate is k/ L_h BHPs per microsecond, and the maximum transmission rate is $(K - k)/ L_b$ data bursts per microsecond. Since $(K-k)/L_b \leq k/L_h$, we have:

$$Lb \geq \frac{(K-k)\text{Lh}}{\text{k}} \qquad (1)$$

We have to consider the guard period, say θ *μs*, for each data burst. The average data burst duration that actually carries user data would be only L_b- θ and the burst overhead is θ / L_b

$$Lb \geq \frac{\theta}{\varepsilon} \qquad (2)$$

Inequalities (1) and (2) will together determine the minimum average data burst length.

Let us look at the impact of the processing speed of the scheduler and switch controller on the burst arrival rate to an *N x N* optical router. Suppose the processing time is divided in three, detection, process and switch time. This processing time can be modeled as a single-server queuing system. Where λ is the average burst arrival rate per fiber. For a stable queuing system, the service rate must be larger than the arrival rate. We have $\lambda<1/\delta d$, $\lambda<1/(\delta s+X)$, and $\lambda<1/(N^*\delta c)$,which leads to:

$$\lambda < \min\left\{\frac{1}{\delta d}, \frac{1}{\delta_s + X'}, \frac{1}{N \cdot \delta c}\right\} \qquad (3)$$

where $X(\delta c < X < N^*\delta c)$ is the average time spent in the switch controller. For distributed control, $X=0$ and $N=1$. If ρ is the data channel utilization, given λ, the average burst duration (in *μs*) can be expressed:

$$Lb = \frac{\rho(K-k)}{\lambda} \qquad (4)$$

Note that inequalities (1) and (2) are from the burst transmission viewpoint, while (3) and (4) consider the switching aspect. The burst arrival rate is also closely related to the burst assembly mechanism used at edge router.

Since in the STOLAS project we are not going to use time offset the dimensioning of the input FDLs has to be done long enough to retain the bursts while the control packet is processed and the scheduler assigns the slot to go out.

Here the main problem is the dimension of the basic unit of timescale D. If the FDL buffer is designed with equally spaced lengths, consecutive multiples of D, small values of D lead to high time resolution and poor buffering capacity, while large values of D lead to large buffering capacity with poor time resolution. There is a trade-off between them to provide optimal burst loss probability.

The dimensioning of the FDL buffer may also influence the level of congestion in the control part of the router, so it can not be carried out without taking into account the control part of the architecture.

In principle the input FDL buffer may be variable or fixed length. The former may be done by an array of FDLs with different length and optical gates selecting the length. The delay could be chosen according to the estimated load in the control part of the router to limit the delay to a required minimum. This solution would significantly increase the complexity of the router. Furthermore, management of the input delays would be an additional task of no negligible logical complexity.

It is easier to assume a fixed delay length FDL, introducing a fixed delay of l μs. In this case, both implementation and management are straightforward. The only problem is to determine the value of l, and this one is closely related to the BCP processing time. If there is only one centralized queuing point for all BCPs, that is at the level of the forwarding table lookup, this point is just modeled as an *M/M/1* system, while the other delays can be characterized by a deterministic delay that can be added to the variable one. To model this variable part of the processing time, we assume that it has an average value equal to υ. So, the processing delay complementary distribution becomes:

$$\Pr(\delta > t) = e^{-(1/\upsilon - N(K-k)\lambda)(t-\upsilon^*)} \qquad (5)$$

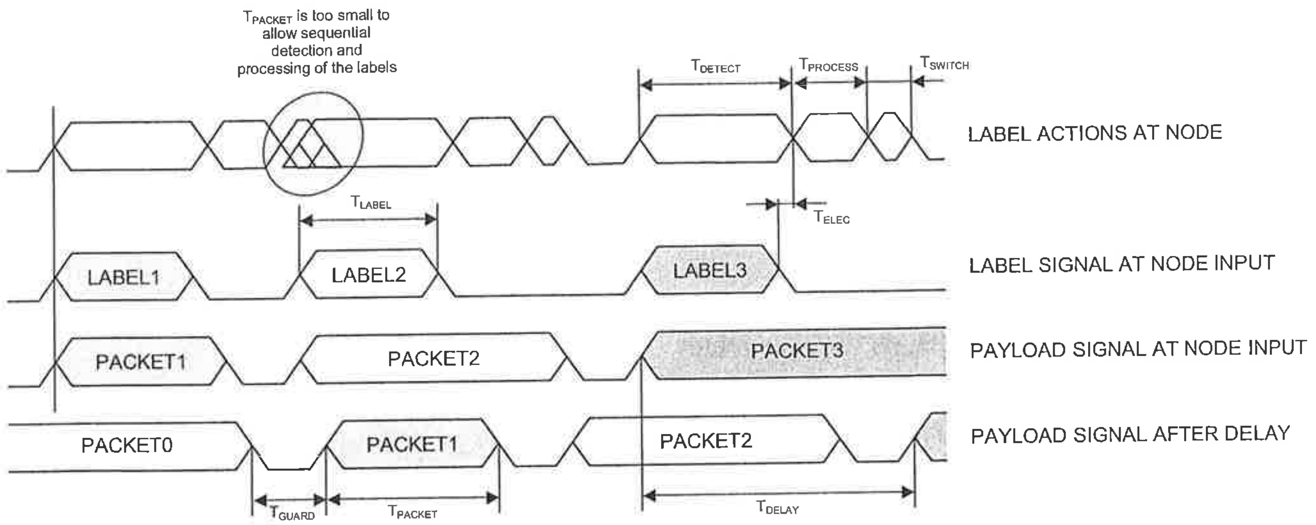
Once the QoS in terms of burst loss probability is given and using this formula, the FDL buffer length is easy to calculate. Anyway, we have to consider all the possibilities to calculate the optical buffer delay, and those are:

- Best case: only one timeslot of delay.
- Statistical: using formula.
- Worst case: upper bound, adding processing times.

The performance of the control channel is strongly related to the average burst length as shown in (1). A limit on the average burst length as a function of the processing capacity in the control logic may be calculated in a similar way, leading to a similar formula. Since the average BCP arrival rate at the burst processor is equal to the burst arrival rate and the BCP service rate is $1/\upsilon$, for a stable system we need $\upsilon < 1/N(K-k)\lambda$, meaning that the average burst arrival should be limited to :

$$\lambda < 1/N(K-k)\upsilon \tag{6}$$

A proposed diagram of timings for the STOLAS OBS concept is as follows:



And according to this diagram the timing equations are:

- $T_{DETECT} + T_{PROCESS} + T_{SWITCH} < T_{DELAY} > T_{DETECT} + T_{PROCESS} + T_{SWITCH} + T_{TOLERANCE}$
- $T_{GUARD} > T_{SWITCH} + T_{TOLERANCE}$
- $T_{PACKET} > T_{DETECT} +$
- $T_{PROCESS}$ (sequential detection and processing of labels)

$T_{PACKET} > T_{LABEL}$ (parallel detection and processing of N labels to be supported)

$$T_{ELEC} + T_{PROCESS} < (N - 1) \times (T_{LABEL} + T_{GUARD})$$

The parameters for STOLAS:

L_{bcp} = 14 bytes
r = 155Mb/s
L_b = 64 Kb (maximum size, variable length)
R = 10 Gb/s
F = 2 fibers
W= K = 4 wavelengths
k = 0 or 4 control channels
ρ = 0.85 or 1 (full loaded, worse case to make calculations)
BER (payload) = 10^{-9}
BER (label) = 10^{-12}
η = 99%
T detect ~1.65 μs
T proc ~ 1μs (including new orders to the TWC)
T switch ~ 1 μs (including TWC electronic)
T tol ~ 0.1 μs
T guard > 1.1μs
T_{bcp} = 0.72 μs
T_b = 51.2 μs
T_{elec} = 0.2 μs
T delay (for fixed length)
Worst case = 3.65 μs
Statistical = 2.65 μs (note: supposing υ=Telec and υ^*= 0.8)
Best case is not possible to calculate yet, more parameters in the node and network to provide different QoS must be established to an accurate assignment.

Note: Timings for detection, processing and switching are parameters fixed by Lucent Tech. since they are using an implemented system out of the STOLAS scope for calculations (picture from Lucent Tech.)

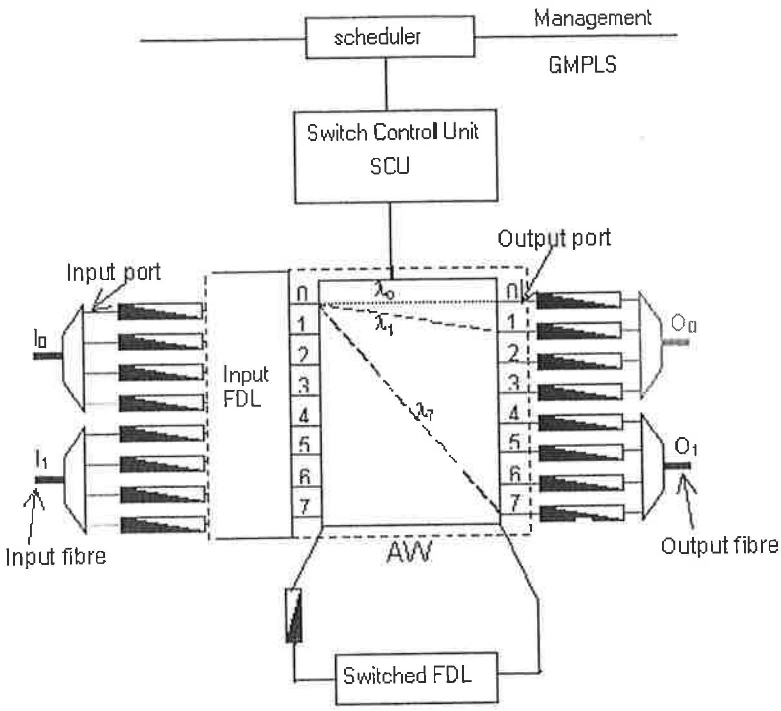


Fig . (13). Node architecture with Input FDLs

1.5. Other issues to take in account

Delay variation between nodes - The time for a packet to travel through a certain distance of the fiber depends on the fiber length, the chromatic dispersion, and the temperature variation. The proposal of using managed inter-node link delays (to make them equal to integer number of time slots) is not yet reasonably applicable with the current technology. When WDM is used the effect of chromatic dispersion has to be taken into consideration. Chromatic dispersion results in different propagation speed for packets transmitted on different wavelengths, therefore different propagation delays occur. For example, with a typical fiber dispersion of 20 ps/nm/km (where ps is the time unit for delay variation, nm the unit for wavelength difference and km the unit for propagation distance), a wavelength variation of 30 nm (consistent with the typical Erbium Doped Fiber Amplifier 1530 – 1560 nm window) and a propagation distance of 100 km, the propagation delay variation would be about 60 ns. If dispersion compensation fibers are used, the above delay variation can be reduced by one order of magnitude. The packet/burst propagation speed also varies with temperature, with a typical figure of 40 ps/0C/km. 100 km of fiber under temperature variation range of 0 – 25 0C means a delay variation of 100 ns. The delay variations mentioned above are relatively slow with respect of time; they can be compensated statically instead of dynamically (on a burst-by-burst basis).

Delay variations inside the nodes - What happens to each packet/burst depends on the switch fabric and contention resolution scheme. In a slotted network that uses fiber delay lines as optical buffers, a packet/burst can take different paths with unequal lengths within the switch fabric. All the considerations given in delay variations in the inter-node links apply here. It is worth mentioning that the fast time jitter induced by dispersion between different wavelengths and unequal optical paths varies from packet/burst to packet/burst at the output of the switch, therefore a fast output synchronization interface might be required. Thermal effects are smaller here because it varies more slowly and can be easily controlled within the node. In an optical switched network each switching node is operating with reference to its own internal clock. As a common practice in SDH/SONET, this clock is derived from a network synchronization signal distributed throughout the network. Phase noise of the oscillators accumulated along the clock distribution and thermal effects on the optical carriers can all contribute to the impairment of the synchronization signal. According to the SDH network synchronization standard, 1 msec is the maximum wander of the local node clock for a time duration larger than 1000 sec. Such slow phase variation has to be taken into consideration.

1.6. Conclusions

1.6.1. Comparison between Generic OBS Node & STOLAS Node concept

- While in Generic OBS there is an offset time between header and payload, in the STOLAS project this concept is avoided. Some implications are related to that decision, use of input FDLs for optical buffering that adds complexity to the control issues, appropriate FDLs lengths, performance of these FDLs, use of fixed or switched ones.

- Since we are going to use BCH channel orthogonal to payload channel: link utilization increases in comparison with other out-of-band signaling systems like SCM. Orthogonal signaling can scale better than SCM signaling. Due to SCM has problems with dispersion and RF processing of subcarriers at higher speeds, this system is not as scalable as STOLAS system. However, at high rates at the STOLAS label swapper the appearance of chirp could be a problem. More study has to be done to verify this issue.

- Link management and protection should be emphasize due to the high level of link utilization.

CHAPTER 2:

MPLS Routing and Signaling for STOLAS

2. MPLS Routing and Signaling. Basis for STOLAS

2.1. Introduction

Throughout this chapter we make an approach to the new protocol proposed for IP-integration-over-WDM and the relationship between MPLS and STOLAS.

The structure of this chapter is as follows: first an introduction to MPLS and the consequences that it supposes. In this first part we make also a comparison between RSVP-TE and CR-LDP signaling protocols for MPLS. The second part is dedicated to MPLambdaS and different control and data planes interworking, and the last part is dedicated to the Link management protocol and the importance of signaling in control management.

The telecom-style network management approaches extended to the IP layer with the cooperation of MPLS Protocol. The appropriate synergy and integration of the two layers is performed with management means capable of performing integrated provisioning of Label Switched paths over Optical paths. If we look into the relationship between MPLS and STOLAS we can see that: STOLAS is MPLS based with the use of OBS and Optical paths (WDM).

Management and adaptation are explained later on in next chapters.

2.2. Definition

Multiprotocol label switching (MPLS) is one of several initiatives to enable delivery on the promise of a converged network. Some of the new protocols' operations span layers 2 and 3. ATM is an example; MPLS is cited in some literature as another one. Strictly speaking, MPLS is a layer 3 protocol in that it does not define the critical function of layer 2 frame delineation. But we will treat it as a multilayer one. By combining the attributes of Layer-2 switching and Layer-3 routing into a single entity, MPLS provides the following benefits:

- enhanced scalability by way of switching technology
- class-of-service (CoS) and quality-of-service (QoS) based services (differentiated services)
- no need for an Internet protocol (IP)–over–asynchronous transfer mode (ATM) overlay model and its associated management overhead
- standards-based solution, promoting interoperability
- enhanced traffic-shaping and engineering capabilities

2.3. Reasons for using label switching

- Speed, delay, and jitter: Label switching is considerably faster than traditional IP forwarding. This speed translates into less delay in transporting traffic through the network. It also translates into less variable delay (jitter).
- Scalability: MPLS allows a large number of IPO addresses to be associated with one or few labels. This approach reduces further the size of address (actually label) tables.
- Resource consumption: label switching networks do not need a lot of the network's resources to execute the control mechanisms to establish label switching paths (LSPs).
- Route control: Most IP-based networks use the concept of destination-based routing, wherein the destination IP address in the IP datagrams determine the route through a network. Destination-routing is not always an efficient operation, and MPLS offers methods to use more efficient route control techniques, thus providing a higher level of service to the user.
- Traffic engineering: As a part of route control, many of the MPLS operations are assigned to allow the network provider to engineer the links and nodes in the network to support different kinds of traffic, as well as constrain the traffic to specific parts of the network.
- Labels and Lambdas: If a label switching is used in optical networks, it is possible to correlate (map) a label or labels to wavelengths, then use a XC switch for forwarding the traffic, thus reducing further the delay and jitter of user payload processing.

2.4. Types of MPLS nodes

In figure (1) the three types of MPLS nodes are shown. They perform the following functions:

- Ingress LSR: Device that operates at the edge of the access network and the MPLS network. Support multiple ports connected to dissimilar networks. Receives native-mode user traffic (for example, IP datagrams), and classifies it into an FEC [5]. It then generates an MPLS header and assigns it an initial label. The IP datagram is encapsulated into the MPLS packet, with the MPLS header attached to the datagram. If it is integrated with a QoS operation the ingress LSR will condition the traffic (such as using different queues for different priorities) in accordance with the DiffServ rules.

- Core/Transit LSR: High speed router that receives the packet and uses the MPLS header/label to make forwarding decisions. It will also perform label swapping (exchange label values). It is not concerned with processing the IP header, only the label header.

- Egress LSR: Performes the decapsulation operation.

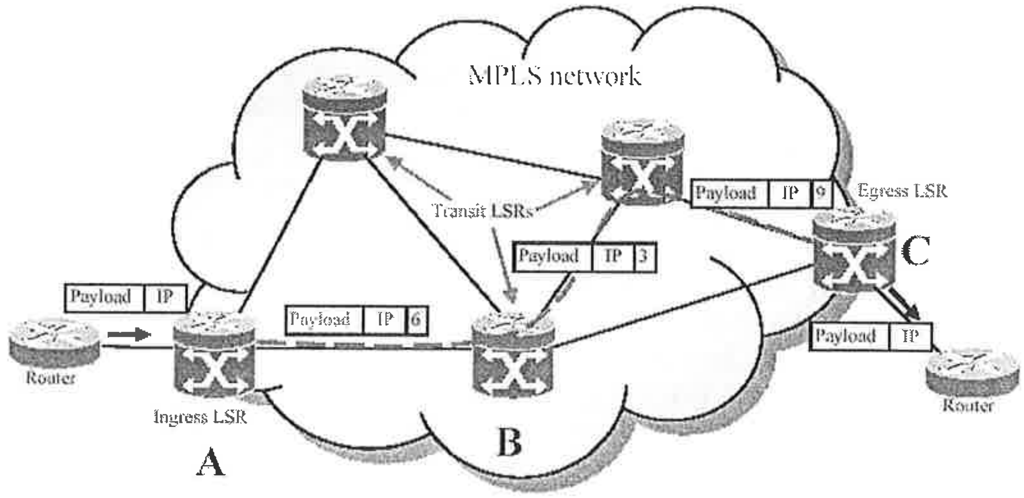


Fig.-1-. Different types of MPLS nodes

2.5. Label Distribution

In order that LSPs can be used, the forwarding tables at each LSR must be populated with the mappings from {incoming interface, label value} to {outgoing interface, label value}. This process is called LSP setup, or Label Distribution.

MPLS does not stipulate a specific label distribution protocol. Since several protocols are currently in operation that can support label distribution, it makes sense to use what is available. Nevertheless, the IETF has developed a specific label distribution protocol to complement MPLS that is called the label distribution protocol (LDP).

Another protocol, the constraint-based LDP (CR-LDP), is an extension to LDP. It allows the network manager to set up explicitly routed Label switched Paths (LSPs). CR-LDP operates independently of any internal gateway protocol (IGP). It is used for delay-sensitive traffic and emulates a circuit-switched network. CR-LDP is also designed to support traffic engineering operations.

RSVP can also be used for label distribution; this extension is called RSVP-TE, it supports label binding and distribution operations. Extensions to border gateway protocols (BGP) are yet another method for advertising and distributing labels. Later we will make a comparison of both protocols to see pros and cons.

Several different approaches to label distribution can be used depending on the requirements of the hardware that forms the MPLS network, and the administrative policies used on the network. The underlying principles are that an LSP is set up either in response to a request from the ingress LSR (downstream-on-demand), or pre-emptively by LSRs in the network, including the egress LSR (downstream unsolicited). It is possible for both to take place at once and for the LSP to meet in the middle.

In all cases, labels are allocated from the downstream direction (where downstream refers to the direction of data flow, and this means that are advertised towards the data source). Some possible options for controlling how LSPs are set up, and the protocols that can be used to achieve them, are described below.

- Hop-by-hop label assignment is the process by which the LSP setup requests are routed according to the next-hop routing towards the destination of the data. LSP setup could be initiated by updates to the routing table, or in response to a new traffic flow. The IETF MPLS Working Group has specified (but not mandated) LDP as a protocol for hop-by-hop label assignment. RSVP and CR-LDP can also be used.

- In Downstream Unsolicited label distribution, the egress LSR distributes the label to be used to reach a particular host. The trigger for this will usually be new routing information received at the egress node. Additionally, if the label distribution method is Ordered Control, each upstream LSR distributes a label further upstream. This effectively builds a tree of LSPs rooted at each egress LSR. LDP is currently the only protocol suitable for this mode of label distribution.

- Once LSPs have been established across the network, they can be used to support new routes as they become available. As the routing protocols (for example BGP) distribute the new routing information upstream, they can also indicate which label (i.e. which LSP) should be used to reach the destinations to which the route refers.

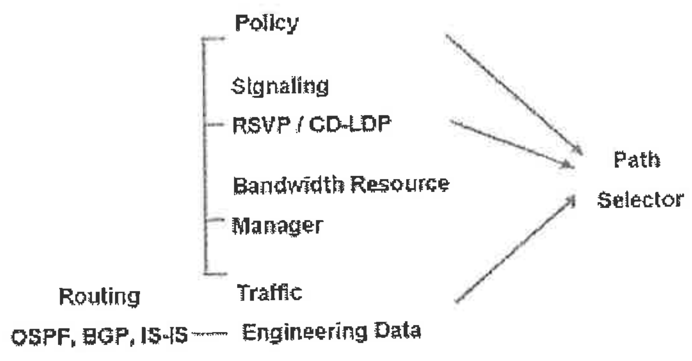


Fig.-2-.MPLS protocols framework for optical WDM Networks

2.6. Comparative study of signaling protocols for GMPLS, RSVP-TE vs. CR-LDP

The choice between RSVP and CR-LDP should be guided by the function of the target system. What LSP setup model will be used?. How stable should the LSP be, do they represent permanent trunks or short-duration calls? How large is the network and how complex is it? Is this a stand-alone network or must the components interwork with other hardware and another networks? A final consideration must be the robustness of the hardware solution. What level of fault tolerance is required? How important is high availability?

Although the two protocols provide a similar level of service, the way they operate is different, and the detailed function they offer is also not consistent. Hardware vendors and network providers need clear information to help them decide which protocol to implement in a Traffic Engineered MPLS network. Each protocol has its champions and detractors , and the specifications are still under development.

Recognizing that the choice of label distribution protocol is crucial for the success of device manufacturers and network designers, we try to explain the similarities and important differences between the two protocols, to help identify which protocol is the right one to use in a particular environment.

Note: all these protocols are still under development, for more information refer to RFC or the IETF drafts.

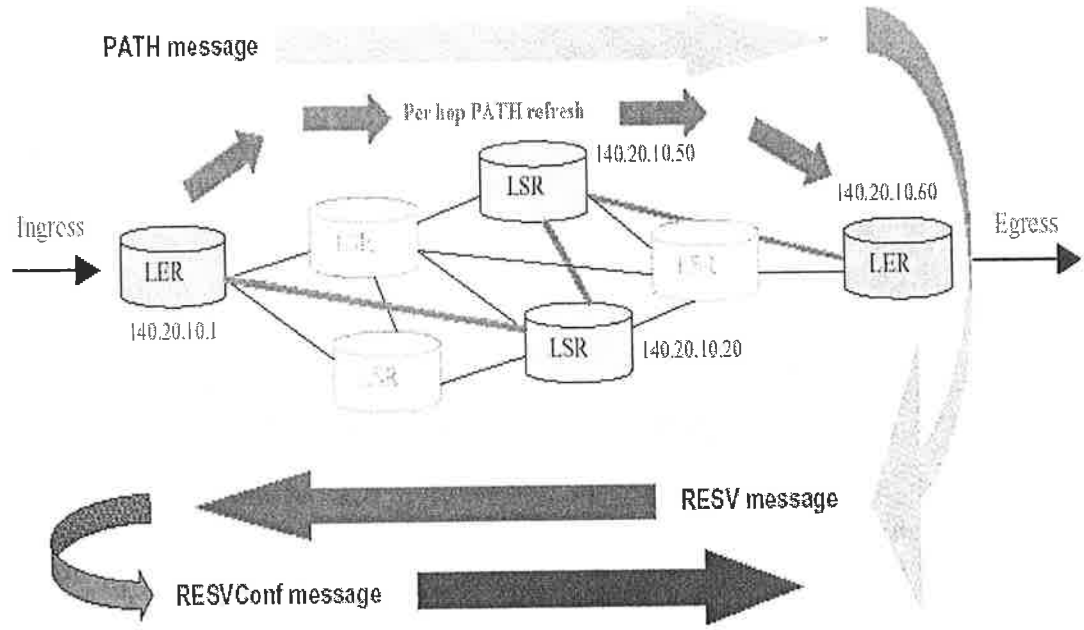
2.6.1. RSVP-TE

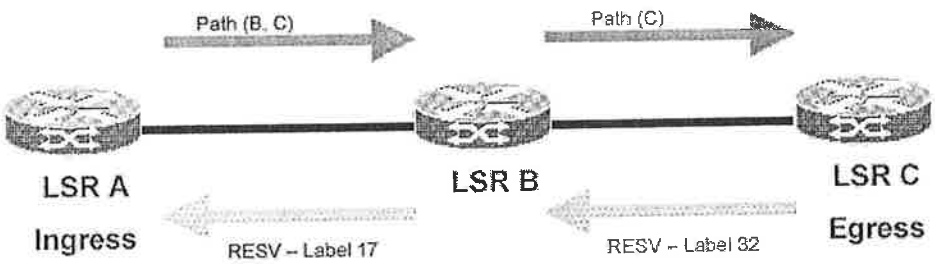
Resource ReSerVation Protocol with extensions of Traffic Engineering.
RSVP defines procedures for signaling QoS requirements and reserving the resources necessary to provide the requested service .

The RSVP-TE extensions enable RSVP to be used for traffic engineering in MPLS environments. The primary extensions add support for assigning MPLS labels and specifying explicit paths as a sequence of loose and strict routes. These extensions are supported by including Label Request and Explicit Route objects in the PATH message. Labels are then subsequently assigned at each node the RESV message traverses.

Other extensions include support for resource affinities, LSP preemption through setup/holding priorities, record route/loop detection options, and rerouting/bandwidth-increase operations.

How it works???





Ingress LER initiates a PATH message to egress LER through each downstream LSR along the path. The traffic parameters required for the session or administrative policies for the network enable LSR A to determine that the route for the new LSP should go through LSR B, which might not be the same as the hop by hop route to LSR C.

LSR B receives the PATH request (message), determines that it is not the egress for this LSP, and forwards the request along the route specified in the request. It modifies the explicit route in the PATH message and passes the message to the next one. Each node receives a PATH message to remember this flow is passing, thus creating a "path-state" or session.

The LSR C determines that it is the egress node for this new LSP, determines from the requested traffic parameters what bandwidth it needs to reserve and allocates the resources required. It selects a label for the new LSP and distributes the label to the LSRB in a RESV message, the egress uses this message to reserve resources with traffic and QoS parameters on each upstream LSR.

Upon receive at the ingress LER a RESV_CONF message is returned to the egress confirming the LSP setup.

After the loose ER-LSP has been established, refresh messages are passing between LER's and LSR's to maintain path and reservation state.

It should be noted that, none of the downstream , upstream or refresh messages between LER and LSR is considered to be reliable, because UDP or raw IP datagrams are used as the communication mechanism. A connection failure will be detected after a neighbor fails to receive a refresh message.

RSVP-TE feature set is robust and provides significant capabilities to provide traffic-engineering services :

- QoS and Traffic parameters- are passed as opaque data to traffic management.
- Failure Notification- upon failure to establish an LSP or loss of an existing one will send failure message, it but relies on timers for refresh message.
- Failure Recovery- "make before break" when rerouting.
- Loop detection- required for loosely routed LSPs only, also supported for re-pathing.
- Multi-protocol support- supports any type of protocol.
- Management- Lsp ID identifies each LSP, thereby allowing ease of management to discrete LSPs.
- Record Route objects- provide the ability to describe the actual setup path to interested parties.
- Path Preemption- the ability to " bump" or discontinue an existing path so that a higher priority tunnel may be established.

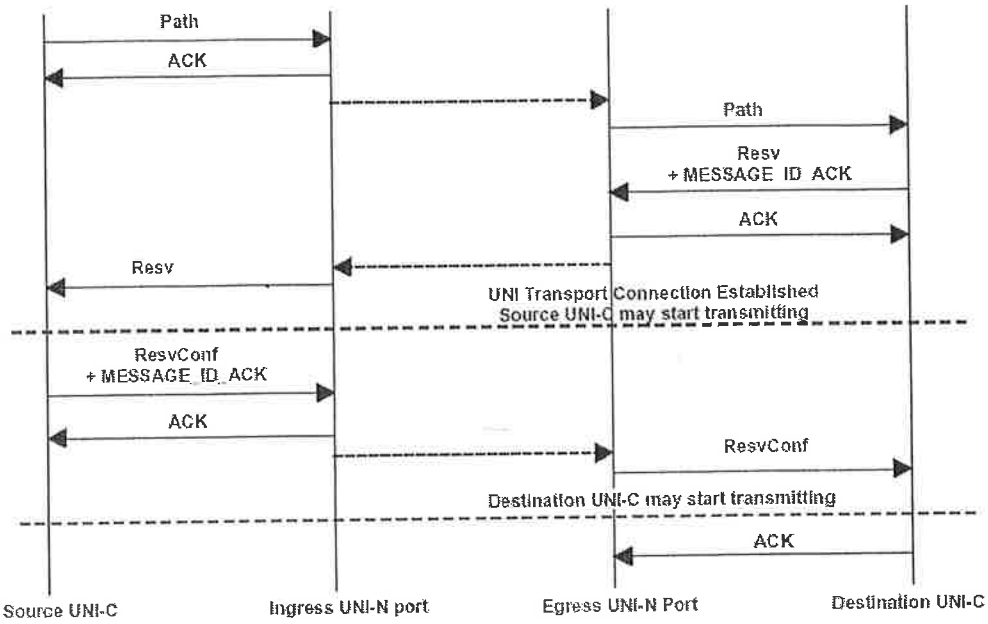


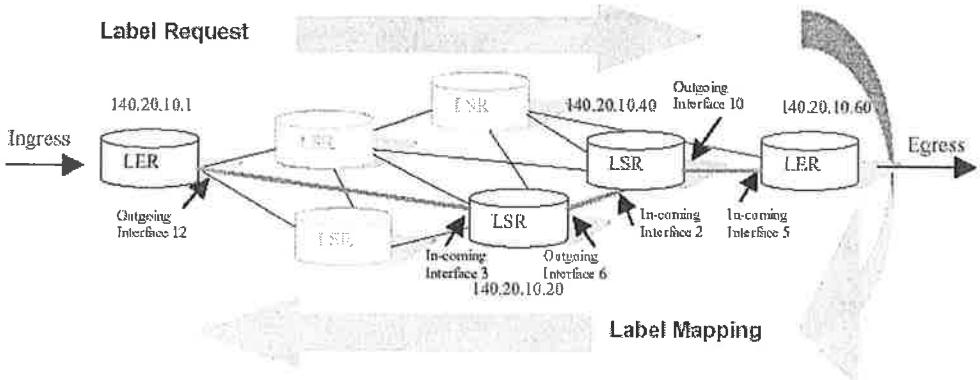
Fig.-3- .A successful connection establishment. Datagram.

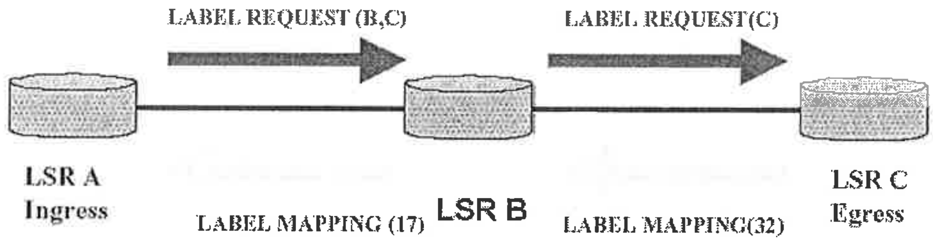
2.6.2. CR-LDP

CR-LDP, constraint based-routed with label distribution protocol, is a simple, scalable, open, non-proprietary, traffic engineering signaling protocol for MPLS networks.

CR-LDP provides mechanisms for establishing explicit routed Label Switched Paths (LSPs) in an MPLS network. These mechanisms are defined as extensions to LDP. Using CR-LDP, resources can also be reserved along a path to guarantee or control QoS for traffic carried on the CR-LSP.

How it works???





The ingress LSR, LSR A, determines that it needs to set up anew LSR to LSR C. The traffic parameters required for the session or administrative policies for the network enable LSR A to determine that the route for the new LSP should go through LSR B, which might not be the same as the hop-by –hop route to LSR C. LSR A builds a LABEL_REQUEST message whit an route of (B,C) and details of the traffic parameters requested for the new route. LSR A reserves the resources it needs for the new LSP, and then forwards the LABEL_REQUEST to LSR B on the TCP session.

LSR B receives the LABEL_REQUEST message, determines that it is not the egress for this LSP, and forwards the request message along the route specified in the message. It reserves the resource for the new LSP, modifies the explicit route in the LABEL_REQUEST message, and passes to LSR C. If necessary, LSR B may reduce the reservation it makes for the new LSP if the appropriate parameters were marked as negotiable in the LABEL_REQUEST.

LSR C determines that it is the egress for this new LSP. It performs any final negotiation on the resource, and makes the reservation for the LSO. It allocates a label to the new LSP and distributes the label to new LSR B in a LABEL_MAPPING message, which contains details of the final traffic parameters reserved for the LSP.

LSR B receives the LABEL_MAPPING and matches it to the original request using the LSP ID contained in both the LABEL_REQUEST and LABEL_MAPPING messages. It finalizes the reservation, allocates a label for the LSP, sets up the forwarding table entry, and passes the new label to LSR A in a LABEL_MAPPING.

The processing at LSR A is similar, but does not have to allocate a label and forward it to an upstream LSR because it is the ingress for the new LSP.

Note: Labels are located from the downstream location, downstream: direction of the data flow, are advertised towards the data source.

Comments:

CR-LDP is built upon LDP, which is already part of MPLS, it is not as mature as RSVP, but it does not require an implementation of an additional protocol and it is only extend if necessary to implement Traffic Engineering.

UDP is used for discovering MPLS peers and TCP is used for control, management, label request and mapping.

If nothing is sent, maintain contact with KeepAlive messages, if it does not arrive (timer) closes the session.

CR-LDP carries the full set of traffic parameters in a Traffic_TLV in the label_request message.

CR-LDP traffic engineering extensions to LDP feature set is comprehensive and is fairly well defined:

- QoS and Taffic Parameters- the ability to define edge rules and per hop behaviors based upon data rates, link bandwidth and weighting given to those parameters.
- Path preemption- the ability to set priorization to allow or not allow preemption by another LSP.

- Path Re-optimization- allows for the capability to re-path loosely routed LSPs based upon traffic pattern changes and includes the option to use route pinning.
- Failure notification- upon a failure to establish an LSP, notification provided on TCP with supporting failure codes.
- Failure recovery- mapping policies to automatic recovery at each device supporting an LSP.
- Multi-Protocol Support- supports any type of protocol.
- Management- LSP ID identifies each LSP, thereby allowing ease of management to discrete LSPs.

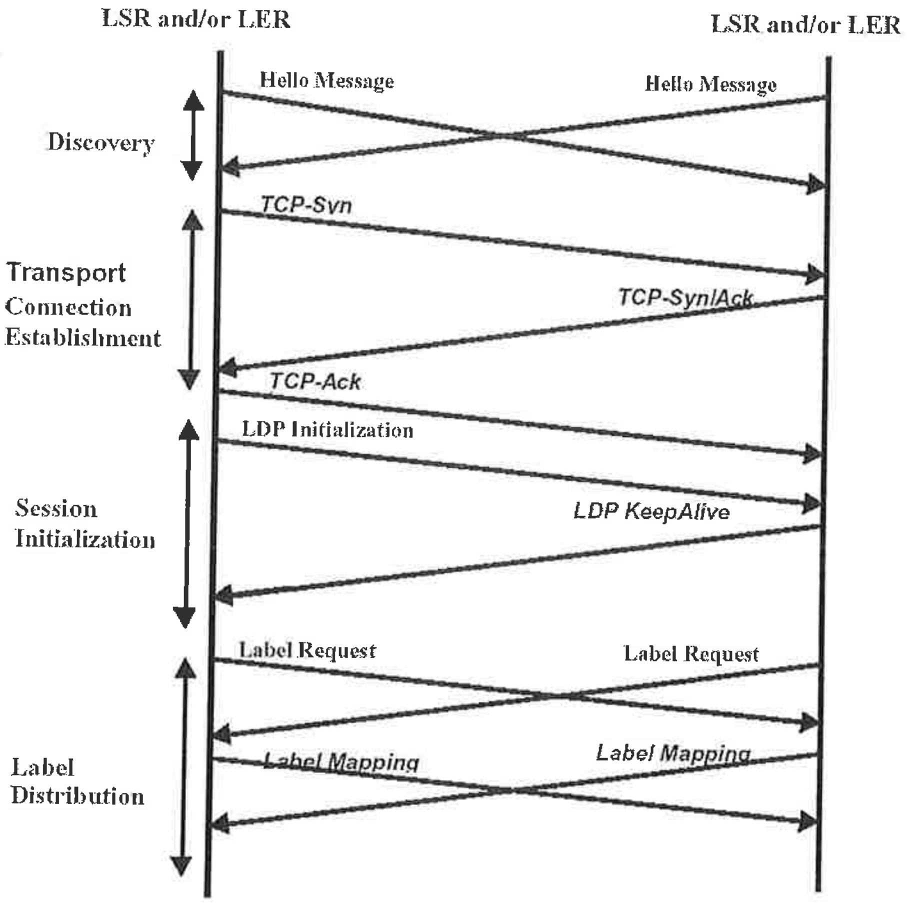


Fig.-4-. LDP initialization

Similarities between TE-RSVP and CR-LDP

Characteristics	CR-LDP	TE-RSVP	Comments
Initiate Setup	Label Request Message	PATH Message containing LABEL_REQUEST object.	
Setup Accomplished	Mapping Message	RESV Message	
Differentiated Services Defined	DIFF-SERV_PSC TLV	DIFFSERV_PSC object	Both contain the DiffServ Code point or DSCP information and included in the setup request message -- see row 1
Support for point to multipoint LSPs	?	?	?
Source Route Capability	Carried in Explicit Route List TLV	Carried in EXPLICIT_ROUTE object	Specify route used to setup switched path.

Differences between TE-RSVP and CR-LDP (adapted from dataplane)

Characteristics	CR-LDP	TE-RSVP	Comments
Development Stage	New	Old with extensions being added, support for legacy networks.	RSVP objects being modified to be used in a MPLS environment
Signaling Transport	UDP for discovery, TCP for sessions	Raw IP datagrams or UDP encapsulation for message	Non-deterministic failure detection with RSVP. TCP failure can have

2.6.3. Comparison of principal features of RSVP-TE and CR-LDP

Reliability:

CR-LDP runs over TCP, while RSVP-TE runs over UDP or IP.
RSVP-TE messages can be lost in transit.
RSVP-TE state refresh allows nodes to recover during fail over events in which local state is lost.
Both use Hello messages, but RSVP-TE Hellos only track whether peer nodes have reset.
CR-LDP Hellos identify each peer's label spaces.

Scalability:

CR-LDP requires less overhead per session than RSVP-TE, because of RSVP-TE's use of soft state.
CR-LDP session maintenance is per-path, not per –LSP.
Intermedite LSRs require less state information with cr-LDP.
RSVP-TE consumes more CPU cycles because of the refresh function.

Interoperability:

Both protocols have undergone interoperability tests (IETF is still working on it).
RSVP-TE has been tested for interoperability with 3com, Bay Networks, Cisco, IBM, and Intel devices.
CR-LDP has been tested for interoperability with Nortel Networks, Ericsson, and GDC.
In September of 1999, ITU-T SG13 decided to use CR-LDP to support IP over ATM using MPLS.
Interworking both are still under development.

Traffic Control:

Both protocols support resource reservation, but RSVP-TE reserves resources on *Resv* arrival while CR-LDP does it on *Label_Request* arrival.
Both support path preemption and have the same setup and holding priority metrics.

2.6.4. Conclusion for protocols

Both CR–LDP and TE–RSVP provide very similar functionality for establishing traffic-engineered, labeled switched paths. Each has its strengths and weaknesses. While LDP is the younger of the two protocols, RSVP has been previously deployed and has operational experience. It is true that there have been extensive enhancements to RSVP in order to support the needs of MPLS. As both CR–LDP and TE–RSVP evolve they will offer more and more similar functionality.

Eventually, MPLS traffic engineering should evolve into a single entity that combines the best-of-breed attributes from both TE–RSVP and CR–LDP. In the meantime, any MPLS implementation by original equipment manufacturers (OEMs) developing LER or LSR platforms should consider supporting both TE–RSVP and CR–LDP to ensure interoperability.

It should be noted that the intrinsic value of having an entry point to providing policy-based management to the core is extremely compelling.

2.7. Multiprotocol Lambda Switching (MPλS)

The framework for internetworking optical networks and MPLS is called MPλS. Both technologies have control mechanisms (a control plane) to manage the traffic. These control planes are shown in figure (5).

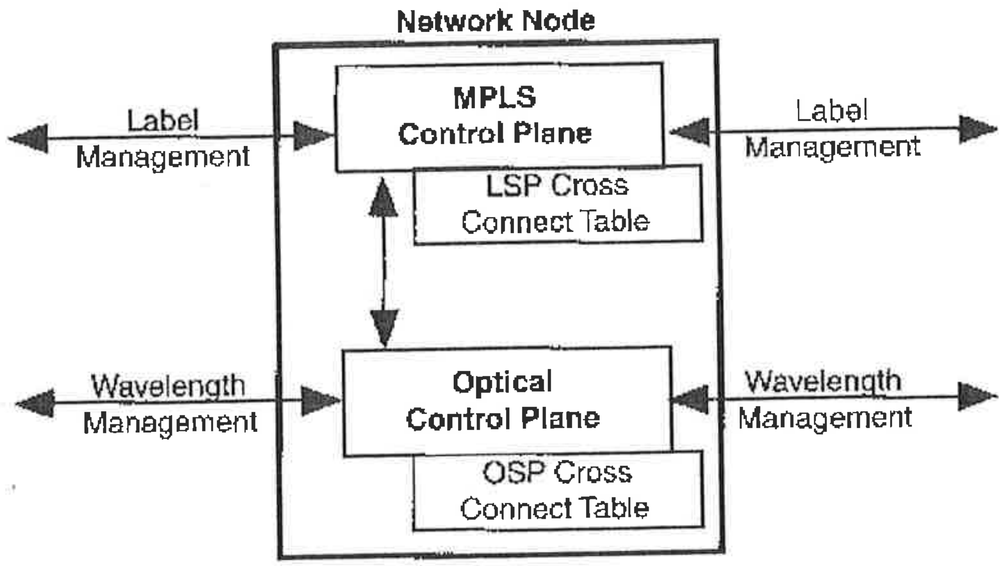


Fig.-5-.The MPLS and optical control planes. Adapted from figure references [2].

The MPLS control plane is concerned with label distribution and binding and end-to-end LSP. The optical control plane is concerned with setting up wavelengths, optical coding schemes, transfer rates, and protection switching options between adjacent nodes.

Some authors [6] disagree with the idea of having different control planes when two technologies must interwork. For example, they propose adapting MPLS TE control plane for OXC. But, due to the arrows between the planes interworking is desirable. We will discuss later about this issue.

2.7.1. OXC and MPLS

An easy way to view the relationship between MPLS and optical networking is to look at the layered model. The optical operations occur in layer 1, and MPLS occur between 2 and 3 (maybe is better to say in a combination of both even that MPLS is a layer 3 protocol).

The data plane of an LSR uses the label swapping to transfer a labeled packet from an input port. The data plane of an OXC uses a switching matrix to connect an optical channel trail from an input port to an output port. An LSR performs label switching by first establishing a relation between an input port and an input label, and output port and output label. Likewise, OXC provisions optical channels establishing relationships between input ports and input wavelengths/channels, and output ports and output wavelengths/channels.

The functions of the control plane include resource discovery, distributed routing control, and connection management. In LSR, it is used to discover, distribute and maintain state information associated to MPLS network, and manage label switched paths (LSPs). In OXC, is used for the same but relative to the OTN, and to establish and maintain optical channel trails under some traffic engineering rules and policies.

A difference between LSRs and OXCs is that, with LSRs, the forwarding information is carried as a part of the labels appended to data packets, and with OXCs, the switching information is implied from the wavelength or optical channel.

How to correlate an MPLS label value with an optical wavelength? at the ingress node the label has been correlated to an appropriate wavelength, that is, an appropriate channel into the network, the transit nodes have been configured to process the wavelength to make the routing decisions. Thus, it is not necessary to know about the MPLS label, as long as all nodes know the relationship of the wavelength that is associated with the label, and its final destination.

We will discuss later in more detail the architecture of an IP and MPLS-based Optical Transport Network.The question is, how the XC is going to be able to:

- Interwork labels with wavelengths.
- How protection switching performed on fibers/wavelengths could be correlated to the labels that are running on the label/wavelength.

The key to interworking effectively the MPLS and optical planes is to carefully distinguish the interaction between not only the control planes, but also between data and control planes.

2.7.2. Control planes

A control plane is a set of software and/or hardware in a node that is used to control several vital operations of the network, such as bandwidth allocations, route discovery, and error recovery. Obviously, the control plane is important.

The control messages are exchangeg between nodes to perform a wide variety of operations. For optical networks, some of the more important tasks for the control plane include:

- Exchange status messages, such alarms and diagnostics.
- Providing timing messages to keep nodes' clocks in synchronization with each other.
- Using messages to download information on which wavelengths will be used between two nodes.
- Building forwarding cross-connect tables to allow the data plane to relay traffic from input port to output port.

The approach for IP/WDM transport networks is to define a separate, dedicated control plane that can operate in any of the following fashions:

- The control plane messages can be exchanged on a separate physical fiber link from those of the user traffic.
- Alternatively, the control messages can be sent on the same fiber link used for the user traffic, as well as on the same wavelength (STOLAS approach).
- Control messages can also be sent on a separate wavelength on the same fiber that is transporting user traffic on the other wavelengths of that fiber.
- Control messages can be sent and received on separate nodes from those that carry the data traffic (complex).

2.7.3. The Optical Control and Data Planes

Figure (6) shows the optical and data planes for IP/WDM transport network. The control plane can be executed with GMPLS or LMP or a combination of both. Whatever the implementation may be, the optical control plane is used to coordinate the use of wavelengths between adjacent optical nodes, as well to insure the nodes are up and running.

Thereafter, the control plane is invoked only for ongoing management operations, diagnostics, recovery, and so on.

The requirements for the optical control plane have established that this control plane must be able to support the following types of connections:

- A permanent optical channel set up by the network management system via network management protocols.
- A soft permanent optical channel set up by the network management system, using network-generated signaling and routing protocols to establish connections.
- A switched optical channel, which can be set up by the customer on demand using signaling and routing protocols.

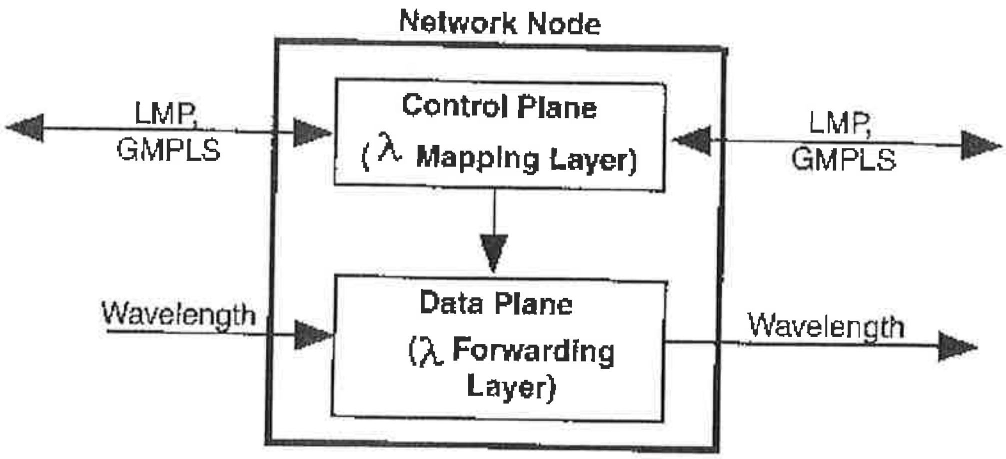


Fig.-6-.The Optical Control and Data Plane

2.7.4. Interworking the three control planes

The separate operations of the IP, MPLS, and Optical Control planes should be coordinated in order to take advantage of:

- The route discovery capabilities of the IP control plane.
- The traffic engineering capabilities of the MPLS control plane.
- The forwarding (switching) speed of the optical data plane.

Figure (7) illustrates how this interworking can be accomplished.

So, the following three events must take place to exploit the powerful capabilities of the three control planes:

- The IP routing protocols advertise and discover addresses as well as the routes to the nodes that are identified by the addresses.
- The MPLS label distribution protocols distribute labels associated with the IP addresses, maps addresses to certain labels.
- The MPLS labels can be mapped to specific wavelengths between adjacent optical nodes.

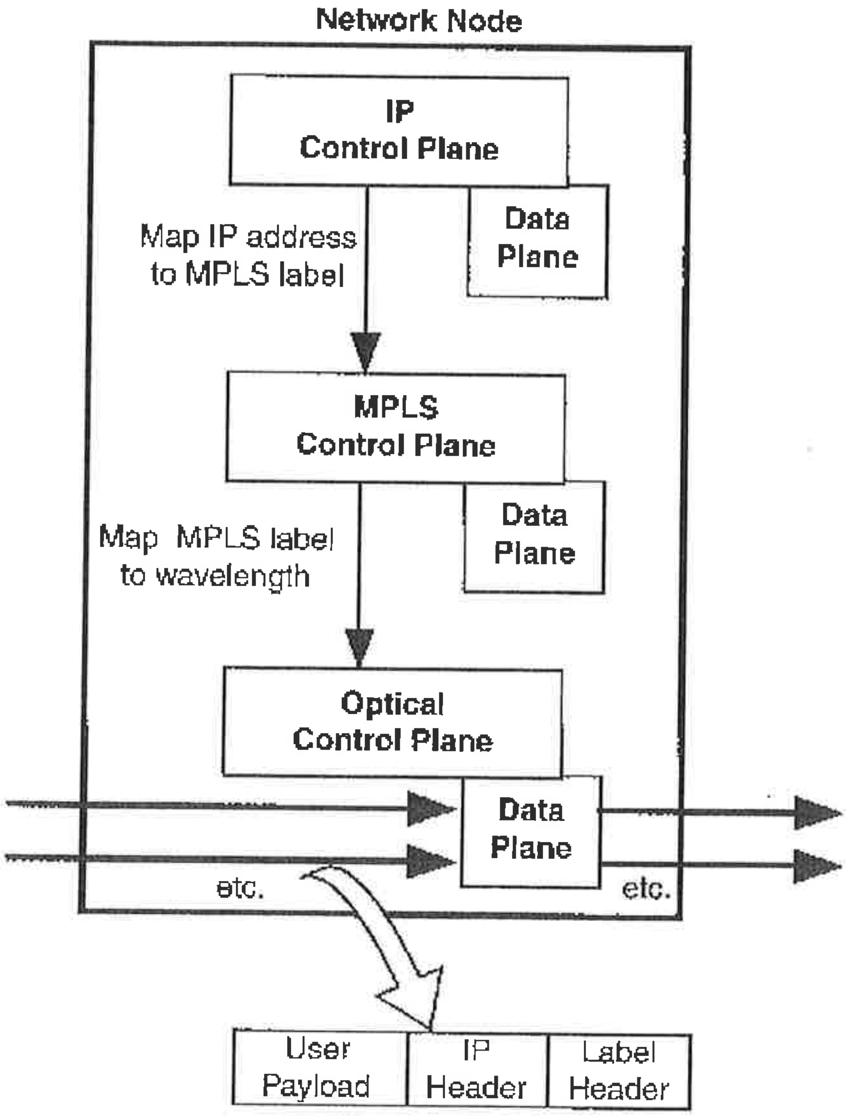


Fig.-7-.Interworking the three control planes. Adapted from figure references [2].

2.7.5. Interconnections for IP over Optical

Given that IP/MPLS over optical can use the Domain or Unified models, the transport of the IP datagrams over an optical network can occur through three kinds of interconnections:

- Peer: Under the peer model, the IP/MPLS layers act as peers of the optical transport network, so that a single control plane runs over both the IP/MPLS and the optical domains. When there is a single optical network involved, presumably a common routing protocol such OSPF or IS-IS, with appropriate extensions, can be used to distribute topology information over the integrated network.

- Overlay: Under the Overlay model, the IP/MPLS routing, topology distribution, and signaling protocols are independent of the routing, topology distribution and signaling protocols at the optical layer, those are defined for the optical domain. Interactions between routing and signaling are accomplished through UNI-defined procedures.

- Augmented: Under augmented model, there are actually separate routing instances in the IP and optical domain, but information from one routing instance is passed through the other routing instance.

2.8. Link Manager Protocol (LMP)

Future networks will consist of photonic switches (PXCs), optical crossconnects (OXCs), routers, switches, DWDM systems, and add-drop multiplexors (ADMs) that use the Generalized MPLS (GMPLS) control plane to dynamically provision resources and to provide network survivability using protection and restoration techniques. A pair of nodes (e.g., two PXCs) may be connected by thousands of fibers, and each fiber may be used to transmit multiple wavelengths if DWDM is used. Furthermore, multiple fibers and/or multiple wavelengths may be combined into a single traffic-engineering (TE) link for routing purposes. To enable communication between nodes for routing, signaling, and link management, a control channel must be established between the node pair. Following references [7] specifies a link management protocol (LMP) that runs between neighboring nodes and is used to manage TE links.

LMP can be used for any type of node, enhancing the functionality of traditional DXCs and routers, while enabling PXCs and DWDMs to intelligently interoperate in heterogeneous optical networks.

In GMPLS, the control channel between two adjacent nodes is no longer required to use the same physical medium as the data-bearing links between those nodes. For example, a control channel could use a separate wavelength or fiber links, and vice-versa. Therefore, a clean separation between the fate of the control channel and data-bearing links must be made. Furthermore, new mechanisms must be developed to manage the data-bearing links, both in terms of link provisioning and fault localization.

A data-bearing link may be either a "port" or a "component link" depending on its multiplexing capability; component links are multiplex capable, whereas ports are not multiplex capable. This distinction is important since the management of such links (including, for example, resource allocation, label assignment, and their physical verification) is different based on their multiplexing capability.

If multiple interfaces are grouped together into a single TE link using link bundling [8], then the link resources must be identified using three levels: TE link Id, component interface Id, and timeslot label.

Resource allocation happens at the lowest level (timeslots), but physical connectivity happens at the component link level. If multiple interfaces are once again grouped together into a single TE link, then link bundling [8] is not required and only two levels of identification are required: TE link Id and port Id. Both resource allocation and physical connectivity happen at the lowest level (i.e. port level). LMP is designed to support aggregation of one or more data-bearing links into a TE link (either ports into TE links, or component links into TE links).

2.8.1. LMP Overview

LMP runs between a pair of nodes and includes a core set of functions; two additional tools are defined in some references [7] to extend the functionality of LMP and are optional. The core function set includes control channel management and link property correlation.

Control channel management is used to establish and maintain control channel connectivity between neighboring nodes. This is done using lightweight Hello messages that act as a fast keep-alive mechanism between the nodes. Link property correlation consists of a LinkSummary message exchange that is used to synchronize the link properties (e.g., local/remote Interface ID mappings) between the adjacent nodes.

LMP requires that a pair of nodes have at least one active bi-directional control channel between them. This control channel may be implemented using two uni-directional control channels that are coupled together using the LMP Hello messages. All LMP messages are IP encoded (except, in some cases, the Test Message which may be limited by the transport mechanism for in-band messaging).

In LMP, multiple control channels may be active simultaneously between a pair of nodes. Each control channel must individually negotiate the control channel parameters, and each active control channel must exchange LMP hello packets to maintain LMP connectivity. If a group of control channels share a common node pair and support the same LMP capabilities, then LMP control messages may be transmitted over any of the active control channels of that group without coordination between the local and remote nodes.

LMP also allows secondary (or backup) control channels to be defined. For example, data-bearing may be used as backup control channels provided control channel traffic has preemptive priority over the data traffic on the link. Secondary control channels only become active control channels when the switchover is complete and they inherit the configuration properties of the primary control channel that is being switched over to it.

Two additional tools are defined that extend the functionality of LMP: link connectivity verification and fault management. These tools are particularly useful when the control channel is transmitted out-of-band from the data-bearing links.

Link connectivity verification is used to verify the physical connectivity between the nodes and exchange the Interface Ids, these Ids are used in GMPLS signaling. The procedure uses in-band Test messages that are sent over the data-bearing links and TestStatus messages that are transmitted over the control channel. The fault management scheme uses ChannelActive and ChannelFail message exchanges between a pair of nodes to localize failures in both opaque and transparent networks, independent of the encoding scheme used for the data. As a result, both local span and end-to-end path protection/restoration procedures can be initiated.

The LMP fault management procedure is based on two message exchanges: ChannelActive and ChannelFail. The ChannelActive message is used to indicate that one or more data-bearing channels are now carrying user data. This is particularly useful for detecting unidirectional channel failures in the transparent case.Receipt of a ChannelActive message must be acknowledged with a ChannelActiveAck message.

The ChannelFail message is used to indicate that one or more active data channels or an entire TE link have failed. Receipt of a ChannelFail message must be acknowledged with either a ChannelFailNack or ChannelFailAck message, depending on if the channel failure is clear or not in the adjacent node.

2.8.2. Control channel management

To initiate an LMP session between two nodes, a bi-directional control channel must be established. The control channel can be used to exchange MPLS control-plane information such as link provisioning and fault isolation information (implemented using a messaging protocol such as LMP), path management and label distribution information (implemented using a signaling protocol such as RSVP-TE or CR-LDP), and network topology and state distribution information (implemented using traffic engineering extensions of protocols such as OSPF and IS-IS). For the purposes of LMP, we do not specify the exact implementation of the control channel; it could be, for example, a separate wavelength or fiber. Our control channel is defined in following chapters.

Furthermore, the control channel messages should be defined to be IP encoded. This allows the control channel implementation to encompass both in-band and out-of-band mechanisms; including the case where the control channel messages are transmitted separately from the associated data link(s).

Note that for in-band signaling, a control channel could be allocated to a data-bearing link; however, this is not true when the control channel is transmitted separately from the data links.

The control channels may also be used for transmitting and receiving signaling and routing messages. Each LMP control channel must individually negotiate the control channel parameters, and each active control channel must exchange LMP Hello packets to maintain LMP connectivity. For LMP, it is essential that at least one control channel is always available.

CHAPTER 3:

Network Management for the STOLAS project: System Analysis and Architecture.

3. NETWORK MANAGEMENT FOR THE STOLAS PROJECT: SYSTEM ANALISYS AND ARCHITECTURE

Note: This chapter is an adaptation of the WINMAN project to the STOLAS project, any information included about WINMAN is EXTREMALY CONFIDENTIAL, request WINMAN Consortium to use any information.

3.1. Introduction

The management functions specifically needed for optical-label-switched networks will be defined. These include the elements management of the advanced optical network elements (i.c. the label swapping wavelength converter and phase modulator circuits, the fast widely tunable laser diodes, the 2R regenerators..), the optical signal monitoring, the interaction with the electrical IP routers, the control of the optical routing, and the control of the labeling functions. Also the requirements on a network management and control communication channel will be identified.

MPLS-controlled optical label switching (OLS) related management requirements are related to both element management and network management. The activities will focus on:

- The definition of the management architecture (logical and physical) for the Element Management System (EMS) and the Network Management System (NMS).
- The definition of the configuration, fault and performance requirements for EMS and NMS.
- The definition of the functional architecture and atomic functions necessary for EMS and NMS.
- Identification of the methodology useful to describe the information model for managing the network element through the EMS, if necessary.
- Description of the basic blocks of the OLS information models useful to satisfy the requirements defined for each functional management area.

The development of a network management system is not a project objective, within STOLAS it is mainly hardware-oriented and focuses on a new optical labeling concept. Nevertheless, to be able to deliver a proof-of-concept and the mapping of electrical labels on the optical stack in a field trial, a limited network management system is required, taking into consideration that without configuration management is difficult to set up paths between the commercial IP routers.

The main objective of the WINMAN-over-STOLAS solution is to provide an integrated network management system, which is capable of providing managed end-to-end IP connectivity services derived from Service Level Agreements (SLA's). In other words, WINMAN will test in field a Network Management System, capable of performing integrated provisioning of IP/MPLS Label Switched Paths over optical paths, as well as integrated multi-layer fault and performance management. WINMAN will capture the requirements, define and specify and open, distributed, and scalable management architecture. Will support multi-vendors, multi-technology environments and evolution scenarios. Therefore, it is a good manager for STOLAS project, moreover WINMAN is a project from IST in which Lucent is involved.

Note that the term domain in this paper, does not refer to administrative domains, but to technology domains , like IP and WDM.

3.2. The Control Plane vs. Management Plane approaches in the IP/WDM Integration

The term control plane is used in the literature to refer to the set of real-time mechanisms and algorithms needed for call or connection control. It deals mainly with the signaling to set-up, supervise and release calls and connections [9]. Although, a detailed de-composition of the control plane and a description of each component is not our purpose, we can safely assume that the signaling protocol for connection set-up and the routing protocols supporting network discovery are the most significant features of the control plane. In that respect, it is significantly easier to follow all the recent advances and proposals about the integration of the IP-electrical world and the WDM-optical world control planes.

Many standardization bodies as well as international for a have addressed the issue of integrating the control lane of the IP MPLS-capable and the WDM network elements (NE). The IETF has proposed the MPLambdaS framework [6], which extends the MPLS ideas to the optical domain, allowing the re-usability of the existing Internet protocols with the appropriate extensions. The OSPF as well as the IS-IS routing protocols have been enhanced to disseminate information relevant to the optical domain ([10], [11]). On the other hand, the OIF and he ODSI forums have made one step towards the definition of the appropriate signaling messages ([12], [13]), which will allow the dynamic set-up of end-to-end connections between IP routers spanning the optical network. Furthermore, [14] and [15] present the mapping between the signaling messages defined in [12] and existing IP/WDM signaling protocols, namely RSVP-TE and CR-LDP.

The latter is an important issue, since the automatic provisioning of end-to-end connection between two IP routers by means of signaling permits to Network Operators

(NO) and Network Service Providers (NSPs) to provide rapidly the requested IP connectivity to customers. This is the main requirement of the NO and NSP in order to cope with the changes in the service-chain and to be able to provide in a cost efficient way the requested services.

On the other hand, the term management plane is used in the literature to refer to the set of near real-time management mechanisms and algorithms related to the system as a whole and to the OAM [9]. It deals mainly with the procedures related to five functional areas, namely Configuration, Fault, Accounting, Performance and Security (FCAPS). The three main management functions of the FCAPS which have similar functionality with the control plane functions, thus competing each other, are path provisioning with routing and QoS support, in the Configuration Management Area, and automatic recovery of failures or performance degradations in the Fault and Performance Management Areas. Network Management functionality mainly exists independently for the IP-electrical world and the WDM-Optical world rather than for the integration of the two worlds. In this direction we will first deal with the advances and proposals for IP layer and WDM layer separately and then we will provide the limited efforts on their integration.

3.2.1. The Control & Management approach (best of two worlds)

It seems clear that comparing the two previous approaches we see that one of them is clearly better in the first phase and the other one is clearly better in the second phase.

What we try to do in this third approach is to define a mixed solution that combines the best of the previous approaches.

The design phase is performed by the management system. The management system has a clear picture of the network and it has access to information about network topology changes faster than any other network element. Because it is the management system that makes the design of the IP paths, it is the management system that decides what resources should be reserved and when. So the management system knows where and what resources are reserved on the network. The information about resources is always up to date on the management system.

By making the design in the management system we maintain the advantages of freeing the edge routers from this high CPU consuming process and having the best-informed entity making the design of the IP paths.

The implementation phase is performed by the control plane, but triggered by the management plane. After the management system has designed the path it triggers the signaling protocol.

The signalling protocol must support explicit routing and must be able of performing resource reservation. The information transmitted by the signalling protocols is given by the management system when it triggers the signaling.

This way we have paths designed by the management system implemented as fast as the ones designed by the control plane (and implemented much faster than the ones implemented by the management system).

The major problem in this approach is the same that we had in the implementation phase of the control plane approach. All the network elements must use compatible signalling protocol implementation. This is very hard to provide in multi-vendors networks nowadays.

We think that with this approach we are using the best of the previous two. We can have IP paths established by the best-informed entity (management system) and implemented by the fastest mechanism (signaling protocols of the control plane).

3.3. High level architecture

A two-layer management architecture has been conceived for the system (multi-layered network), the first layer contains technology dependent managers, that is one for the IP network and another manager for the WDM optical network. These two management system will interact with existing element management systems or even subnetworks called southbound WINMAN (Fig.1), can be deployed as a stand-alone management system capable of managing a WDM network by its own.

On top of this layer there is a second one, called INMS (Integrated Network Management System), whose purpose is the integration of the above mentioned management systems to make the best use of the underlying transport technologies.

Goals in WINMAN architecture:

Open, flexible, modular, scalable, distributed (transparent), independent, minimization of interfaces.

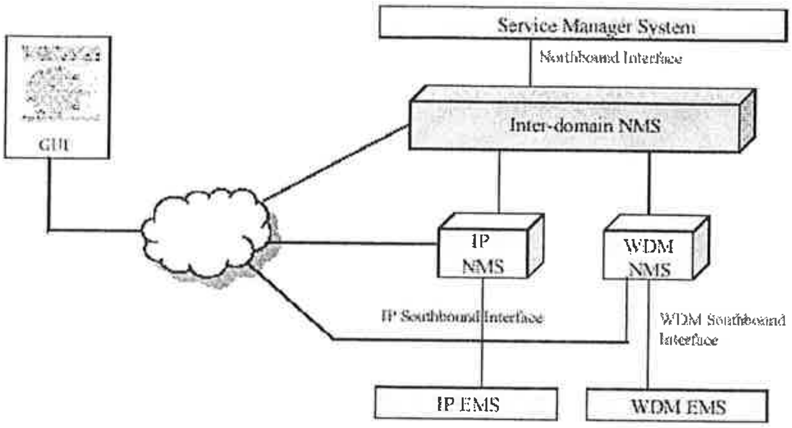


Fig 1. Overview of WINMAN high level architecture.

3.4. Systems description

3.4.1. Integrated Network Management System (IMNS)

Through a properly abstracted information model, the INMS is able to use the domain NMSs for the provisioning and monitoring of end-to-end connections without any knowledge about network details except for the layered structure.

The main functions that the IMNS performs are in the area of configuration management, fault management and performance management (end-to-end).

- Serving of requests from the Service Management Systems.
- Configuration of the end-to-end connections and services, designing the route as an ordered set of subnetwork connections.
- Presentation of an end-to-end view to the operator, resources.
- Maintenance of an inventory of resources.
- Co-ordination of the restoration process and Qos.
- Provides updates.
- Inter domain alarm correlation.
- Analysis of data and assessment of the QoS delivered.

3.4.2. WDM-NMS

The WDM Network Management System carries out the management of the WDM transport network layer .

The main functions are:

- Provision of optical paths, including protection.
- Wavelength routing, transmission and implementation of the optical path in the network elements.
- Discovery of network resources in terms of network elements (ports, fiber connections, wavelengths per fiber, transport capacity per wavelength…).
- Co-ordination.
- Gathering of data about the performance levels (power, BER, Eb/No….).
- Control of alarms.

3.4.3. IP-NMS

The IP policy-based Network Management System carries out the management of the IP network layer , adapted to the connection oriented approach on which WINMAN focuses.

The main functions are:

- Design of connections inside its own subnetworks.
- Establishment and maintenance of end-to-end QoS connectivity services, including the setting up of MPLS LSPs.
- Discovery of network topology.
- Gathering of performance information (link load, network congestion, effective throughput,…).

3.4.4. GUI (Graphic User Interface)

The graphical user interface interacts with the three previously described systems.It supports the access to information and functions of the INMS, WDM-NMS, IP-NMS. Should be as thin as possible in order to be WEB enabled.

3.5. Domain model and Network views

WINMAN will focus on a connection-oriented IP domain model, corresponding to MPLS technology. We identify 2 layers, MPLS and WDM, so we consider only MPLS entities and we replace everything related to IP with MPLS.

We suppose MPLS domain starts at the Customer Edge IP equipment, so all the traffic that enters in WINMAN is marked as MPLS. However, the MPLS domain should be further developed since this domain is not completed established and is still under updating.

3.5.1. Common Entities

- Connectivity Service (CS):
  Is the basic service delivered by the MPLS or WDM network and managed by the Management system. The service is delivered either by the MPLS network meeting some QoS goals resembling a connection-oriented circuit or by the WDM network.

- Element Management System (EMS):
  The EMS represents the abstraction of the sub-network(s) managed by the EMS managed domain and the element management system itself.

  - A termination point (TP) shall be a logical abstraction of an end –point (actual or potential) of a topological (physical) link, or a subnetwork connection. A TP is contained within a managed element.

- Physical Termination Point (PTP):
  A termination point that is an actual or potential endpoint of a topological (physical) link shall be abstracted as a physical termination point (PTP). Essentially, is the representation of a physical port.

- Connection termination Point;
  Actual or potential end point of a subnetwork connection.

- Topological Link:
  A Topological Link is a physical link between two PTPs. A topological link has a name and references to the two PTPs. A topological link reported bya n ERMS to NMS will be between two managed elements (MEs) managed by the same EMS.

- Link connection:
  Represents the transparent capacity of transfer information characterized by a given signal identification between two fixed points.

- Managed element:
  Is an abstract class used to represent Network Elements visible across the interfaces.

3.5.2. MPLS entities

Some of the most important are:

- MPLS Connectivity Service Connection (MPLS-CS):
  CS is the basic service delivered by the MPLS
- IP Element Management System (IP-EMS), manages both IP and MPLS protocols.
- MPLS Connection Termination Point (CTP), At the MPLS layer it corresponds to a specific label and port.
- MPLS Cross-Connect table, is a set of associations (FEC).
- MPLS Trail or Label Switched Path, transport entity.
- MPLS Physical Termination Point, representation of a physical port, belonging to a Label Switch Router (LSR).

3.5.3. WDM entities

- Optical Connectivity Service Connection (OCS), provides OCh-SNCs.
- WDM Element Manager System (WDM-EMS).
- Optical Topological Link.
- WDM Physical Termination point (WDM PTP), representation of a physical port, belonging to an Optical Managed Element.
- Optical Connection Point (OCP), in an optical network it is a representation of a specific wavelength of an optical port.
- Optical Managed Element (OME),used to represent network elements, we should include: OADM, OXC, Optical Amplifiers, Terminal Multiplexer.

3.5.4. External and Internal network view

(modules referred to next figure 2.)

(1) Module (INMS):This external interface should be as abstract as possible. The INMS need to show to SMS only the edge Physical Term. Points, which are the ones interconnecting the provider and the customer.
PTPs are always MPLS PTPs and not WDM. CTPs are inside PTPs.

(2) Module (IP NMS): PTP are edge MPLS-PTPs which were also shown towards the SMS and the MPLS over WDM PTPs, which interconnect to the WDM network. Between CTP exist MPLS subnet. connections (MPLS – SNC). Note that the MPLS CTP interconnecting to the WDM network are marked with ('), while the corresponding CTPs in WDM will be later marked with (").

3.5.5. Multiple views in a single picture

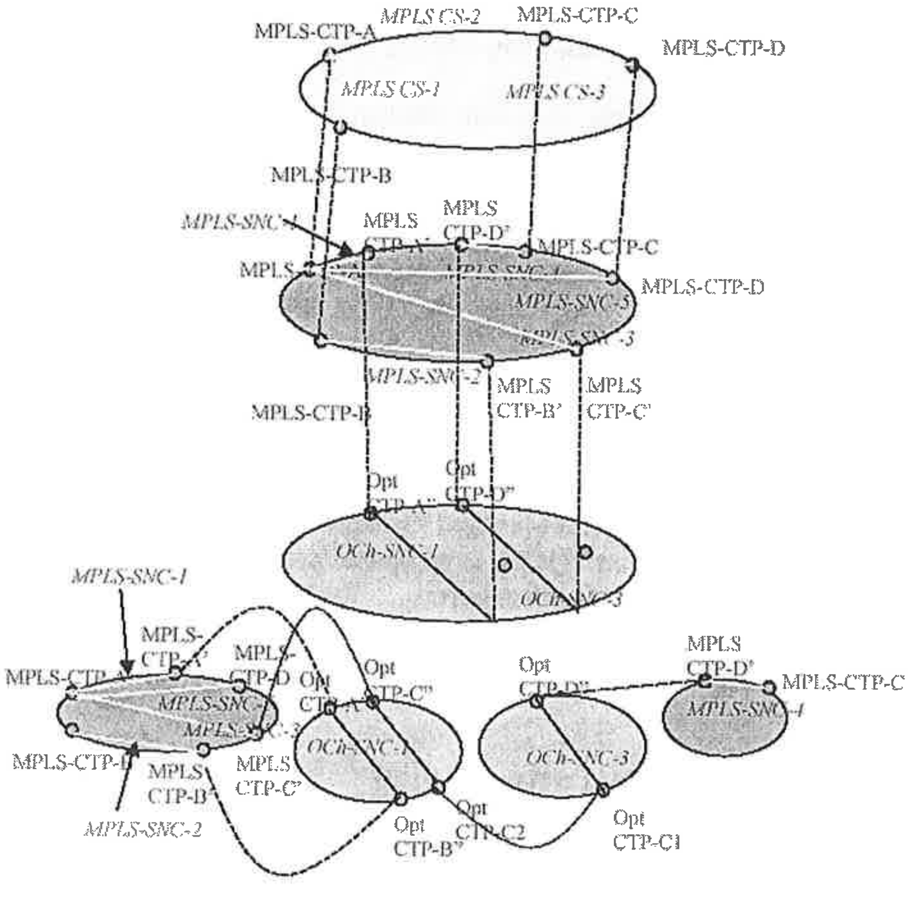


Fig 2. Multiple SMS-INMS, WDM NMS-INMS, and IP NMS-IMNS. Adapted from figure references [4].

The previous picture depicts the multiple views between SMS-INMS, WDM NMS-INMS, IP NMS-INMS and IP EMS-IP NMS, WDM EMS-WDM NMS in a single picture.

Analysing for example the MPLS-CS –1, it is composed by:

MPLS CS-1= MPLS-SNC-1+LC+Och-SNC-1+LC+MPLS-SNC-2

- The MPLS-SNC1 between MPLS CTP-A and CTP-A'
- The optical link connection between MPLS CTP-A' and optical CTP-A"
- The Och-SNC1 between MPLS CTP-A" and optical CTP-B"

- The optical link connection between optical-CTP-B" and MPLS-CTP-B' and finally
- The MPLS-SNC2 between the MPLS-CTB' and MPLS-CTP-B.

We can see also:

MPLS CS-2= MPLS-SNC-3+LC+Och-SNC-#+LC+MPLS-SNC-4
MPLS CS-3= MPLS-SNC-5 (pure MPLS).

In order to interconnect with a real example and test-bed configuration, next figure (3) shows the Element Management Layer of the MPLS and Optical Sub-networks together with the corresponding technology equipment.

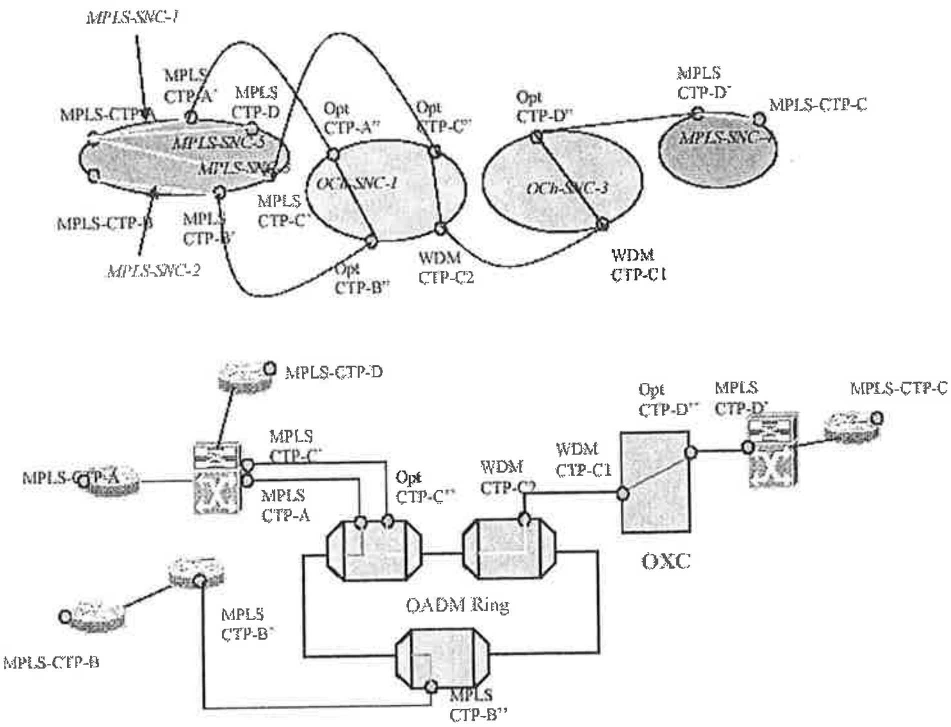


Fig. 3. Element Management Layer of the MPLS and Optical Sub-networks together with the corresponding technology equipment. Adapted from figure references [4].

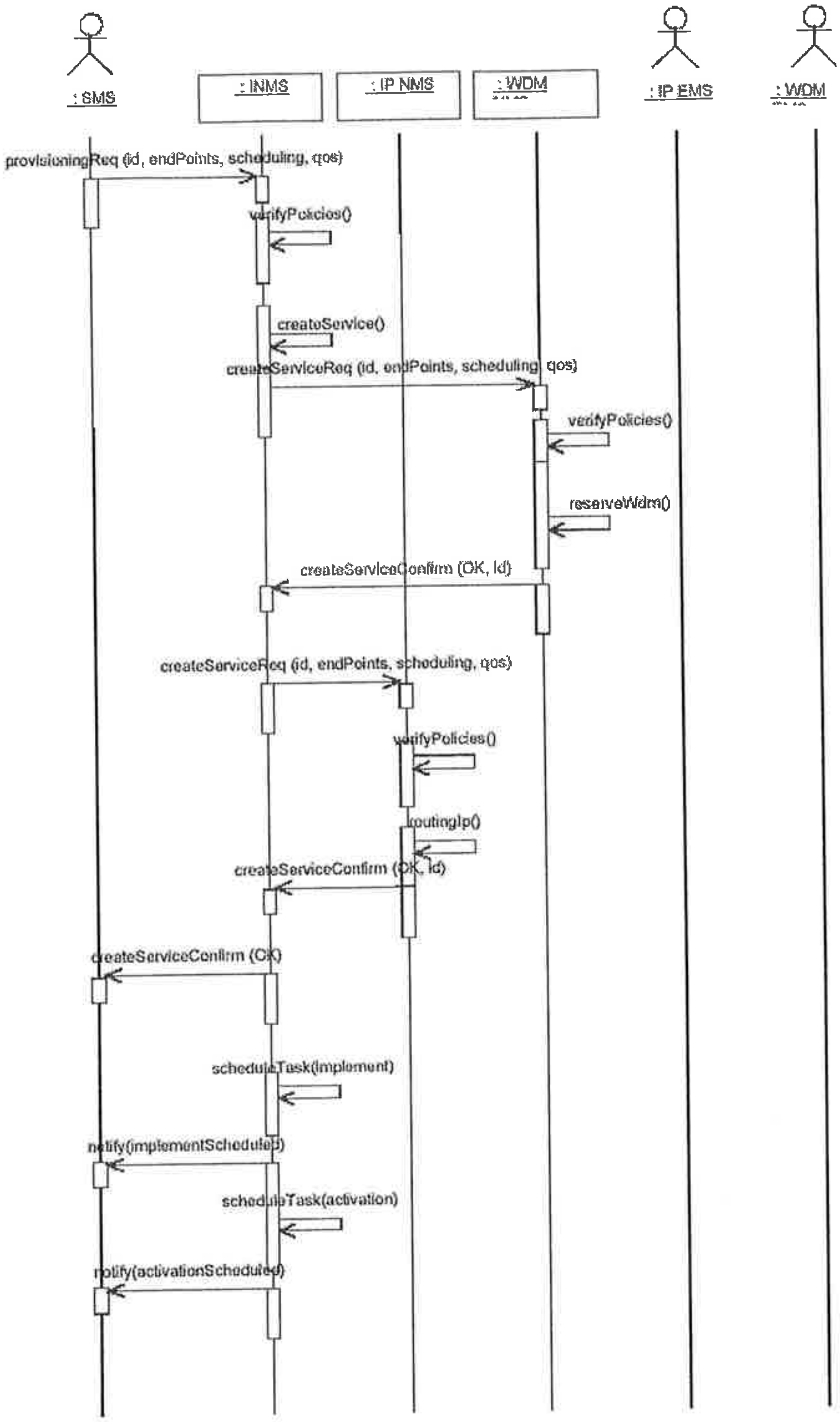
The network that is being displayed is composed by the 2 Optical Sub-networks and 2 MPLS sub-networks. The first optical sub-network is a ring of Optical Add-Drop-Multiplexers and the second one is an Optical Cross-Connect. In the first case, OChSNC are a concatenation of optical link connections (inside fibers) and cross-connects inside the OADM, while in the second case the OChSNC is generated only in the cross-connect inside the OXC. On the other hand the MPLS network is composed by multiple Label Switch Routers, either playing the role of ingress, core or egress LSRs.

Note that we have selected 3 MPLS CS connections, 2 of them spanning both domains, while 1 of them is a pure MPLS one (between MPLS-CTP-A and MPLS-CTP-D).

3.6. System scenarios

In this section is intended to show the responsibilities of each system and the main interactions among them.

One of these is, for example , provision of ICS involving creation of new lightpath. Corresponds with the request received from the Service Management System for the provisioning of an IP connectivity service with NLA parameters and scheduling data. Includes the following steps:


: SMS
: INMS
: IP NMS
: WDM
: IP EMS
: WDM
provisioningReq (id, endPoints, scheduling, qos)
verifyPolicies()
createService()
createServiceReq (id, endPoints, scheduling, qos)
verifyPolicies()
reserveWdm()
createServiceConfirm (OK, Id)
createServiceReq (id, endPoints, scheduling, qos)
verifyPolicies()
routingIp()
createServiceConfirm (OK, Id)
createServiceConfirm (OK)
scheduleTask(implement)
notify(implementScheduled)
scheduleTask(activation)
notify(activationScheduled)

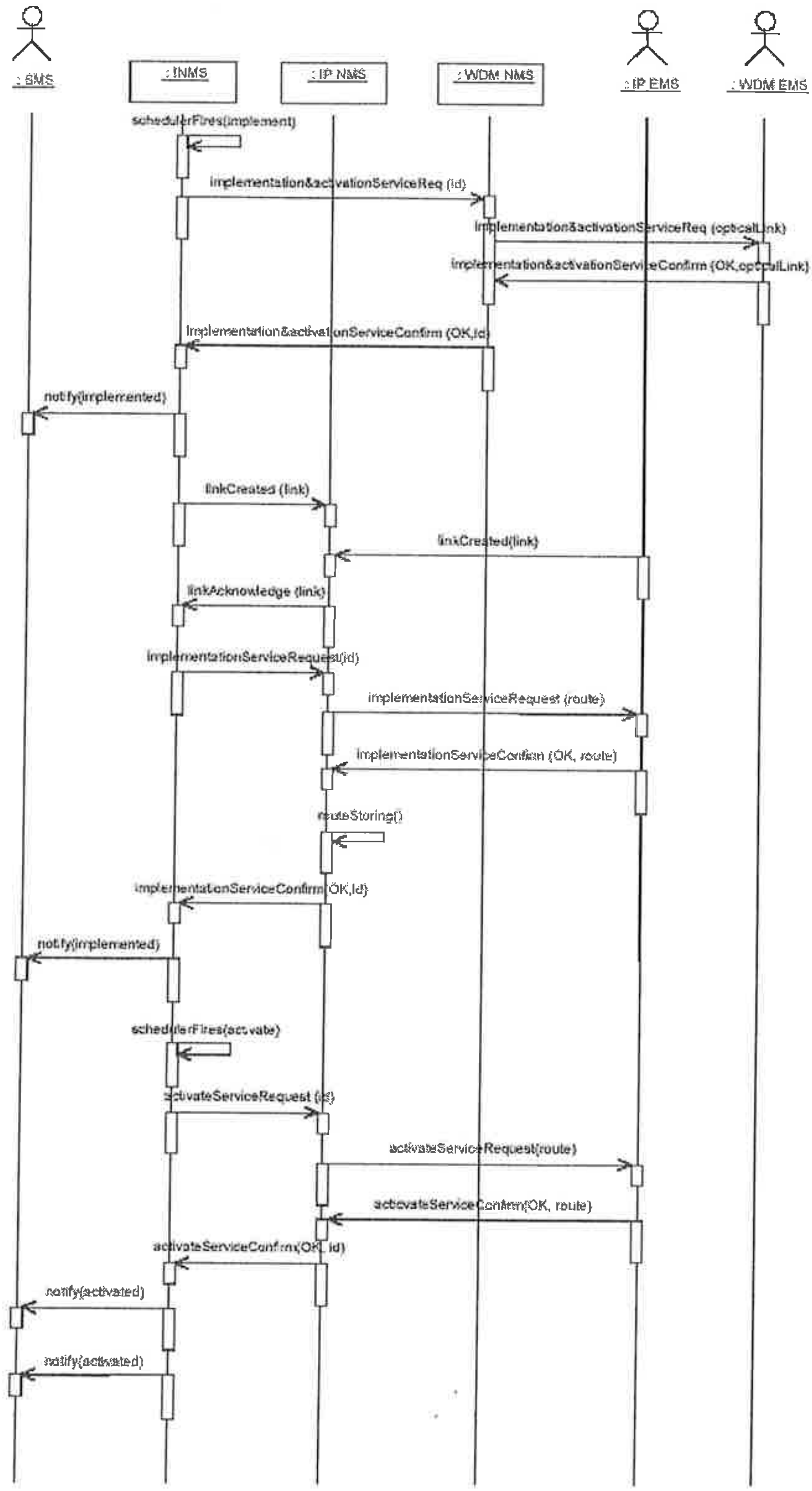


Fig.4. Provision of ICS involving creation of new lighpath. (WINMAN)

3.6.1. Description

- SMS sends a provisioning request to the INMS, specifying ICS identifier and new parameters for the service.
- INMS verifies the request and determines the validity of the requested service by applying policy rules. In the current step only INMS policy rules are verified.
- INMS makes the design of the network connections at INMS level, cause this system have enough info for that sort of decisions.
- Once INMS have made the design, it sends a request to the WDM NMS to design the WDM path in the WDM domain.
- WDM NMS proceeds to verify WDM policy rules and then to design and reserve the WDM path.
- WDM NMS sends response to INMS pointing success.
- INMS sends a request to the IP NMS to design the IP route.
- IP NMS verifies the request against policies and proceeds to the creation of the IP route.
- IP NMS sends a service created confirmation to the INMS.
- The implementation task is scheduled.
- A notification is sent after the schedule.
- The activation task is scheduled.
- A notification is sent after the schedule.
- At the scheduling time, INMS send a request to the WDM for the implementation and activation of the designed optical path.
- WDM NMS forwards that request to the correspondent EMSs, indicating the optical link to be implemented and activated.
- The EMSs responds to the WDM NMS indicating success in the operation.
- The WDM NMS forwards that confirmation to the INMS.
- INMS notifies to the IP NMS that a new link has been created, thus the IP topology has changed.
- The creation of a new link in the optical domain can be detected by the network elements, and notified upwards. In that case the EMS sends a notification to the WDM NMS informing about the new link created. Only the first notification of the new link is processed by the IP-NMS. If this notification from the EMS were not received, the IP NMS would send to the appropriate EMSs a new link notification.
- IP NMS sends notification to the INMS indicating that IP NMS has been informed about the new link.
- INMS sends an implementation request to the IP NMS, indicating the service identifier.
- IP NMS forwards the request to the appropriate EMSs.
- EMS sends a response to the IP NMS.
- IP NMS stores the route in the database.
- IP NMS sends a confirmation to the INMS about the service implementation.
- INMS notifies to the SMS that the service has been implemented.

- At the scheduled time the INMS sends to IP NMS an order to activate the given connection.
- IP NMS sends to EMS the order to activate the new route and receives its confirmation.
- IP NMS notifies the activation confirmation to the INMS (for WDM).
- IP NMS notifies the activation confirmation to the INMS (for IP).

3.7. STOLAS Network Management

The STOLAS Management System follows to a certain extend the outline of the general management architecture.

The OLS-EMS has knowledge about the OLS-NEs and is able to create/delete a communication through the OLS network. The OLS-GUI is the point where the operator has access to the OLS management network.

The interfaces between the management system are all CORBA (Common Object Request Broker Architecture) based with a possible exception for the IP-routers (Nes under IP_GUI). The interfaces have to be specified in IDL (interface definition language).

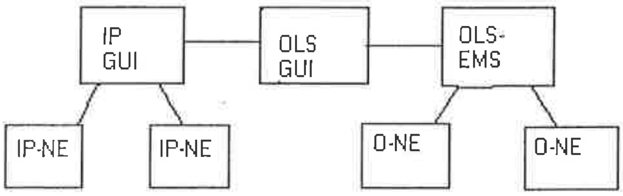


Fig.5. STOLAS network Management System.

3.8. STOLAS Network Elements

3.8.1. Optical Edge Router

The OER (optical edge router)has on his access side data interfaces like GigabitEthernet or SDH (input ports and output ports). On his network side the OER has optical interfaces (input ports and output ports) with the optical label added. For management purpose the edge router has an Ethernet interface. The data ports are PTP (physical termination points).

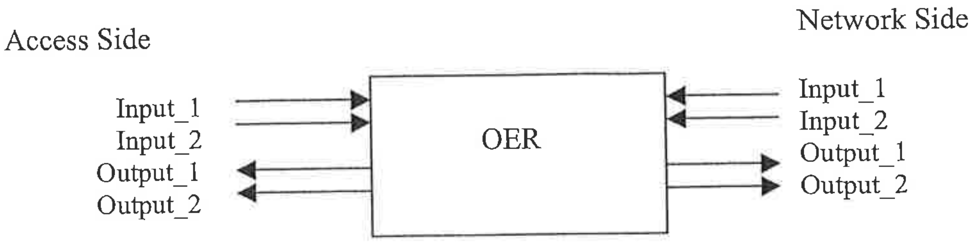


Fig.6 . Data ports of OER

3.8.2. Optical Cross-Connect

The OXC is placed in the optical network. It has inputports, outputports, addports and drpports. There is a possibility of muti-cast from inputports to dropports and/or outputports. For management purpose the OXC has an Ethernet interface. The add/drop ports can be connected to network side ports of the OER.

3.8.3. Optical Add-Drop Multiplexer

The OADM is placed in the optical network. An OADM consist of two modules that have each one inputport, outputport, addport and dropport. There is multi-cast within a module possible. For management purpose the OADM has an Ethernet interface.

Physically OADM-NE and OER-NE are in the same box. The addports/dropports can be connected to the network side of the OER.

3.8.4. STOLAS internal Structure of OLS-EMS

For the testbed the EMS is constituted by a number of classes.

The OLS-EMS will manage a number of network elements (4 in the testbed). The focus is on the configuration management (i.e. establishing and removing subnetwork connections). To achieve this the OLS-EMS should be able to instruct the optical NEs to set up or delete OLS-WDM lighpaths. The configuration of the network and its elements should be represented.

This configuration consists of managed elements, topological links, physical termination points and connection termination points, logically separated in a number of subnetworks. This structure will be presented to the EMS-operator using a GUI.

Fault management of the optical nodes (laser error, fibre cut, etc.), and performance parameters (like power measurements and S/N ratio) should be added and defined into the WDM-EMS. The structure of the WDM-EMS is shown in figure (7).

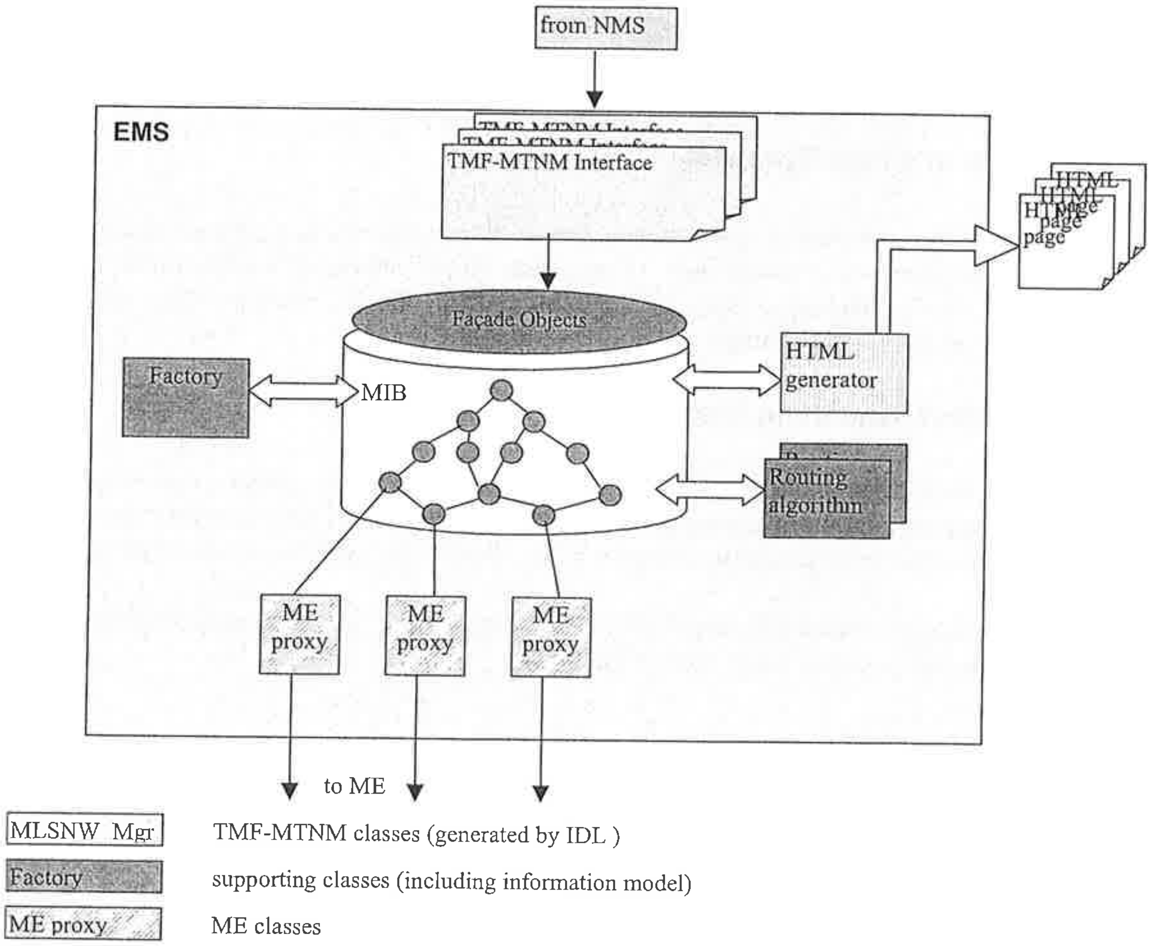


Fig.7 Structure of the OLS-EMS

3.9. Information Model

The basic and dynamic configuration of the subnetwork managed by the EMS is described by the Information Model (IM). This model gives the objects that can be present in the subnetwork, and their relationships. Instances of these objects are contained in the database of the EMS. This database has a static and a dynamic part. In the static part, the physical configuration is represented, like the network elements, their physical ports and the links between those ports. The dynamic configuration contains the connections that are present in the network elements, together with the connection termination points. If a new optical network element is added, the static configuration changes; if a new subnetwork connection is created, the dynamic configuration is altered. The instances of the information model will be contained in a List of Top objects, which is a data member of the Facade objects. These objects are the facade objects of the MIB, through which all of the operations on the MIB are initiated.

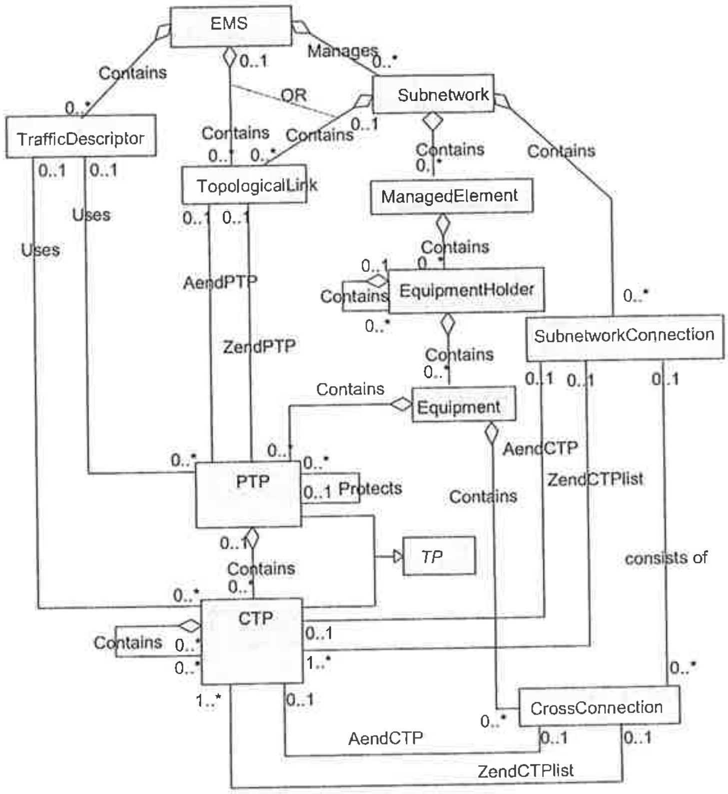


Fig.8 Object classes constituting the OLS-network information model (STOLAS-WINMAN)

3.9.1. TopObjects

Each object in the information model inherits from the Top object. Attributes will be added in the constructor of the object, which also gives the attributes an initial value. With the **set** operation, attributes can be assigned a value. If a set operation is invoked using a name that is not valid (i.e. for which no **addAttribute** operation is invoked), the set operation will fail. With the parentId and the containmentList the object instance tree can be created and searched.

Object Name	TopObjects	
Object Description	TopObjects is the base class for all STOLAS based information model classes	
Super Class	Top	
ParentId	none	
ChildIdList	All managedObject, which represents EMS System	
Attributes	globalDefs::NamingAttributes_T	name
	String	UserLabel
	String	NativeEMSName
	String	Owner
	globalDefs::NVSList_T	additionalInfo
Operations	set(TattributeId aId, const Tval& tval)	

Object Name	EMS	
Object Description	This object represents the aggregation of the networks as managed by the EMS.	
Super Class	TopObjects and emsMgr::EMS_T	
ParentId	none	
ChildIdList	Subnetwork, TopologicalLink, TrafficDescriptor	
Attributes	String	EmsVersion
	String	type
Operations	getAllTopLevelSubnetworks	
	getAllTopLevelTopologicalLinks	
	getAllManagedElements	
	createTrafficDescriptor	
	deleteTrafficDescriptor	
	getAllTrafficDescriptors	

Object Name	Subnetwork

Object Description	This object represents a logical grouping or partitioning of NetworkElements in a way that is entirely determined by the EMS	
Super Class	TopObjects and multiLayerSubnetwork::MultiLayerSubnetwork_T	
ParentId	EMS object	
ChildIdList	TopologicalLink, SubnetworkConnection,ManagedElement	
Attributes	Topology_T	subnetworkType
	TransmmissionParameters::LayerRateList_T	supportedRates
Operations	activateSNC	
	createAndActivateSNC	
	createSNC	
	deactivateAndDeleteSNC	
	deactivateSNC	
	deleteSNC	
	getAllEdgePoints	
	getAllManagedElements	
	getAllSubnetworkConnections	
	getAllTopologicalLinks	

Object Name	**ManagedElement**	
Object Description	The managedElement represents an abstraction of a set of co-located physical resources (i.e. OADM, Switch) managed as a single entity by the an EMS	
Super Class	TopObjects and managedElement::ManagedElement_T	
ParentId	EMS	
ChildIdList	EquipmentHolder	
Attributes	String	Location
	String	Version
	String	productName
	CommunicationState_T	communicationState
	TransmissionParameters::LayerRate List_T	SupportedRates
Operations	createSNC (This is considering as CC)	
	getAllEquipment	
	getAllPTPs	
	getAllCrossConnections	
	setTPData	

Object Name	**CTP**

Object Description	a CTP(Connection Termination Point) is an actual or potential end point of a subnetwork connection. For a DWDM Och CTPs, it contains an attribute specifying the wavelength it occupies.	
Super Class	TP	
ParentId	PTP or CTP	
ChildIdList	TPIdList(containing lower order CTP's)	
Attributes	none	
Operations	getContainingTPs	

Object Name	EquipmentHolder	
Object Description	This object class represents resources of the NE that are capable of holding other physical components. Specific resources that are represented by instances of this entity are racks(bays), shelves, and slots.	
Super Class	TopObjects and equipment::EquipmentHolder_T	
ParentId	ManagedElement	
ChildIdList	Equipment , EquipmentHolder(with lower order)	
Attributes	Boolean	alarmReportingIndicator
	EquipmentHolderType_T	holderType
	NVSList	expectedOrInstalledEquipment
	EquipmentObjectTypeList_T	acceptableEquipmentTypeList
	HolderState_T	holderState
Operations	getAllEquipment	

Object Name	EquipmentHolder	
Object Description	This object class represents resources of the NE that are capable of holding other physical components. Specific resources that are represented by instances of this entity are racks(bays), shelves, and slots.	
Super Class	TopObjects and equipment::EquipmentHolder_T	
ParentId	ManagedElement	
ChildIdList	Equipment , EquipmentHolder(with lower order)	
Attributes	Boolean	alarmReportingIndicator
	EquipmentHolderType_T	holderType
	NVSList	expectedOrInstalledEquipment
	EquipmentObjectTypeList_T	acceptableEquipmentTypeList
	HolderState_T	holderState

Operations	getAllEquipment

Object Name	**CC**	
Object Description	A cross-Connection represents a physical connection within a managed element. A cross-Connection is atomic and is identified, similarly to an SNC in a singleton subnetwork, based on its external shape(A end(s), Z end(s), "SNC" type, and directionality).	
Super Class	TopObject and subnetworkConnection::CrossConnect_T	
ParentId	Equipment	
ChildIdList	none	
Attributes	Boolean	active
	ConnectionDirection_T	direction
	SNCType_T	ccType
	NVSList_T	aEndNameList
	NVSList_T	zEndNameList
Operations	none	

Object Name	**TrafficDescriptor**	
Object Description	Represents a collection of attributes that are used to define bandwidth and QoS characteristics on a CTP.	
Super Class	TopObject and trafficDescriptor::TrafficDecriptor_T	
ParentId	EMS	
ChildIdList	none	
Attributes	ServiceCategory_T	serviceCategory
	TrafficParameterList_T	trafficParameters
	String	conformanceDefinition
Operations	getAssociatedCTPs	

Note: These object classes are from IST-WINMAN project and is confidential information, where says TopObjects before was written LAMPIONTop, but was changed to adapt it to the STOLAS project . There are more classes but they are not in the scope of this report. For more information refer to WINMAN.

As we are going to consider each node as a Network Element, each node should be characterized. All the components of the node should be described to have a perfect performance, that means characterize the following components:

- Tunable transmitters
- 2R regenerators
- Wavelength converters
- Amplitude modulators
- Label detection modules
- Interfaces
- Optical switch controllers for the delay lines (FDLs)
- Label swappers
- Optical supervisory channel

Note: It is possible that many other components should be added or deleted to/from the list since the STOLAS project is still under development.

3.10. Optical Supervisory Channel

The OSC of the STOLAS project will use a 1310 nm wavelength that will be transported over the same fibre. The OSC is terminated electrically on every node in the ring. The OSC provides communication capabilities between Network Element (NE) and the EMS. The interface between the NE and the EMS is based on CORBA IIOP, which uses IP as transport protocol. All those interfaces and characteristics should be programmed on a C++ or Java base.

Note: under discussion within the STOLAS project.

3.11. Data Bases

The STOLAS/WINMAN solution requirements should consist of 3 types of data bases used for network configuration purposes and for fault and performance data:

Network inventory or Physical database:

The network inventory database maintains the physical inventory of the network resources, represents the equipment in the network. At least the following attributes should be found:

-Cards
-Ports
-Fiber conduits
-Location

Logical data base:

The logical data base maintains the logical inventory of the network resources, and consists of logical paths that are supported by physical resources of the network, and has at least the following attributes:

- Protection of the path
- Available bandwidth
- Network quality: Delay and Delay variation

Connectivity database:

All active or the ones that are going to be activated connections in the network are stored with their attributes like:

- Connection status
- Logical route
- Connection bandwidth
- Connection quality and delay

Fault topology database:

Should have at least the following attributes:

- Fault case
- Fault severity
- Location

3.12. Conclusions

The focus of this chapter has been done under the possibility of a general application of the STOLAS Management System for the demonstrator that is going to be used to probe the STOLAS scopes, but to be readapted to a large network, where many other components and issues could be necessary.

As an adaptation to the demonstrator where there are only three basic nodes: OADM, OXC and PhaSar, many blocks could be avoided. In that case, the use of data bases is not necessary, protection systems are not specified since we have not decided what kind of protection is needed for STOLAS, only monitoring of the components' performance, for that issue signaling protocols are needed, also signaling allows to discover failures, but between nodes. For a short testbed signaling protocols should be readapted. An adaptation of RSVP-TE signaling protocol has been decided for STOLAS since the project scope is not on resource reservation and it will be used for path discovery ,so an adaptation of that protocol has to be done.

Further work has to be done to characterize a large network within STOLAS Management System concept, that is, in case that a large network would be implemented, connection between internal node control and distributed network management system should be discussed.

CONCLUSIONS and FURTHER WORK

4. Conclusions

This report has described the STOLAS node architecture, signaling and routing concepts of MPLS and MPλS and the Manager System for the STOLAS project.

Summary conclusions are presented regarding the performance of the concepts.

4.1. STOLAS node architecture

After studying the STOLAS node architecture we conclude that the main points to take in account are:

Considering slotted operation, we know which input ports are active at the start time of every slot. It has to be said that the wavelength assignment algorithm should be done for each output fiber independently from each other. This assumes that each packet/burst has a unique output fiber, determined by a lookup table (case of GMPLS), and excludes the possibility of a routing mechanism.

The key component for the switch control unit is the scheduler, for our STOLAS node and after some comparisons and simulations made by IMEC we will use LAUC-VF as a scheduling algorithm.

Due to specific novel orthogonal modulation scheme to label signals in STOLAS, then the problem arises what to do when no data is preceding the burst, if there is nothing to FSK on it is impossible to send the information, one option is to use "dummy" packets or bursts with no information, but our option is not to use offset between header and payload and to use an input buffer made of FDLs. The dimensioning of the FDL buffer influences the level of congestion in the control part of the router, so a key issue is the correct dimensioning of the FDL buffer. Some implications are related to that decision, use of input FDLs for optical buffering that adds complexity to the control issues, appropriate FDLs lengths, performance of these FDLs, use of fixed or switched ones.

Since we are going to use BCH channel orthogonal to payload channel: link utilization increases in comparison with other out-of-band signaling systems like SCM. Orthogonal signaling can scale better than SCM signaling. Due to SCM has problems with dispersion and RF processing of subcarriers at higher speeds, this system is not as scalable as STOLAS system. However, at high rates at the STOLAS label swapper the appearance of chirp could be a problem. More study has to be done to verify this issue.

Link management and protection should be emphasize due to the high level of link utilization.

4.2. MPLS signaling, basis for STOLAS

After studying MPλS signaling protocols we conclude that:

Both CR–LDP and TE–RSVP provide very similar functionality for establishing traffic-engineered, labeled switched paths. Each has its strengths and weaknesses. While LDP is the younger of the two protocols, RSVP has been previously deployed and has operational experience. It is true that there have been extensive enhancements to RSVP in order to support the needs of MPLS. As both CR–LDP and TE–RSVP evolve they will offer more and more similar functionality.

Eventually, MPLS traffic engineering should evolve into a single entity that combines the best-of-breed attributes from both TE–RSVP and CR–LDP. In the meantime, any MPLS implementation by original equipment manufacturers (OEMs) developing LER or LSR platforms should consider supporting both TE–RSVP and CR–LDP to ensure interoperability.

For the STOLAS project RSVP-TE signaling protocol has been chosen, but since we are not going to make resource reservations, this protocol will be used only for path discovery, so some adaptations should be done to limit its functions.

The control channels may also be used for transmitting and receiving signaling and routing messages. Each LMP control channel must individually negotiate the control channel parameters, and each active control channel must exchange LMP Hello packets to maintain LMP connectivity. For LMP, it is essential that at least one control channel is always available.

4.3. STOLAS Management System

The focus of the management system has been done under the possibility of a general application of the STOLAS Management System for the demonstrator that is going to be used to probe the STOLAS scopes, but to be readapted to a large network, where many other components and issues could be necessary.

As an adaptation to the demonstrator where there are only three basic nodes: OADM, OXC and PhaSar, many blocks could be avoided. In that case, the use of data bases is not necessary, protection systems are not specified since we have not decided what kind of protection is needed for STOLAS, only monitoring of the components' performance, for that issue signaling protocols are needed, also signaling allows to discover failures, but between nodes. For a short testbed signaling protocols should be readapted. An adaptation of RSVP-TE signaling protocol has been decided for STOLAS since the project scope is not on resource reservation and it will be used for path discovery ,so an adaptation of that protocol has to be done.

Further work has to be done to characterize a large network within STOLAS Management System concept, that is, in case that a large network would be implemented, connection between internal node control and distributed network management system should be discussed.

4.4. Recommendations and further work

- Performance simulations between "classic" STOLAS node architecture, the multistage architecture and loop-back configurations in order to describe pros and cons of each architecture.
- Accurate design for the scheduling algorithm.
- Improving of header processing times; detection, process and switching.
- Readapting of RSVP-TE signaling protocol.
- Implementation of the data bases for the Management System (programming).

Abbreviations and References

Abbreviations

STOLAS	Switching Technologies for Optically Labeled Signals
WDM	Wavelength Division Multiplexing
IP	Internet protocol
RSVP	Resource Reservation Protocol
CR-LDP	Constrain route Label Distribution Protocol
MPLambdaS	Multiprotocol Lambda Switching
OBS	Optical Burst Switching
CoS	Class of Service
QoS	Quality of Service
ATM	Asynchronous transfer Mode
LSP	Labeled Switched Path
XC	Cross-Connect
LSR	Labeled Switched Route
IETF	Internet Engineering Task Force
IGP	Internal Gateway Protocol
BGP	Border Gateway Protocol
SCU	Switch Control Unit
FDL	Fiber Delay Line
BHP	Burst Header Packet
DCG	Data Channel Group
CCG	Control Channel Group
FIFO	First in first out
AWG	Array Waveguide Gratings
TWC	Tunable Wavelength Converter
LAUC-VF	Latest available unused channel with void filling
OXC	Optical Cross Connect
OTN	Optical Transport Network
LMP	Link Management Protocol
ADM	Add-Drop Multiplexer
TE	Traffic Engineering
OSPF	Open shorter path first
EMS	Element Management System
OLS	Optical Label Switching
NMS	Network Management System
SLA	Service Level Agreement
IST	Information Society Technologies
NO	Network Operator
NSP	Network service Provider
INMS	Integrated Network Management System
GUI	Graphic User Interface

CS	Connectivity Service
TP	Termination point
PTP	Physical termination point
OCh	Optical Channel
FEC	Forwarding Equivalence Classes
OME	Optical Management Element
OADM	Optical Add-Drop Multiplexer
OCP	Optical Connection point
SNC	Sub-network Connection
CTP	Connection Termination Point
OER	Optical Edge Router
CORBA	Common Object Request Broker Architecture
SDH	Synchronous Digital Hierarchy
IDL	Interface Definition Language
IM	Information Model
MIB	Management Information Base
OSC	Optical Supervisory Channel

References:

[1] J. Turner, "Terabit burst switching".
Journal High Speed Networks, vol 8, pp. 3-16. 1999.

[2] P. Gambini, et al., "Transparent Optical Packet Switching: Network architectura and demonstrators in the KEOPS project", IEEE J. Selec. Areas Commun., vol.16, pp. 1245-1259, Sept. 1998.

[3] M.K. Smith, C. van Dam. "Phasar-based WDM devices: principles, design and applications, IEEE J. Selec. Topics Quant. Electron, Vol. 2, 236-250, 1996.

[4] F. Callegati, H. Cankaya, Y. Xiong, and M. Vanderhoute, "Dessign of optical IP routers for Internet backbone applications". IEEE Commun. Magazine, pp. 124-128, Dec. 1999.

[5] A. Raboso, Graduation Thesis. "Performance Analysis of combined DPSK and ASK modulation Formats for Labeling of optical signals". Sept. 2001.pp.3-12.

[6] D. Awduche, et al., "Multiprotocol Lambda Switching: Combining MPLS Traffic Engineering Control with Optical Cross-Connects", draft-awduche-mpls-te-optical-03, April 2001.

[7] draft-ietf-mpls-lmp-02.txt

[8] draft-ietf-mpls-bundle-03.txt

[9] ITU-T Recommendation I.321 (1991)
B-ISDN Protocol reference model and its applications.

[10] K.Kompella et al., "OSPF Extensions to support MPLambdaS".
Draft-Kompella-ospf-mpls-extensions-00.txt.

[11] K.Kompella et al., "IS-IS Extensions to support MPLambdaS".
Draft-Kompella-isis-mpls-extensions-00.txt.

[12] OIF Contribution #2000.125.2 "User network interface 1.0 Proposal", 31/10/2000.

[13] Y.Cao, Y.Xu, "ODSI Signaling specifications_optical Gateway Protocol, version 1.0", July 2000.

[14] OIF Contribution #2000.171.0 "LDP extension for optical UNI signaling", November 2000.

[15] OIF Contribution #2000.140.1 "RSVP extension for optical UNI signaling", August 2000.

[16] Chunming Qiao, "Labeled Optical Burst Switching for IP-over-WDM Integration", IEEE Communication Magazine, September 2000, pp.104-114.

[17] D.J. Blumenthal, et al., "All-optical label swapping networks and technologies", IEEE J. of Lightwave Technol., Dec. 2000, pp. 2058-2075.

Figure references and other references:

[1] figures

- Jan Cheyns, Chris Develder, Erik v. Breusegem, Piet Demeester, "Routing in an AWG Based Optical Packet Switch". Submission for Photonic Network Communications. March 2002.

- www.netplane.com

- www.dataconnection.com

[2] figures

- Uyless Black. "Optical Networks, third generation transport systems". Prentice Hall,2002.

[3] figures

- Y. Xiong, Marc Vanderhoute. "Control architecture in Optical Burst-Switched WDM Network". J. Selected Areas Comm. Vol. 18. Oct 2000.

- "IP over DWDM Networks supported by GMPLS-based LOBS deploying Combined modulation Format".
  Sulur, Ton Koonen, Idelfonso Tafur Monroy, Huug de Waardt, Jean Jennen, Opticomm 2001, August 2001, Denver, Colorado (USA)

- "Optical packet routing in IP over WDM networks deploying two-level optical labeling".
  Ton Koonen, Geert Morthier, Jean Jennen, Huug de Waardt, Piet Demeester, ECOC 2001, October 2001, Amsterdam, The Netherlands.

[4] figures

- IST-WINMAN project, Deliverables (internal documents).

Eindhoven University of Technology
Faculty of Electrical Engineering
Division of Telecommunication Technology and Electromagnetics

Architecture and Management Of Optical Burst Switching Nodes

Gustavo Gil Lopez
Graduation Thesis

Performed at:	Electro-Optical Communication Group
Period:	October 2001-June 2002
Graduate Professor:	Prof. ir. A. M. J. Koonen
Supervisors:	Dr. ir. I. Tafur Monroy
	Dr. ir. H. de Waardt

Abstract

With the exponential growth in Internet traffic, scaling of Internet Protocol (IP) data networks far beyond their present performances is required. In order to meet this challenge, Multiprotocol Label Switching (MPLS) is rapidly emerging as the technology that allows for enhanced speed and scalability as well as service providing capabilities in the Internet. On the other hand, recent developments in WDM technology have dramatically increased the traffic capacities of optical networks. As the electronic processing is becoming the weak point for the bit-rate (speed) of the networks, research is ongoing to introduce more intelligence in the control plane of the optical transport systems, which will make them more survivable, flexible, controllable and open to traffic engineering techniques. The research efforts are focusing on the efficient interworking and integration of higher layers IP and WDM transport layer. Direct integration and transport is supported by Multiprotocol Lambda Switching (MPλS), which is based on MPLS, and generalizes the use of label-switching concept to incorporate optical label switching.

One promising technique to support IP over WDM networks is Optical Burst Switching (OBS). OBS, exploiting wavelength division multiplexing (WDM) fits well into the MPλS concept, while offering flexibility, enabling scaling to very high bit rates and eliminating potential bottlenecks in the electronic header processors like in Optical Packet Switching. In optical burst switched networks, MPLS protocol-related signaling can be carried using control channels. For instance, in the recent IST-STOLAS project, cross-connect nodes can make use of MPLS messages to control burst switching, bandwidth administration, delays reservations and wavelength conversion.

In this report, firstly, a general view of the OBS concept is introduced. In order to describe the STOLAS optical burst switching node architecture, an analogy between the OBS node and the STOLAS node has been included. After this description, we have focused on the main building blocks of the STOLAS burst witching node, array waveguide grating (AWG), fiber delay lines (FDL), tunable laser sources and controllers, to find the operation parameters for its performance according to the specification adopted within the STOLAS project. This in it turns, means to consider the node architecture, specifically issues related to the technology being use, and the parameters related to the data traffic characteristics. In order to realize such a description a deep study on the OBS node techniques and the STOLAS specific characteristics has been carried out. This study has shown that it is not possible to use the typical OBS scheme for the STOLAS project, since within the project time offset and orthogonal modulation are not compatible. However, due to the specific novel orthogonal modulation scheme to label the signals in STOLAS, a number of adaptations should be done.

Secondly, MPLS routing and signaling protocols have been proposed for STOLAS. After a brief introduction to the signaling protocols in MPLS, we described what could be best solution for the signaling problem inside the network. The proposed solution is based on an exhaustive comparison between RSVP-TE and CR-LDP protocols. Moreover, a part of this report is also dedicated to MPλS and the control and

data plane internetworking. A conclusion drawn form this study is that because STOLAS is not using resource reservation an adaptation of signaling protocols has to be carried out.

Thirdly, a complete study of the characteristics of the Management System for STOLAS was accomplished. As the development of a network management system is not the focus of the project we adapted IST-WINMAN Management System approach for STOLAS. A list of new Elements (specific to STOLAS) has been made. These new elements have to be controlled inside the nodes and along the network based on a system of data bases that should be implemented for the STOLAS implementation.

Finally, this report presents a list of conclusions and recommendations for further work has been summarized. Guidelines on how to improve the performance of the burst switching node and management system are also included.

My own contribution within the project was, after assisting to many workshops and discussions about network and node architecture, to make an accurate inventory of the node architecture, also proposing some configurations. I have done a deep study on MPLS protocols and a first step on the adaptation for the Management System in STOLAS.

Index

0. Introduction

The main overall objective of the STOLAS project is to improve the throughput of packet switched networks, by providing novel optical routing techniques. Additional overall objectives are to develop key optical functionalities for optical packet routing, and to validate the technical feasibility in a limited-scale testbed.

The amount of packet-based data traffic in today's telecommunication networks is increasing steeply, and in many networks has already overtaken circuit-switched traffic. In order to answer the exploding demand for network capacity, wavelength division multiplexing (WDM) is being introduced consecutively in core networks and in metropolitan area networks, and will penetrate in access networks in the near future.

A novel approach termed lambda labeling or Multi-Protocol Lambda Switching (MPλS) has been presented for direct IP-over-WDM transport. The scheme has been derived from the IP-MPLS protocol and the label-switching concept to provision wavelength-switched channels; the label-switched paths (LSPs) are associated with optical channels. Optical packets (or bursts of packets) are switched to other wavelengths (label swapping or switching) at the transmitters in the edge nodes; wavelength routers in the core nodes guide the packets along the appropriate network paths. The label swapping or switching can be done by wavelength converters of which the pump signal is generated by a wavelength-tunable laser.

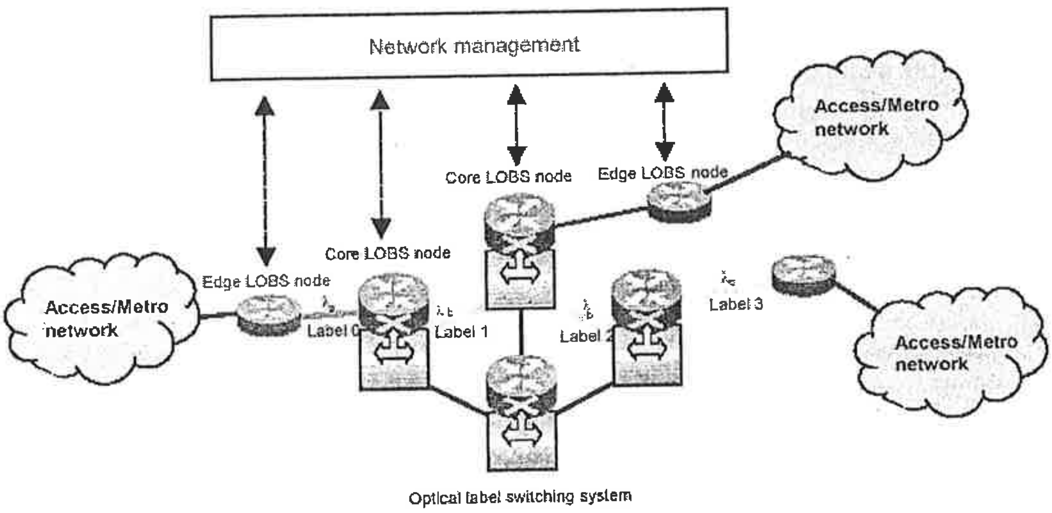


Fig.1. Burst IP packets over WDM network supported the GMPLS and OBS

In addition to optically labeling a packet (or burst of packets) by assigning a particular wavelength to it, a next-level label can be attached by using a particular modulation scheme. The latter can be done by modulating the label information orthogonally to the packet data. Having two optical labels (wavelength and DPSK/FSK) is in agreement with the label-stacking as foreseen in MPLS (i.e., 2 or more labels corresponding with a packet stream). The orthogonally modulated label also facilitates the implementation of multi-casting the streams. Furthermore, the orthogonal channel can be used as an embedded network management and control channel, e. g., for supervising the routing (crossconnect and add/drop) functions.

The STOLAS project will focus on some key issues for optical packet/burst routed networks:

- Tunable lasers for wavelength labeling which can be wavelength-switched very fast, including the control electronics.
- Wavelength converters for optical label swapping on two levels (wavelength, and an orthogonal modulation domain such as phase or frequency), including the label processing electronics.
- Optical-label-controlled crossconnects and add/drop nodes.
- 2R multi-wavelength regenerator
- Management and control of the optical labeling and routing functions, and the validation of the optical routing techniques in a small-scale testbed containing a few edge routers and core routers.

GMPLS includes wavelength switching next to label switching and fiber (space) switching. It provides quick and efficient forwarding of IP packets because it uses only a single forwarding algorithm based on label swapping (removal and reinsertion). The label is a short, fixed length value that is contained in each IP packet or forwarding equivalence class (FEC) and used to forward the packet through the network. Therefore, IP/GMPLS is a low latency, low overhead routing technique that simplifies packet forwarding and enables scaling to terabit rates.

GMPLS-based LOBS [16] provides an effective IP over WDM networking due to the advantages of optical burst switching (OBS). OBS, which represents a balance between circuit and packet switching, has advantages in comparison with optical packet switching (OPS), namely, a label and its corresponding burst data in LOBS can be much more loosely coupled in both space (by using separate label and data wavelength) and time (by using a nonzero offset time) than a label and its payload are in OPS, the requirements for processing label and synchronizing between burst (as well as between burst data and its label) in OBS can be much less stringent than those for processing label and synchronizing between payload (as well as between payload and its label) in OPS. In addition, OBS can provide differentiated services (priority scheme) by using an extra offset time. OBS can avoid optical buffer, as long the offset time is longer than the total delay from router followed by burst IP packet. OBS is able to offer delayed reservation through the offset time.

In the GMPLS-based LOBS, burst data can be composed by assembling several IP packets (in ingress LOBS node). It may contain several megabytes of data. However, a small size of bust data is still possible to occur, therefore, for efficiency of network throughput, the GMPLS-based LOBS with segmented burst data is preferred.

Two main approaches have been investigated for labeling of optical signals: bit serial label and optical sub-carrier multiplexed label. The former multiplexes a fixed bit rate label in front of the IP packet while the latter multiplexes the label information that is modulated a RF sub-carrier in the IP packets. Both approaches put the label at the same wavelength of the IP packet [17]. We propose a new approach, based on orthogonal modulation of the label information with respect to the modulation format of the IP payload.

CHAPTER 1:

Architecture for an Optical Burst Switch Node

1. NODE ARCHITECTURE FOR AN OPTICAL BURST SWITCH ROUTER

1.1. Generic Node Architecture

In a generic optical burst router three different parts can be distinguished: the input interface, the switching element and the output interface. Figure (1) shows the modular structure of a typical optical burst switch router.

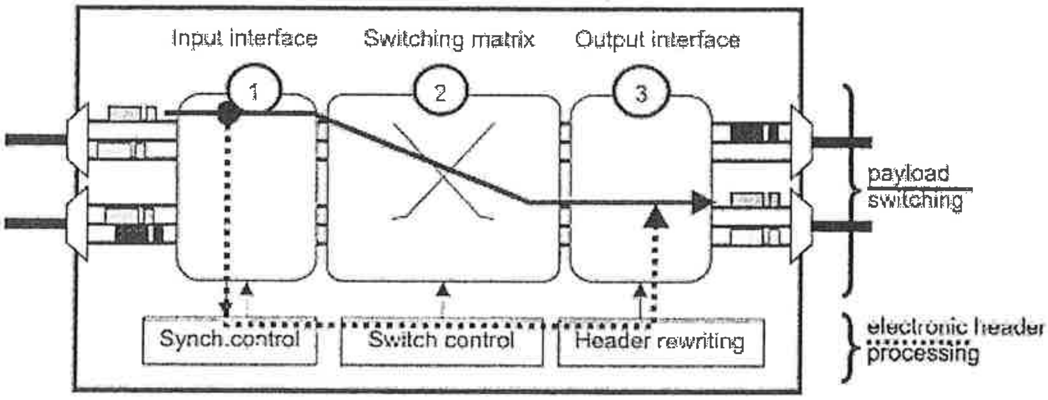


Fig (1). Generic Optical Burst router

1.T he input interface consists of: the demultiplexer or channel mapper and the input FDLs, Later on, this chapter goes deeper is some of the elements.

2. The switching element consists: the switching matrix, the input wavelength converters, the SCU, the scheduler, the drop/multicast element and the output wavelength converters. The focus of this chapter is on this part of the router: the switching element.

3.T he output interface that consists of: the header rewriter and the multiplexer.

Although the control element is placed in the switching element, it should be said that this component is connected to all the components.

Through this chapter we will make an analogy between the typical optical burst switch and the STOLAS core router, trying to identify the elements present in both and the consequences of their use.

The architecture of an N x M optical core router is shown in Fig. (2), which mainly consists of input FDLs (fiber delay lines), an optical switching matrix, a switch control unit (SCU), also called switch matrix controller, and routing and signaling processors. Data channels are connected to the optical switching matrix and control channels are terminated at the SCU. Channel mapping logically decouples the channels from physical fibers wavelengths. The input FDLs are used to delay the arriving data bursts and synchronize them, thus allowing the SCU to have enough time to process the associated BHPs. Data bursts still remain in the form of optical signals in the core routers. The optical buffers of FDLs are used to resolve data burst contentions on outgoing DCGs (data channel groups). The use of electronic buffers instead of FDL optical buffers was considered in [1]. Note that there are J incoming DCGs and J' outgoing DCGs in Fig.(2). A typical example of the general architecture is a symmetric router with equal number of input and output fibers is given in fig (2), where each fiber has one DCG of n wavelength channels and one CCG (control channel group) of *k* channels.

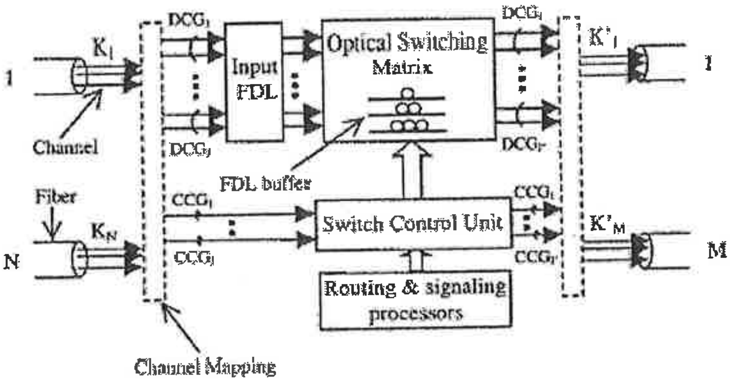


Fig (2). Optical Core Router architecture (general)
Addapted from figure references [3].

A block diagram of an N x N non blocking optical switching matrix is given in Fig.(3) where the spatial switch is able to switch a data burst from any incoming wavelength/channel to any FDL as long as it does not overlap with other data bursts. Each optical buffer has B WDM FDLs with i^{th} FDL being able to delay Qi time slots, $1<=I<=B$, and it is assumed that $Q1<Q2<....<Qb$. Note that an FDL in Fig. (2) has $(K-k)$ wavelengths. By default there is always an FDL with zero delay time, denoted by 0 with $Q0 = 0$.

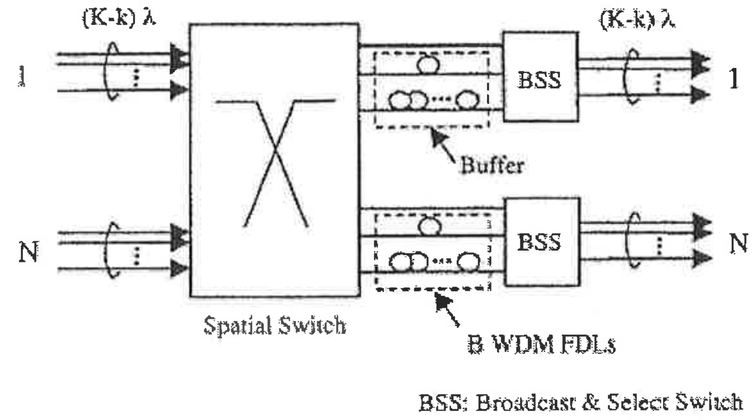


Fig. (3). Block diagram of a non-blocking (symmetric) optical switching matrix. Addapted from figure rferences [3].

The function of the SCU in Fig.(2) is similar to that of a conventional electronic router. The routing processor runs routing and control protocols for the whole OBS network. It creates and maintains a routing table and computes the forwarding table for the SCU. Forwarding can be connectionless or connection- oriented (prior path establishment through signaling). Apart from forwarding table lookup, the SCU decides on which outgoing DCG and CCG to forward each arriving data burst and its BHP. If free data and control channels are available from these groups, either when the data burst arrives at the optical switching matrix or after some delay in an FDL buffer, the SCU will select the FDL and configure the optical switching matrix to let the data burst pass through. Otherwise, the data burst is dropped. While arranging the transfer of a data burst and its corresponding BHP in the optical switching matrix and SCU, respectively, the SCU tries to resynchronize the data burst and the BHP by keeping the offset time τ as close as possible to $\tau 0$. If a data burst enters the optical switching matrix before its BHP has been processed (this phenomenon is called *early burst arrivals*), the burst is simply "dropped." This is because data bursts are optical analog signals. If no path is set up when a data burst enters the optical switching matrix, it is lost. Since a BHP and its data burst are switched in the SCU and the optical switching matrix, respectively, the delay Δ introduced by the input FDL should be properly engineered such that under the normal traffic condition data bursts are rarely dropped due to early arrivals.

The block diagram of a SCU is given in Fig.(4). Depending on the optical switching matrix, the SCU can have either a centralized configuration as shown in Fig. (4) or a distributed configuration.

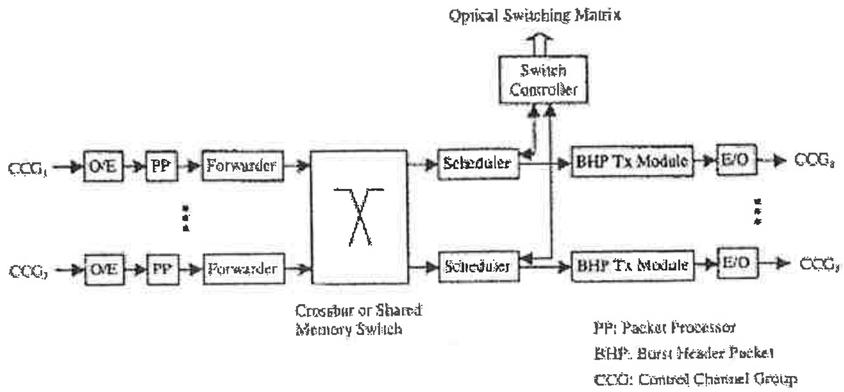


Fig. (4). block diagram of the switch control unit (centralized control). Addapted from figure references [3]

In a distributed configuration, each scheduler has its own switch controller. Distributed configuration could be applied to the broadcast-and-select type switches [2]. Here we focus on the description of a centralized configuration. The functionality of each building block in Fig. (4) is detailed below.

The forwarder performs the forwarding table lookup to decide on which outgoing CCG to forward the BHP. The associated data burst will be forwarded to the corresponding DCG. The mapping of logical channels to physical wavelengths and fibers is done in the forwarder. The forwarder then simply forward BHPs across the switch in a certain order (e.g., FIFO). To reduce the switch delay, it is preferred to use a switch with output-queueing discipline. To support multicast traffic, the switch requires native multicast capability. Otherwise, copies of a multicast BHP are made in the forwarder.

The scheduler in Fig.(4) is responsible for both the scheduling of the switch of the data burst on an outgoing data channel and the scheduling of the transmission of its BHP on an outgoing control channel. The scheduler is optical switching matrix specific. For the nonblocking optical switching matrix in Fig. (3), there is one scheduler for each DCG and CCG pair, and each scheduler only needs to keep track of the busy/idle periods of a single outgoing DCG and an outgoing CCG.

The scheduler works as follows. It first reads the time-stamp and the data burst duration information from a BHP to determine when the corresponding data burst will enter the optical switching matrix and how long the data burst will last. It then searches for an idle outgoing data channel time slot to carry the data burst, making potential use of the FDLs to delay the data burst. Once the idle outgoing data channel is found and the FDL to be used (if necessary) is determined, the scheduler knows the departure time of the data burst from the optical switching matrix. Subsequently, it schedules the time to send out the BHP on the outgoing CCG, trying to resynchronize the BHP with the data burst. After successfully scheduling the transfer of the data burst and its BHP, the scheduler will send the configuration information to the switch controller which will in turn configure, just-in-time, the optical switching matrix accordingly to let the data burst pass through. The configuration information includes incoming data channel identifier, outgoing data channel identifier, time to switch the data burst, duration of the data burst, and the FDL buffer identifier.

The scheduler is bi-directionally connected to the switch controller. After processing the configuration information sent by the scheduler, the switch controller sends back an acknowledgment to the scheduler. The scheduler then updates the state information of the DCG and CCG, modifies the BHP (e.g., the offset time and the data channel identifier) and passes it along with the time-to-send BHP information to the BHP transmission (Tx) module.

One of the reasons we need an acknowledgment from the switch controller is to limit the number of configuration requests from the schedulers. Hence, the maximum time for the switch controller to process the request can be estimated.

1.2. STOLAS node architecture

A possible architecture for such a node is made combining Array Waveguide Gratings (AWG) and all-optical tunable wavelength converters (TWC). We will see that an inherent problem of this switch is its internal blocking. This drawback can be greatly overcome by using an intelligent and efficient wavelength algorithm within the node.

The core element of the switching node is the switching matrix, and is also the most critical for switching speed. The time between the controller having calculated a new configuration for the switching matrix and the moment when the switch is stable in the new state must be short. Crucial for this element is also its transparency. The switch should just guide the light containing the payload information from input to output, without using any information on bitrate, data coding technique....

Firstly, we present the principle of operation of AWG, secondly we present several node architectures for the STOLAS node using AWG's plus tunable laser sources.

1.2.1. The Array Waveguide Grating with Tunable Wavelength Converter

An Array Waveguide Grating (AWG) is a diffractive element, and as such it doesn't contain any active devices, making it a very reliable component. It is sometimes also referred to as a PHASed ARray (phasar). We will shortly describe the functionality of an AWG, without giving any physical details (these can be found in e.g. [3]). The dashed box in Fig. (5) shows a black box model of an 8x8 AWG. Depending on the wavelength at input port i, the light comes out at output port j. This allows us to write out a simple matrix M, where the input ports are written vertically (index i) and the output ports horizontally (index j). An element Mji in this matrix is the wavelength, which will come out on port j, if it entered the device at input port i. All these wavelengths are equally spaced by $\Delta\lambda$. An AWG has a cyclic nature: if $Mji=\lambda a$, there exists a λcyc for which $\lambda b=\lambda\alpha+k.\lambda cyc$ ($k \in \mathbf{Z}$, λcyc is called the period) shows the same behaviour as λa. In other words, if λb enters on port i, it will also exit on port j, just as it was the case for λa. It is possible to design the AWG in such a way that for an NxN AWG, the period is $N*\Delta\lambda$. Putting all of this together, we come to a table like the one shown in Fig. 6. In this figure another property of the AWG can be seen. Let $\lambda i=\lambda 0+i.\Delta\lambda$. If from a certain input port λi goes to output port k, then $\lambda i+1$ will go to output port (k+1)mod N. This is a physical constraint that has an impact on the available configuration options, since it limits the degree of freedom to a large extent.

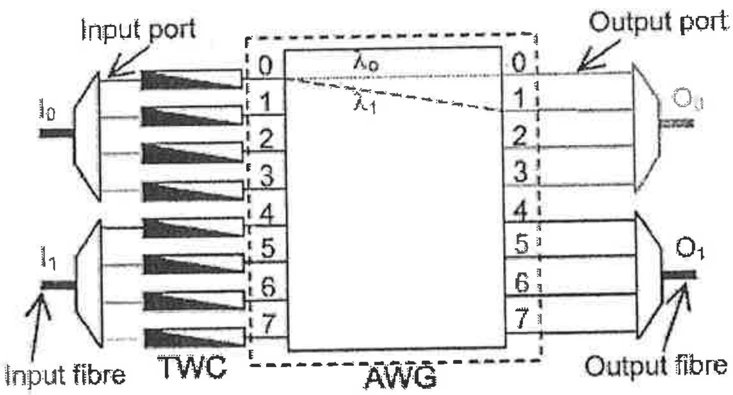


Fig. (5) STOLAS node: classic configuration.
Adapted from figure references [1].

		Input ports							
		0	1	2	3	4	5	6	7
Output ports	0	λ_0	λ_7	λ_6	λ_5	λ_4	λ_3	λ_2	λ_1
	1	λ_1	λ_0	λ_7	λ_6	λ_5	λ_4	λ_3	λ_2
	2	λ_2	λ_1	λ_0	λ_7	λ_6	λ_5	λ_4	λ_3
	3	λ_3	λ_2	λ_1	λ_0	λ_7	λ_6	λ_5	λ_4
	4	λ_4	λ_3	λ_2	λ_1	λ_0	λ_7	λ_6	λ_5
	5	λ_5	λ_4	λ_3	λ_2	λ_1	λ_0	λ_7	λ_6
	6	λ_6	λ_5	λ_4	λ_3	λ_2	λ_1	λ_0	λ_7
	7	λ_7	λ_6	λ_5	λ_4	λ_3	λ_2	λ_1	λ_0

Fig. 3 An AWG table representation

As an example two configurations of the AWG are shown on both Fig.(5) and Fig.(6), in order to clarify this notation further. The signal at input port 4 can be sent to output port 4 if it were on λ0, (dotted line) while it would end up at output port 5 if the wavelength is λ1(full line). I will continue working with this example of an 8x8 AWG, and a switch based upon it.

The table learns that only converting these input ports to either *l1* or *l3* can do this. So we have a problem, since three ports are contending for only two wavelengths, resulting in the loss of one of the input ports, since this loss is not due to contention (problems between wavelengths and packets destined in a same output fiber) is due to internal blocking of this configuration of the switch, conversion to some wavelengths can degrade the function of the node. So it is clear that the way the output ports are combined into an out put fiber is very important and has a severe impact on the performance of the node.

Now that the concept of the node is clear, let's move on to the operation of the node. At a certain moment in time a header will arrive, indicating what to do with the associated payload. First the header is processed to determine the output fiber. Using the table for that specific output fiber, we need to find out to which wavelength has the signal to be converted by the TWC in order to switching to that correct output fiber. Thus, a wavelength assignment algorithm is needed, considering slotted operation, we know which input ports are active at the start time of every slot. It has to be said that the wavelength assignment algorithm should be done for each output fiber independently from each other. This assumes that each packet/burst has a unique output fiber, determined by a lookup table (case of GMPLS), and excludes the possibility of a routing mechanism.

Different approaches can be taken for this design, heuristic, random or maximum matching, for example:

- LUW-LFP (heuristic)
- Maximum Matching
- RORA (random)

In figure (7), a comparison of the blocking probability performance is illustrated, for a Poisson traffic and load of 0.8. Clearly both the heuristic and the maximum matching have about the same performance, if we consider the time of simulation and performance, heuristic is a lot slower than maximum matching, so we will consider Maximum Matching as our wavelength assignment algorithm.

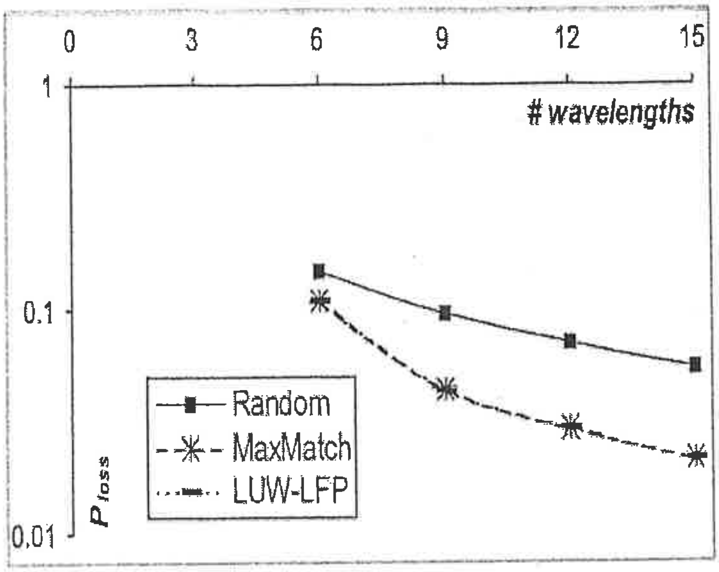


Fig. (7). Performance comparison of thed ifferent wavelength algorithm, depending on then umber of wavelengths per fiber. (IMEC simulations)

But in our "classic" STOLAS node, and because of the internal structure of the TWC, it is not available the conversion to the original wavelength, that means that the we have some unfairness performance of the node because some ports have less blocking probability than others. At low loads it provokes a severe deterioration of the node performance, even for higher loads. See figure (9).

	in							
out	0	1	2	3	4	5	6	7
0						λ_3		λ_1
1	λ_1	λ_0					λ_3	λ_2
2	λ_2		λ_0					
3	λ_3	λ_2	λ_1	λ_0				
4		λ_3		λ_1				
5			λ_3	λ_2	λ_1	λ_0		
6					λ_2		λ_0	
7					λ_3	λ_2	λ_1	λ_0

Fig (8).Table when conversion to original wavelength is not possible

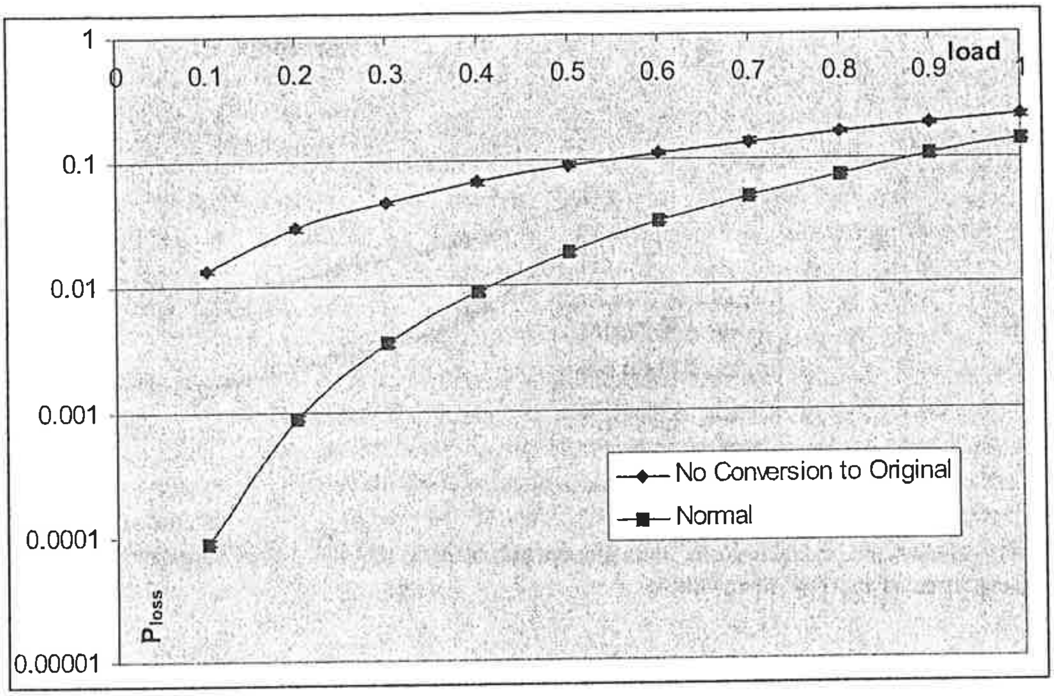


Fig. (9). Performance of a 2-fibre, 4-wavelengths STOLAS Node with and without Original Conversion. (IMEC simulations).

How can we solve the problems? let's look for different configurations of the node. There are other possibilities.

1. An alternative for this could be the over-dimensioning of the internal wavelengths in the node.

Instead of using only the W wavelengths, which are present on a fiber, use FW wavelengths (the dimension of the AWG). In this way the table of the AWG in the STOLAS node becomes:

		output ports							
		0	1	2	3	4	5	6	7
input ports	0	0	1	2	3	4	5	6	7
	1	7	0	1	2	3	4	5	6
	2	6	7	0	1	2	3	4	5
	3	5	6	7	0	1	2	3	4
	4	4	5	6	7	0	1	2	3
	5	3	4	5	6	7	0	1	2
	6	2	3	4	5	6	7	0	1
	7	1	2	3	4	5	6	7	0

The dark and light shadow denote a certain output fiber, note that any combination is now good in contrary to the "classic" case. We see that we can direct any input signal (less or equal than 4 of course) to different output ports. Only problem is that we no longer have wavelengths 0 to 3 at the output ports. So an extra wavelength conversion stage would be needed. And if they would be completely flexible, the dual labeling process would again become feasible. The node would then look like :

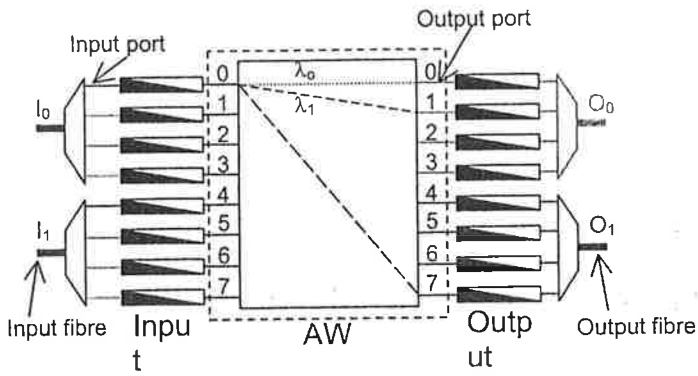


Fig. (10). Over-dimensioning of the internal wavelengths.
Adapted from figure references[1].

The downside of this architecture is probably its scalability, an FWxFW AWG would be needed (can this behavior be emulated by some connection of smaller AWGs??). Notice however that although the number of TWCs needed is doubled, it stays a linear function!

Upgradability is also poor: if more wavelengths are to be used, the range of the TWC has to be increased and a larger AWG is needed. However this also seems the case for the classic STOLAS node architecture

2.Anoth er configuration that could avoid the scalability problems and technological limits is the multistage architecture. A 2-stage non- blocking solution, based on a three stage Clos switch is presented in figure (11).

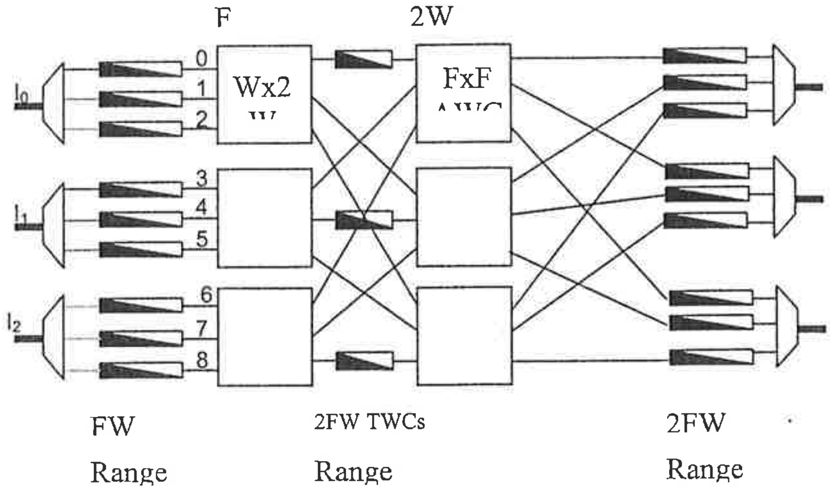


Fig. (11). 2-stage STOLAS architecture

This architecture is non-blocking, in the first stage an input port can be routed to any second stage switch, in these second stage, the output port of the AWG immediately determines the output fiber. Note that in the Clos three stage switch there would also be an AWG in the third stage, however switching signals into the same output fiber is done by the third stage TWCs.

1.3. Contention avoiding

For the contention problem in STOLAS, the proposal is to use a combination of accurate scheduling and retransmission (loop-back configuration) with FDL. We have to be sure that the edge nodes send data only at "good" moments in time, and refrain from sending at "bad" moments. That means that for each slot time every node has a schedule of how many wavelengths go to a certain output.

The scheduler is the key component in the SCU. The design of this block poses a new challenge, as it has to work at extremely high speed (100-200 ns per BHP) and the gaps introduced by the FDL will greatly complicate the design.

The core part of this scheduler is the data channel scheduling algorithm. For instance, LAUC-VF (*latest available unused channel with void filling*) algorithm can be used for this purpose. How it works? The gap between two data bursts in a channel is unused channel capacity, this algorithm try to fill the voids with new arriving data bursts.

The implementation complexity will depend on the DCG size, the BHP delay in the SCU, the maximum data burst length, the switched FDL design, and the data burst characteristics. To make the scheduling more strict many rules could be imposed on it in order to avoid interclass contention, like avoid first class traffic over x %. To do so, the scheduler keeps track of the utilization of the control and data channels.

To adapt the switch FDL module for back-loop configuration to the node the AWG should be over dimensioned with one or several inputs and outputs ports. In that way, data bursts affected by contention could be routed to those outputs to be looped-back through the AWG. The delay of the switched FDL will be set to accommodate the burst, depending on the length, by the SCU and the scheduler since all the necessary data is contained on the BHP. After the FDL module there should be another TWC capable to convert to all the different wavelengths in the node. The final scheme is shown in figure (12).

Another issue to take into account is that the scheduler has to be updated periodically managed by the management system, that means that the delay related to this operation could be a big concern.

One of the main parameters in OBS is the time offset between the header and the payload, that offset tries to avoid burst losses. In the STOLAS project there is a big handicap because our novelty is to send the header information in the signal phase orthogonal to the signal amplitude. But then the problem arises what to do when "no light" (no data) is preceding the burst, meaning that we have no amplitude to perform FSK on. This rises new issues.

One proposal solution is to use no-time offset and to use an optical buffer to store the payload while the header is being processed. So this buffer should be carefully dimensioned to support the different CoS and the variable burst length.

Within the STOLAS project further work should be done in this aspect to decide what kind of scheduling and control is the best solution for our system and node architecture.

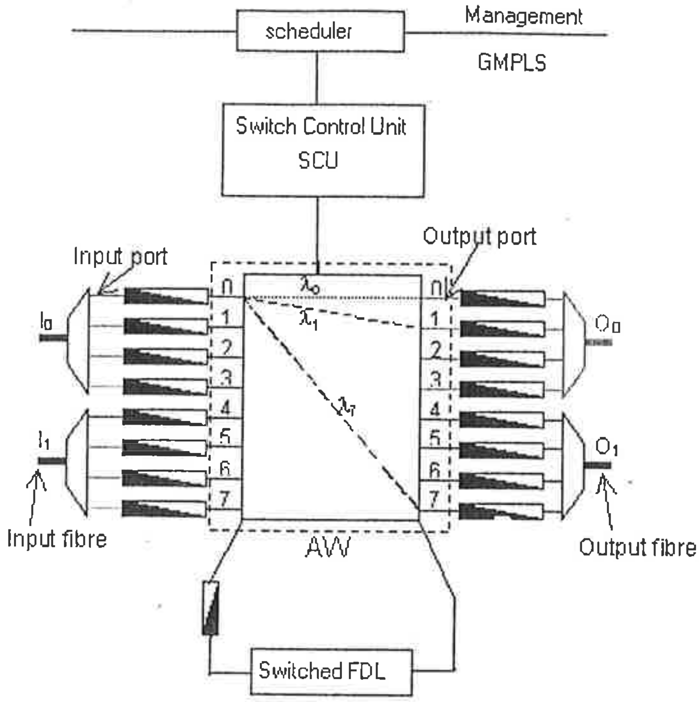


Fig. (12). Loop-back configuration for STOLAS node.

1.4. Burst length, timings and dimensioning

Except for the separate transmission of headers and payloads and being switching in different domains, there is no fundamental difference between packet switching and OBS.

The link utilization of the OBS network will largely depend on the number of channels dedicated to transmitting BHP (as well as other control packets) and the guards in each data burst.

Consider a WDM link having K channels with k control channels, and $K-k$ data channels, $1 \leq k \leq K$. Suppose the data channel rate is R Gb/s and the control channel rate is r *Gb/s*. The maximum link utilization will be η, defined as:

$$\eta = \frac{(K-k)R}{(K-k)R + kr}$$

As a data burst can be sent out on a data channel only if its BHP can be sent out on a control channel, there is a minimum requirement for the average data burst length in order to prevent congestion on control channels [4], so the performance of the control channel is strongly related to the average burst length. Since we will often deal with time domain issues in OBS, it is convenient to use time duration instead of bytes to represent lengths. The basic time unit is µs.

Let L_b be the average duration of a data burst (or $R*L_b$ *kbits* in length) and L_h the average duration of a BHP. Consider that both control and data channel are fully loaded. Under this situation, the maximum average BHP transmission rate is k/L_h BHPs per microsecond, and the maximum transmission rate is $(K-k)/L_b$ data bursts per microsecond. Since $(K-k)/L_b \leq k/L_h$, we have:

$$Lb \geq \frac{(K-k)\mathrm{Lh}}{\mathrm{k}} \qquad (1)$$

We have to consider the guard period, say θ *µs*, for each data burst. The average data burst duration that actually carries user data would be only L_b- θ and the burst overhead is θ / L_b

$$Lb \geq \frac{\theta}{\varepsilon} \qquad (2)$$

Inequalities (1) and (2) will together determine the minimum average data burst length.

Let us look at the impact of the processing speed of the scheduler and switch controller on the burst arrival rate to an $N \times N$ optical router. Suppose the processing time is divided in three, detection, process and switch time. This processing time can be modeled as a single-server queuing system. Where λ is the average burst arrival rate per fiber. For a stable queuing system, the service rate must be larger than the arrival rate. We have $\lambda<1/\delta d$, $\lambda<1/(\delta s+X)$, and $\lambda<1/(N*\delta c)$,which leads to:

$$\lambda < \min\left\{\frac{1}{\delta d}, \frac{1}{\delta s + X'}, \frac{1}{N \cdot \delta c}\right\} \qquad (3)$$

where $X(\delta c < X < N*\delta c)$ is the average time spent in the switch controller. For distributed control, $X=0$ and $N=1$. If ρ is the data channel utilization, given λ, the average burst duration (in *µs*) can be expressed:

$$Lb = \frac{\rho(K-k)}{\lambda} \qquad (4)$$

Note that inequalities (1) and (2) are from the burst transmission viewpoint, while (3) and (4) consider the switching aspect. The burst arrival rate is also closely related to the burst assembly mechanism used at edge router.

Since in the STOLAS project we are not going to use time offset the dimensioning of the input FDLs has to be done long enough to retain the bursts while the control packet is processed and the scheduler assigns the slot to go out.

Here the main problem is the dimension of the basic unit of timescale D. If the FDL buffer is designed with equally spaced lengths, consecutive multiples of D, small values of D lead to high time resolution and poor buffering capacity, while large values of D lead to large buffering capacity with poor time resolution. There is a trade-off between them to provide optimal burst loss probability.

The dimensioning of the FDL buffer may also influence the level of congestion in the control part of the router, so it can not be carried out without taking into account the control part of the architecture.

In principle the input FDL buffer may be variable or fixed length. The former may be done by an array of FDLs with different length and optical gates selecting the length. The delay could be chosen according to the estimated load in the control part of the router to limit the delay to a required minimum. This solution would significantly increase the complexity of the router. Furthermore, management of the input delays would be an additional task of no negligible logical complexity.

It is easier to assume a fixed delay length FDL, introducing a fixed delay of l μs. In this case, both implementation and management are straightforward. The only problem is to determine the value of l, and this one is closely related to the BCP processing time. If there is only one centralized queuing point for all BCPs, that is at the level of the forwarding table lookup, this point is just modeled as an *M/M/1* system, while the other delays can be characterized by a deterministic delay that can be added to the variable one. To model this variable part of the processing time, we assume that it has an average value equal to υ. So, the processing delay complementary distribution becomes:

$$\Pr(\delta > t) = \mathrm{e}^{-(1/\upsilon - N(K-k)\lambda)(t-\upsilon^*)} \qquad (5)$$

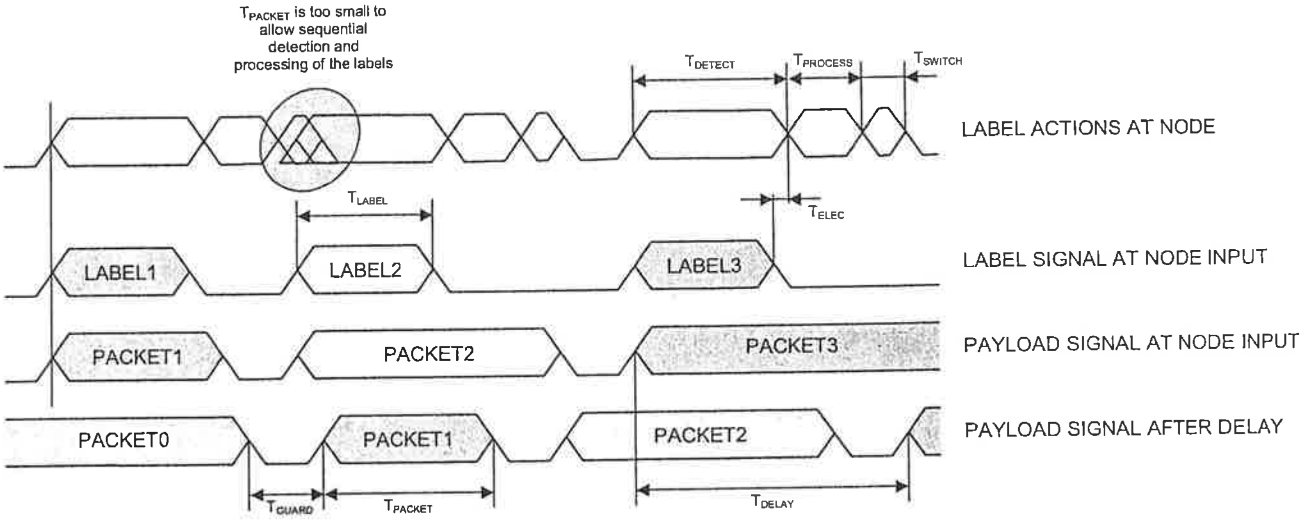
Once the QoS in terms of burst loss probability is given and using this formula, the FDL buffer length is easy to calculate. Anyway, we have to consider all the possibilities to calculate the optical buffer delay, and those are:

- Best case: only one timeslot of delay.
- Statistical: using formula.
- Worst case: upper bound, adding processing times.

The performance of the control channel is strongly related to the average burst length as shown in (1). A limit on the average burst length as a function of the processing capacity in the control logic may be calculated in a similar way, leading to a similar formula. Since the average BCP arrival rate at the burst processor is equal to the burst arrival rate and the BCP service rate is $1/\upsilon$, for a stable system we need $\upsilon < 1/N(K-k)\lambda$, meaning that the average burst arrival should be limited to :

$$\lambda < 1/N(K-k)\upsilon \qquad (6)$$

A proposed diagram of timings for the STOLAS OBS concept is as follows:



And according to this diagram the timing equations are:

- $T_{DETECT} + T_{PROCESS} + T_{SWITCH} < T_{DELAY} > T_{DETECT} + T_{PROCESS} + T_{SWITCH} + T_{TOLERANCE}$
- $T_{GUARD} > T_{SWITCH} + T_{TOLERANCE}$
- $T_{PACKET} > T_{DETECT} +$
- $T_{PROCESS}$ (sequential detection and processing of labels)

$T_{PACKET} > T_{LABEL}$ (parallel detection and processing of N labels to be supported)

$$T_{ELEC} + T_{PROCESS} < (N - 1) \times (T_{LABEL} + T_{GUARD})$$

The parameters for STOLAS:

L_{bcp} = 14 bytes
r = 155Mb/s
L_b = 64 Kb (maximum size, variable length)
R = 10 Gb/s
F = 2 fibers
W= K = 4 wavelengths
k = 0 or 4 control channels
ρ = 0.85 or 1 (full loaded, worse case to make calculations)
BER (payload) = 10^{-9}
BER (label) = 10^{-12}
η = 99%
T detect ~1.65 μs
T proc ~ 1μs (including new orders to the TWC)
T switch ~ 1 μs (including TWC electronic)
T tol ~ 0.1 μs
T guard > 1.1μs
T_{bcp} = 0.72 μs
T_b = 51.2 μs
T_{elec} = 0.2 μs
T delay (for fixed length)
Worst case = 3.65 μs
Statistical = 2.65 μs (note: supposing υ=Telec and υ*= 0.8)

Best case is not possible to calculate yet, more parameters in the node and network to provide different QoS must be established to an accurate assignment.

Note: Timings for detection, processing and switching are parameters fixed by Lucent Tech. since they are using an implemented system out of the STOLAS scope for calculations (picture from Lucent Tech.)

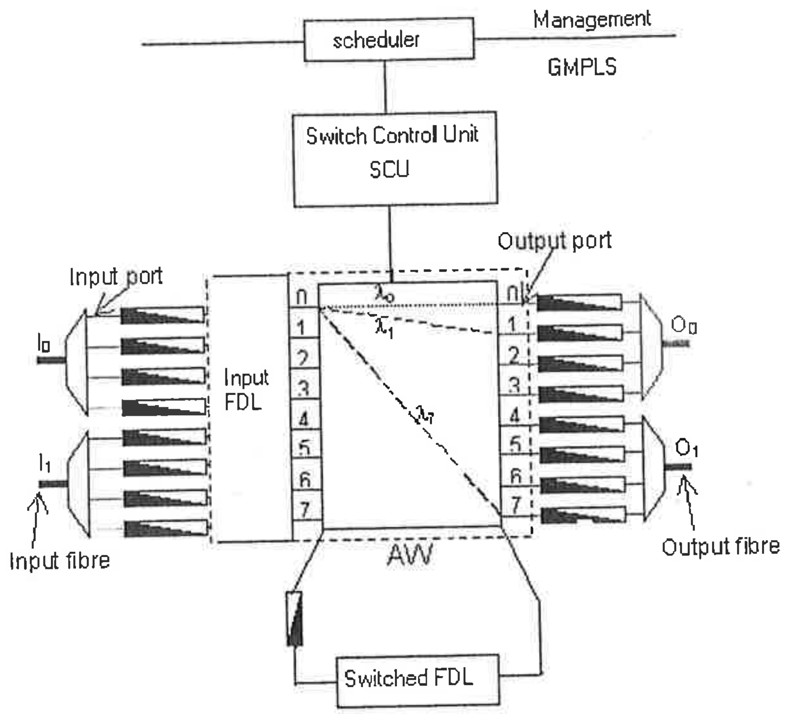


Fig . (13). Node architecture with Input FDLs

1.5. Other issues to take in account

Delay variation between nodes - The time for a packet to travel through a certain distance of the fiber depends on the fiber length, the chromatic dispersion, and the temperature variation. The proposal of using managed inter-node link delays (to make them equal to integer number of time slots) is not yet reasonably applicable with the current technology. When WDM is used the effect of chromatic dispersion has to be taken into consideration. Chromatic dispersion results in different propagation speed for packets transmitted on different wavelengths, therefore different propagation delays occur. For example, with a typical fiber dispersion of 20 ps/nm/km (where ps is the time unit for delay variation, nm the unit for wavelength difference and km the unit for propagation distance), a wavelength variation of 30 nm (consistent with the typical Erbium Doped Fiber Amplifier 1530 – 1560 nm window) and a propagation distance of 100 km, the propagation delay variation would be about 60 ns. If dispersion compensation fibers are used, the above delay variation can be reduced by one order of magnitude. The packet/burst propagation speed also varies with temperature, with a typical figure of 40 ps/0C/km. 100 km of fiber under temperature variation range of 0 – 25 0C means a delay variation of 100 ns. The delay variations mentioned above are relatively slow with respect of time; they can be compensated statically instead of dynamically (on a burst-by-burst basis).

Delay variations inside the nodes - What happens to each packet/burst depends on the switch fabric and contention resolution scheme. In a slotted network that uses fiber delay lines as optical buffers, a packet/burst can take different paths with unequal lengths within the switch fabric. All the considerations given in delay variations in the inter-node links apply here. It is worth mentioning that the fast time jitter induced by dispersion between different wavelengths and unequal optical paths varies from packet/burst to packet/burst at the output of the switch, therefore a fast output synchronization interface might be required. Thermal effects are smaller here because it varies more slowly and can be easily controlled within the node. In an optical switched network each switching node is operating with reference to its own internal clock. As a common practice in SDH/SONET, this clock is derived from a network synchronization signal distributed throughout the network. Phase noise of the oscillators accumulated along the clock distribution and thermal effects on the optical carriers can all contribute to the impairment of the synchronization signal. According to the SDH network synchronization standard, 1 msec is the maximum wander of the local node clock for a time duration larger than 1000 sec. Such slow phase variation has to be taken into consideration.

1.6. Conclusions

1.6.1. Comparison between Generic OBS Node & STOLAS Node concept

- While in Generic OBS there is an offset time between header and payload, in the STOLAS project this concept is avoided. Some implications are related to that decision, use of input FDLs for optical buffering that adds complexity to the control issues, appropriate FDLs lengths, performance of these FDLs, use of fixed or switched ones.

- Since we are going to use BCH channel orthogonal to payload channel: link utilization increases in comparison with other out-of-band signaling systems like SCM. Orthogonal signaling can scale better than SCM signaling. Due to SCM has problems with dispersion and RF processing of subcarriers at higher speeds, this system is not as scalable as STOLAS system. However, at high rates at the STOLAS label swapper the appearance of chirp could be a problem. More study has to be done to verify this issue.

- Link management and protection should be emphasize due to the high level of link utilization.

CHAPTER 2:

MPLS Routing and Signaling for STOLAS

2. MPLS Routing and Signaling. Basis for STOLAS

2.1. Introduction

Throughout this chapter we make an approach to the new protocol proposed for IP-integration-over-WDM and the relationship between MPLS and STOLAS.

The structure of this chapter is as follows: first an introduction to MPLS and the consequences that it supposes. In this first part we make also a comparison between RSVP-TE and CR-LDP signaling protocols for MPLS. The second part is dedicated to MPLambdaS and different control and data planes interworking, and the last part is dedicated to the Link management protocol and the importance of signaling in control management.

The telecom-style network management approaches extended to the IP layer with the cooperation of MPLS Protocol. The appropriate synergy and integration of the two layers is performed with management means capable of performing integrated provisioning of Label Switched paths over Optical paths. If we look into the relationship between MPLS and STOLAS we can see that: STOLAS is MPLS based with the use of OBS and Optical paths (WDM).

Management and adaptation are explained later on in next chapters.

2.2. Definition

Multiprotocol label switching (MPLS) is one of several initiatives to enable delivery on the promise of a converged network. Some of the new protocols' operations span layers 2 and 3. ATM is an example; MPLS is cited in some literature as another one. Strictly speaking, MPLS is a layer 3 protocol in that it does not define the critical function of layer 2 frame delineation. But we will treat it as a multilayer one. By combining the attributes of Layer-2 switching and Layer-3 routing into a single entity, MPLS provides the following benefits:

- enhanced scalability by way of switching technology
- class-of-service (CoS) and quality-of-service (QoS) based services (differentiated services)
- no need for an Internet protocol (IP)–over–asynchronous transfer mode (ATM) overlay model and its associated management overhead
- standards-based solution, promoting interoperability
- enhanced traffic-shaping and engineering capabilities

2.3. Reasons for using label switching

- Speed, delay, and jitter: Label switching is considerably faster than traditional IP forwarding. This speed translates into less delay in transporting traffic through the network. It also translates into less variable delay (jitter).
- Scalability: MPLS allows a large number of IPO addresses to be associated with one or few labels. This approach reduces further the size of address (actually label) tables.
- Resource consumption: label switching networks do not need a lot of the network's resources to execute the control mechanisms to establish label switching paths (LSPs).
- Route control: Most IP-based networks use the concept of destination-based routing, wherein the destination IP address in the IP datagrams determine the route through a network. Destination-routing is not always an efficient operation, and MPLS offers methods to use more efficient route control techniques, thus providing a higher level of service to the user.
- Traffic engineering: As a part of route control, many of the MPLS operations are assigned to allow the network provider to engineer the links and nodes in the network to support different kinds of traffic, as well as constrain the traffic to specific parts of the network.
- Labels and Lambdas: If a label switching is used in optical networks, it is possible to correlate (map) a label or labels to wavelengths, then use a XC switch for forwarding the traffic, thus reducing further the delay and jitter of user payload processing.

2.4. Types of MPLS nodes

In figure (1) the three types of MPLS nodes are shown. They perform the following functions:

- Ingress LSR: Device that operates at the edge of the access network and the MPLS network. Support multiple ports connected to dissimilar networks. Receives native-mode user traffic (for example, IP datagrams), and classifies it into an FEC [5]. It then generates an MPLS header and assigns it an initial label. The IP datagram is encapsulated into the MPLS packet, with the MPLS header attached to the datagram. If it is integrated with a QoS operation the ingress LSR will condition the traffic (such as using different queues for different priorities) in accordance with the DiffServ rules.

- Core/Transit LSR: High speed router that receives the packet and uses the MPLS header/label to make forwarding decisions. It will also perform label swapping (exchange label values). It is not concerned with processing the IP header, only the label header.

- Egress LSR: Performes the decapsulation operation.

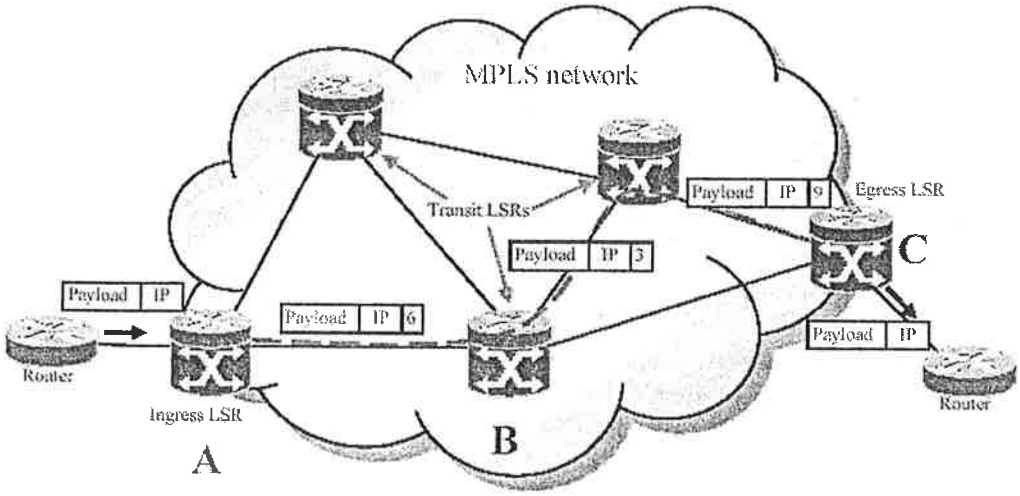


Fig.-1-. Different types of MPLS nodes

2.5. Label Distribution

In order that LSPs can be used, the forwarding tables at each LSR must be populated with the mappings from {incoming interface, label value} to {outgoing interface, label value}. This process is called LSP setup, or Label Distribution.

MPLS does not stipulate a specific label distribution protocol. Since several protocols are currently in operation that can support label distribution, it makes sense to use what is available. Nevertheless, the IETF has developed a specific label distribution protocol to complement MPLS that is called the label distribution protocol (LDP).

Another protocol, the constraint-based LDP (CR-LDP), is an extension to LDP. It allows the network manager to set up explicitly routed Label switched Paths (LSPs). CR-LDP operates independently of any internal gateway protocol (IGP). It is used for delay-sensitive traffic and emulates a circuit-switched network. CR-LDP is also designed to support traffic engineering operations.

RSVP can also be used for label distribution; this extension is called RSVP-TE, it supports label binding and distribution operations. Extensions to border gateway protocols (BGP) are yet another method for advertising and distributing labels. Later we will make a comparison of both protocols to see pros and cons.

Several different approaches to label distribution can be used depending on the requirements of the hardware that forms the MPLS network, and the administrative policies used on the network. The underlying principles are that an LSP is set up either in response to a request from the ingress LSR (downstream-on-demand), or pre-emptively by LSRs in the network, including the egress LSR (downstream unsolicited). It is possible for both to take place at once and for the LSP to meet in the middle.

In all cases, labels are allocated from the downstream direction (where downstream refers to the direction of data flow, and this means that are advertised towards the data source). Some possible options for controlling how LSPs are set up, and the protocols that can be used to achieve them, are described below.

- Hop-by-hop label assignment is the process by which the LSP setup requests are routed according to the next-hop routing towards the destination of the data. LSP setup could be initiated by updates to the routing table, or in response to a new traffic flow. The IETF MPLS Working Group has specified (but not mandated) LDP as a protocol for hop-by-hop label assignment. RSVP and CR-LDP can also be used.

- In Downstream Unsolicited label distribution, the egress LSR distributes the label to be used to reach a particular host. The trigger for this will usually be new routing information received at the egress node. Additionally, if the label distribution method is Ordered Control, each upstream LSR distributes a label further upstream. This effectively builds a tree of LSPs rooted at each egress LSR. LDP is currently the only protocol suitable for this mode of label distribution.

- Once LSPs have been established across the network, they can be used to support new routes as they become available. As the routing protocols (for example BGP) distribute the new routing information upstream, they can also indicate which label (i.e. which LSP) should be used to reach the destinations to which the route refers.

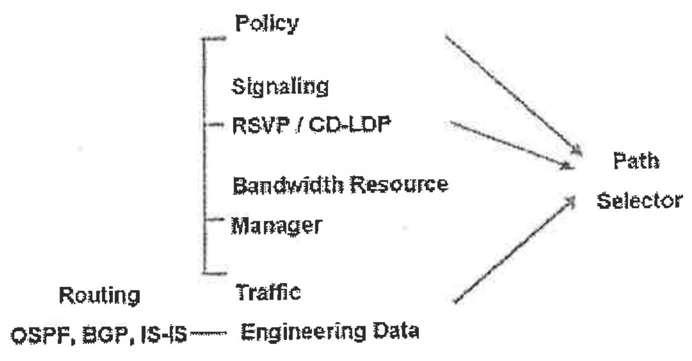


Fig.-2-.MPLS protocols framework for optical WDM Networks

2.6. Comparative study of signaling protocols for GMPLS, RSVP-TE vs. CR-LDP

The choice between RSVP and CR-LDP should be guided by the function of the target system. What LSP setup model will be used?. How stable should the LSP be, do they represent permanent trunks or short-duration calls? How large is the network and how complex is it? Is this a stand-alone network or must the components interwork with other hardware and another networks? A final consideration must be the robustness of the hardware solution. What level of fault tolerance is required? How important is high availability?

Although the two protocols provide a similar level of service, the way they operate is different, and the detailed function they offer is also not consistent. Hardware vendors and network providers need clear information to help them decide which protocol to implement in a Traffic Engineered MPLS network. Each protocol has its champions and detractors , and the specifications are still under development.

Recognizing that the choice of label distribution protocol is crucial for the success of device manufacturers and network designers, we try to explain the similarities and important differences between the two protocols, to help identify which protocol is the right one to use in a particular environment.

Note: all these protocols are still under development, for more information refer to RFC or the IETF drafts.

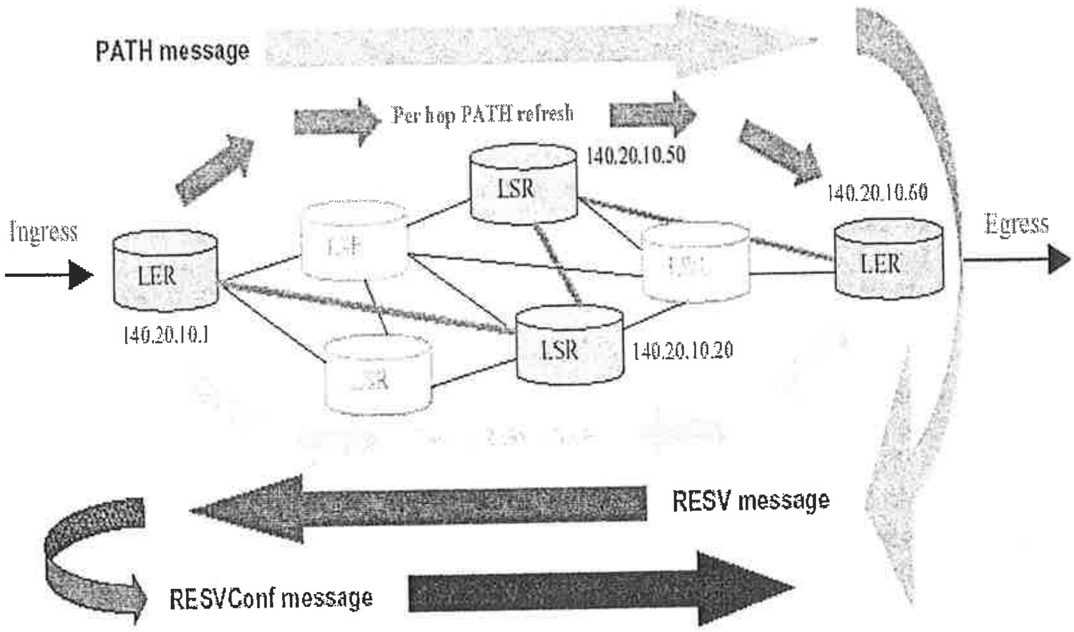
2.6.1. RSVP-TE

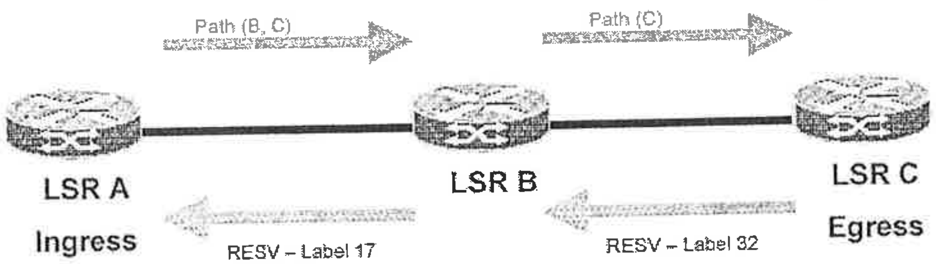
Resource ReSerVation Protocol with extensions of Traffic Engineering.
RSVP defines procedures for signaling QoS requirements and reserving the resources necessary to provide the requested service .

The RSVP-TE extensions enable RSVP to be used for traffic engineering in MPLS environments. The primary extensions add support for assigning MPLS labels and specifying explicit paths as a sequence of loose and strict routes. These extensions are supported by including Label Request and Explicit Route objects in the PATH message. Labels are then subsequently assigned at each node the RESV message traverses.

Other extensions include support for resource affinities, LSP preemption through setup/holding priorities, record route/loop detection options, and rerouting/bandwidth-increase operations.

How it works???





Ingress LER initiates a PATH message to egress LER through each downstream LSR along the path. The traffic parameters required for the session or administrative policies for the network enable LSR A to determine that the route for the new LSP should go through LSR B, which might not be the same as the hop by hop route to LSR C.

LSR B receives the PATH request (message), determines that it is not the egress for this LSP, and forwards the request along the route specified in the request. It modifies the explicit route in the PATH message and passes the message to the next one. Each node receives a PATH message to remember this flow is passing, thus creating a "path-state" or session.

The LSR C determines that it is the egress node for this new LSP, determines from the requested traffic parameters what bandwidth it needs to reserve and allocates the resources required. It selects a label for the new LSP and distributes the label to the LSRB in a RESV message, the egress uses this message to reserve resources with traffic and QoS parameters on each upstream LSR.

Upon receive at the ingress LER a RESV_CONF message is returned to the egress confirming the LSP setup.

After the loose ER-LSP has been established, refresh messages are passing between LER's and LSR's to maintain path and reservation state.

It should be noted that, none of the downstream , upstream or refresh messages between LER and LSR is considered to be reliable, because UDP or raw IP datagrams are used as the communication mechanism. A connection failure will be detected after a neighbor fails to receive a refresh message.

RSVP-TE feature set is robust and provides significant capabilities to provide traffic-engineering services :

- QoS and Traffic parameters- are passed as opaque data to traffic management.
- Failure Notification- upon failure to establish an LSP or loss of an existing one will send failure message, it but relies on timers for refresh message.
- Failure Recovery- "make before break" when rerouting.
- Loop detection- required for loosely routed LSPs only, also supported for re-pathing.
- Multi-protocol support- supports any type of protocol.
- Management- Lsp ID identifies each LSP, thereby allowing ease of management to discrete LSPs.
- Record Route objects- provide the ability to describe the actual setup path to interested parties.
- Path Preemption- the ability to " bump" or discontinue an existing path so that a higher priority tunnel may be established.

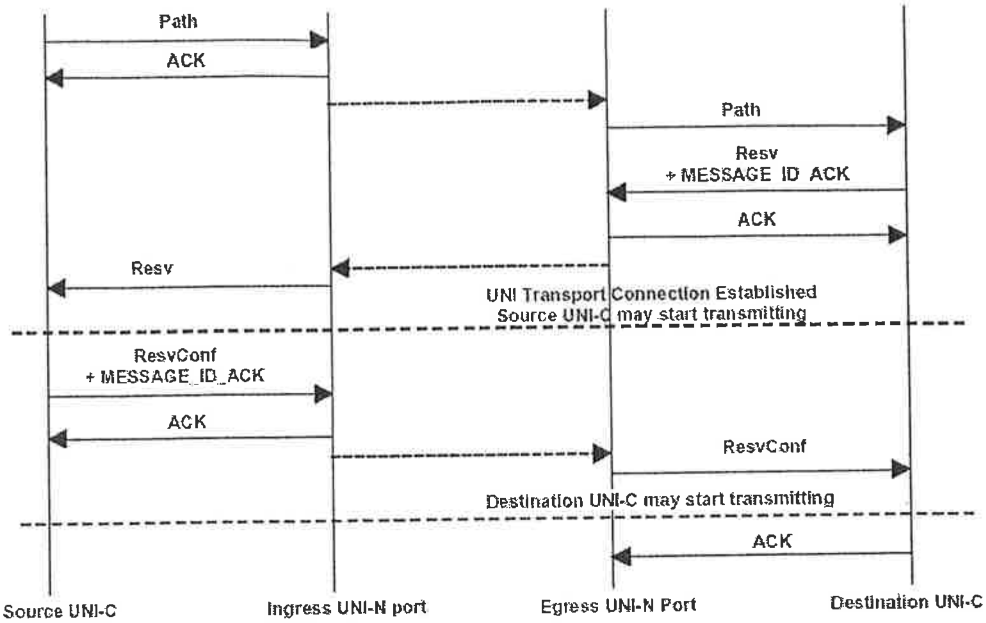


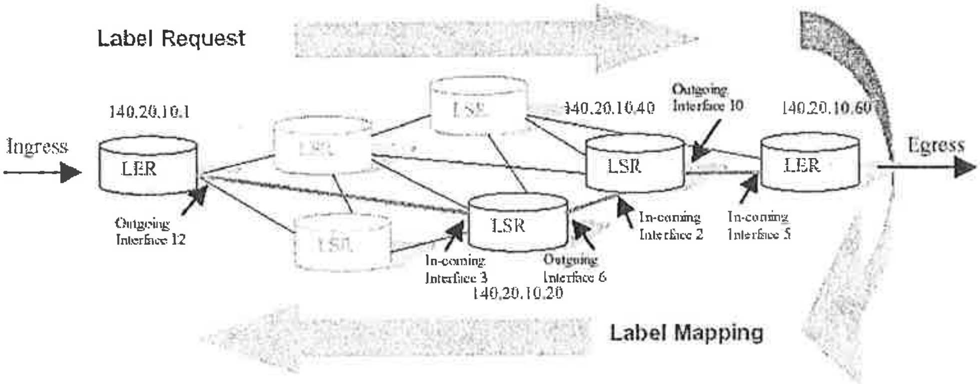
Fig.-3- .A successful connection establishment. Datagram.

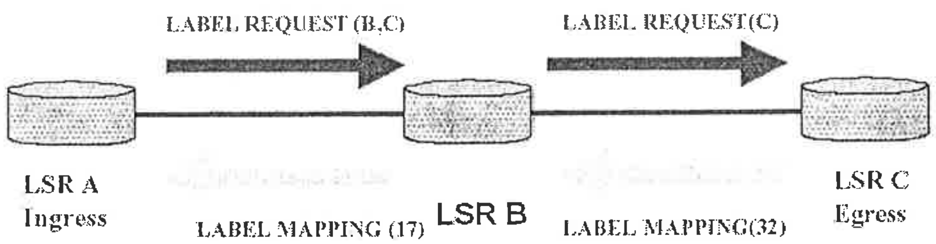
2.6.2. CR-LDP

CR-LDP, constraint based-routed with label distribution protocol, is a simple, scalable, open, non-proprietary, traffic engineering signaling protocol for MPLS networks.

CR-LDP provides mechanisms for establishing explicit routed Label Switched Paths (LSPs) in an MPLS network. These mechanisms are defined as extensions to LDP. Using CR-LDP, resources can also be reserved along a path to guarantee or control QoS for traffic carried on the CR-LSP.

How it works???





The ingress LSR, LSR A, determines that it needs to set up anew LSR to LSR C. The traffic parameters required for the session or administrative policies for the network enable LSR A to determine that the route for the new LSP should go through LSR B, which might not be the same as the hop-by -hop route to LSR C. LSR A builds a LABEL_REQUEST message whit an route of (B,C) and details of the traffic parameters requested for the new route. LSR A reserves the resources it needs for the new LSP, and then forwards the LABEL_REQUEST to LSR B on the TCP session.

LSR B receives the LABEL_REQUEST message, determines that it is not the egress for this LSP, and forwards the request message along the route specified in the message. It reserves the resource for the new LSP, modifies the explicit route in the LABEL_REQUEST message, and passes to LSR C. If necessary, LSR B may reduce the reservation it makes for the new LSP if the appropriate parameters were marked as negotiable in the LABEL_REQUEST.

LSR C determines that it is the egress for this new LSP. It performs any final negotiation on the resource, and makes the reservation for the LSO. It allocates a label to the new LSP and distributes the label to new LSR B in a LABEL_MAPPING message, which contains details of the final traffic parameters reserved for the LSP.

LSR B receives the LABEL_MAPPING and matches it to the original request using the LSP ID contained in both the LABEL_REQUEST and LABEL_MAPPING messages. It finalizes the reservation, allocates a label for the LSP, sets up the forwarding table entry, and passes the new label to LSR A in a LABEL_MAPPING.

The processing at LSR A is similar, but does not have to allocate a label and forward it to an upstream LSR because it is the ingress for the new LSP.

Note: Labels are located from the downstream location, downstream: direction of the data flow, are advertised towards the data source.

Comments:

CR-LDP is built upon LDP, which is already part of MPLS, it is not as mature as RSVP, but it does not require an implementation of an additional protocol and it is only extend if necessary to implement Traffic Engineering.

UDP is used for discovering MPLS peers and TCP is used for control, management, label request and mapping.

If nothing is sent, maintain contact with KeepAlive messages, if it does not arrive (timer) closes the session.

CR-LDP carries the full set of traffic parameters in a Traffic_TLV in the label_request message.

CR-LDP traffic engineering extensions to LDP feature set is comprehensive and is fairly well defined:

- QoS and Taffic Parameters- the ability to define edge rules and per hop behaviors based upon data rates, link bandwidth and weighting given to those parameters.
- Path preemption- the ability to set priorization to allow or not allow preemption by another LSP.

- Path Re-optimization- allows for the capability to re-path loosely routed LSPs based upon traffic pattern changes and includes the option to use route pinning.
- Failure notification- upon a failure to establish an LSP, notification provided on TCP with supporting failure codes.
- Failure recovery- mapping policies to automatic recovery at each device supporting an LSP.
- Multi-Protocol Support- supports any type of protocol.
- Management- LSP ID identifies each LSP, thereby allowing ease of management to discrete LSPs.

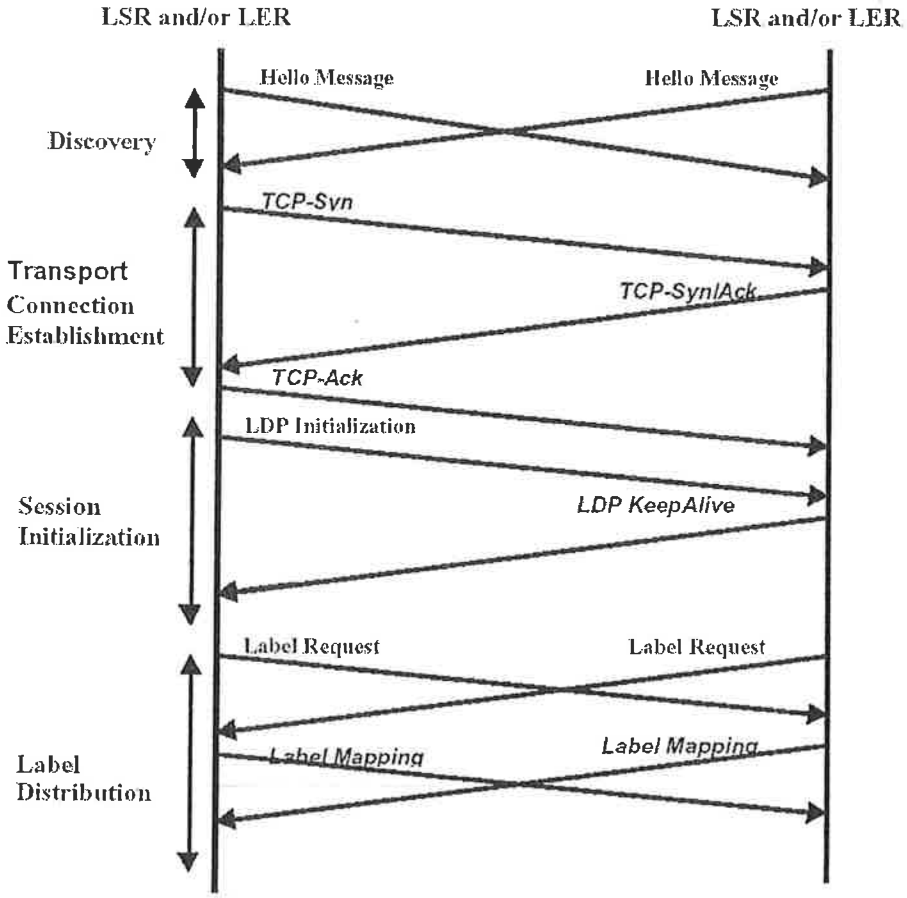


Fig.-4-. LDP initialization

Similarities between TE-RSVP and CR-LDP

Characteristics	CR-LDP	TE-RSVP	Comments
Initiate Setup	Label Request Message	PATH Message containing LABEL_REQUEST object.	
Setup Accomplished	Mapping Message	RESV Message	
Differentiated Services Defined	DIFF-SERV_PSC TLV	DIFFSERV_PSC object	Both contain the DiffServ Code point or DSCP information and included in the setup request message – see row 1
Support for point to multipoint LSPs	?	?	?
Source Route Capability	Carried in Explicit Route List TLV	Carried in EXPLICIT_ROUTE object	Specify route used to setup switched path.

Differences between TE-RSVP and CR-LDP (adapted from dataplane)

Characteristics	CR-LDP	TE-RSVP	Comments
Development Stage	New	Old with extensions being added, support for legacy networks.	RSVP objects being modified to be used in a MPLS environment
Signaling Transport	UDP for discovery, TCP for sessions	Raw IP datagrams or UDP encapsulation for message	Non-deterministic failure detection with RSVP. TCP failure can have

2.6.3. Comparison of principal features of RSVP-TE and CR-LDP

Reliability:

CR-LDP runs over TCP, while RSVP-TE runs over UDP or IP.
RSVP-TE messages can be lost in transit.
RSVP-TE state refresh allows nodes to recover during fail over events in which local state is lost.
Both use Hello messages, but RSVP-TE Hellos only track whether peer nodes have reset.
CR-LDP Hellos identify each peer's label spaces.

Scalability:

CR-LDP requires less overhead per session than RSVP-TE, because of RSVP-TE's use of soft state.
CR-LDP session maintenance is per-path, not per –LSP.
Intermedite LSRs require less state information with cr-LDP.
RSVP-TE consumes more CPU cycles because of the refresh function.

Interoperability:

Both protocols have undergone interoperability tests (IETF is still working on it).
RSVP-TE has been tested for interoperability with 3com, Bay Networks, Cisco, IBM, and Intel devices.
CR-LDP has been tested for interoperability with Nortel Networks, Ericsson, and GDC.
In September of 1999, ITU-T SG13 decided to use CR-LDP to support IP over ATM using MPLS.
Interworking both are still under development.

Traffic Control:

Both protocols support resource reservation, but RSVP-TE reserves resources on *Resv* arrival while CR-LDP does it on *Label_Request* arrival.
Both support path preemption and have the same setup and holding priority metrics.

2.6.4. Conclusion for protocols

Both CR–LDP and TE–RSVP provide very similar functionality for establishing traffic-engineered, labeled switched paths. Each has its strengths and weaknesses. While LDP is the younger of the two protocols, RSVP has been previously deployed and has operational experience. It is true that there have been extensive enhancements to RSVP in order to support the needs of MPLS. As both CR–LDP and TE–RSVP evolve they will offer more and more similar functionality.

Eventually, MPLS traffic engineering should evolve into a single entity that combines the best-of-breed attributes from both TE–RSVP and CR–LDP. In the meantime, any MPLS implementation by original equipment manufacturers (OEMs) developing LER or LSR platforms should consider supporting both TE–RSVP and CR–LDP to ensure interoperability.

It should be noted that the intrinsic value of having an entry point to providing policy-based management to the core is extremely compelling.

2.7. Multiprotocol Lambda Switching (MPλS)

The framework for internetworking optical networks and MPLS is called MPλS. Both technologies have control mechanisms (a control plane) to manage the traffic. These control planes are shown in figure (5).

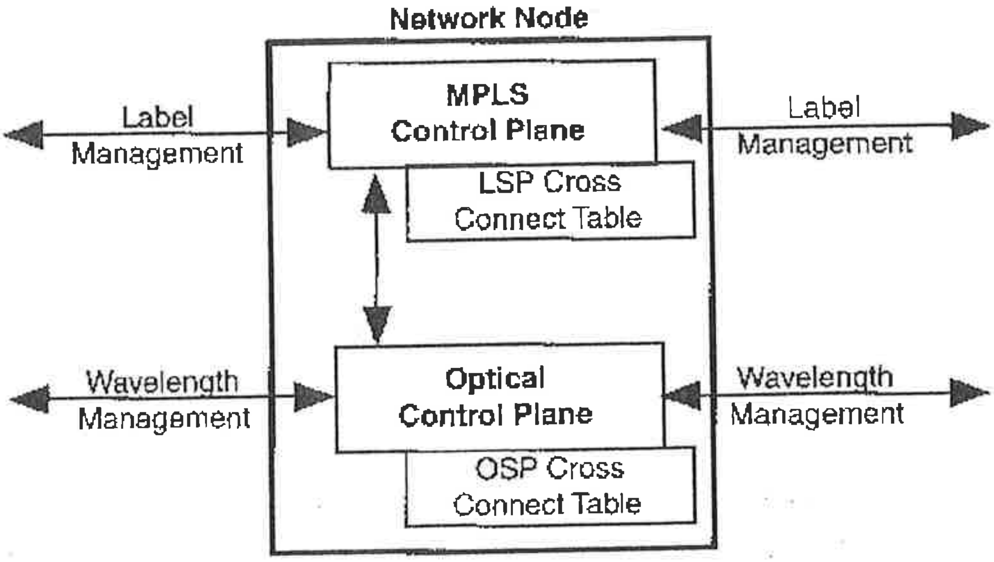


Fig.-5-.The MPLS and optical control planes. Adapted from figure references [2].

The MPLS control plane is concerned with label distribution and binding and end-to-end LSP. The optical control plane is concerned with setting up wavelengths, optical coding schemes, transfer rates, and protection switching options between adjacent nodes.

Some authors [6] disagree with the idea of having different control planes when two technologies must interwork. For example, they propose adapting MPLS TE control plane for OXC. But, due to the arrows between the planes interworking is desirable. We will discuss later about this issue.

2.7.1. OXC and MPLS

An easy way to view the relationship between MPLS and optical networking is to look at the layered model. The optical operations occur in layer 1, and MPLS occur between 2 and 3 (maybe is better to say in a combination of both even that MPLS is a layer 3 protocol).

The data plane of an LSR uses the label swapping to transfer a labeled packet from an input port. The data plane of an OXC uses a switching matrix to connect an optical channel trail from an input port to an output port. An LSR performs label switching by first establishing a relation between an input port and an input label, and output port and output label. Likewise, OXC provisions optical channels establishing relationships between input ports and input wavelengths/channels, and output ports and output wavelengths/channels.

The functions of the control plane include resource discovery, distributed routing control, and connection management. In LSR, it is used to discover, distribute and maintain state information associated to MPLS network, and manage label switched paths (LSPs). In OXC, is used for the same but relative to the OTN, and to establish and maintain optical channel trails under some traffic engineering rules and policies.

A difference between LSRs and OXCs is that, with LSRs, the forwarding information is carried as a part of the labels appended to data packets, and with OXCs, the switching information is implied from the wavelength or optical channel.

How to correlate an MPLS label value with an optical wavelength? at the ingress node the label has been correlated to an appropriate wavelength, that is, an appropriate channel into the network, the transit nodes have been configured to process the wavelength to make the routing decisions. Thus, it is not necessary to know about the MPLS label, as long as all nodes know the relationship of the wavelength that is associated with the label, and its final destination.

We will discuss later in more detail the architecture of an IP and MPLS-based Optical Transport Network.The question is, how the XC is going to be able to:

- Interwork labels with wavelengths.
- How protection switching performed on fibers/wavelengths could be correlated to the labels that are running on the label/wavelength.

The key to interworking effectively the MPLS and optical planes is to carefully distinguish the interaction between not only the control planes, but also between data and control planes.

2.7.2. Control planes

A control plane is a set of software and/or hardware in a node that is used to control several vital operations of the network, such as bandwidth allocations, route discovery, and error recovery. Obviously, the control plane is important.

The control messages are exchangeg between nodes to perform a wide variety of operations. For optical networks, some of the more important tasks for the control plane include:

- Exchange status messages, such alarms and diagnostics.
- Providing timing messages to keep nodes' clocks in synchronization with each other.
- Using messages to download information on which wavelengths will be used between two nodes.
- Building forwarding cross-connect tables to allow the data plane to relay traffic from input port to output port.

The approach for IP/WDM transport networks is to define a separate, dedicated control plane that can operate in any of the following fashions:

- The control plane messages can be exchanged on a separate physical fiber link from those of the user traffic.
- Alternatively, the control messages can be sent on the same fiber link used for the user traffic, as well as on the same wavelength (STOLAS approach).
- Control messages can also be sent on a separate wavelength on the same fiber that is transporting user traffic on the other wavelengths of that fiber.
- Control messages can be sent and received on separate nodes from those that carry the data traffic (complex).

2.7.3. The Optical Control and Data Planes

Figure (6) shows the optical and data planes for IP/WDM transport network. The control plane can be executed with GMPLS or LMP or a combination of both. Whatever the implementation may be, the optical control plane is used to coordinate the use of wavelengths between adjacent optical nodes, as well to insure the nodes are up and running.

Thereafter, the control plane is invoked only for ongoing management operations, diagnostics, recovery, and so on.

The requirements for the optical control plane have established that this control plane must be able to support the following types of connections:

- A permanent optical channel set up by the network management system via network management protocols.
- A soft permanent optical channel set up by the network management system, using network-generated signaling and routing protocols to establish connections.
- A switched optical channel, which can be set up by the customer on demand using signaling and routing protocols.

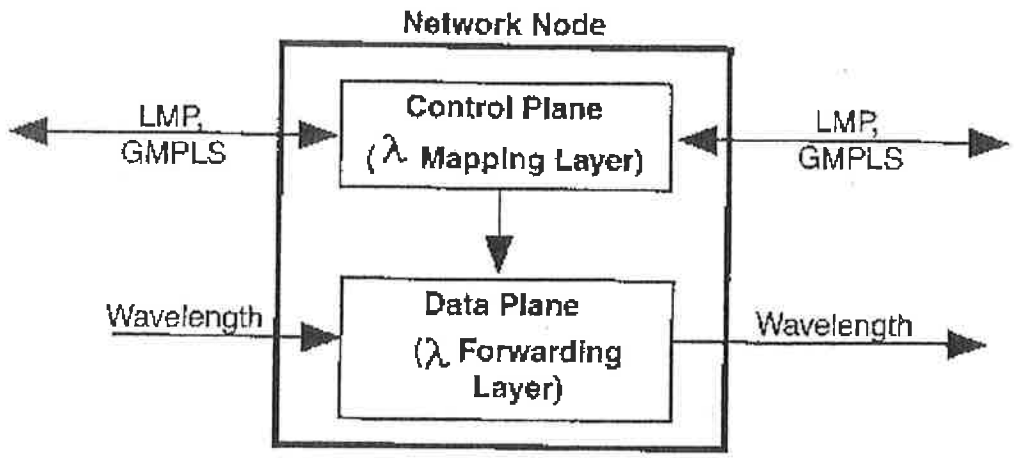


Fig.-6-.The Optical Control and Data Plane

2.7.4. Interworking the three control planes

The separate operations of the IP, MPLS, and Optical Control planes should be coordinated in order to take advantage of:

- The route discovery capabilities of the IP control plane.
- The traffic engineering capabilities of the MPLS control plane.
- The forwarding (switching) speed of the optical data plane.

Figure (7) illustrates how this interworking can be accomplished.

So, the following three events must take place to exploit the powerful capabilities of the three control planes:

- The IP routing protocols advertise and discover addresses as well as the routes to the nodes that are identified by the addresses.
- The MPLS label distribution protocols distribute labels associated with the IP addresses, maps addresses to certain labels.
- The MPLS labels can be mapped to specific wavelengths between adjacent optical nodes.

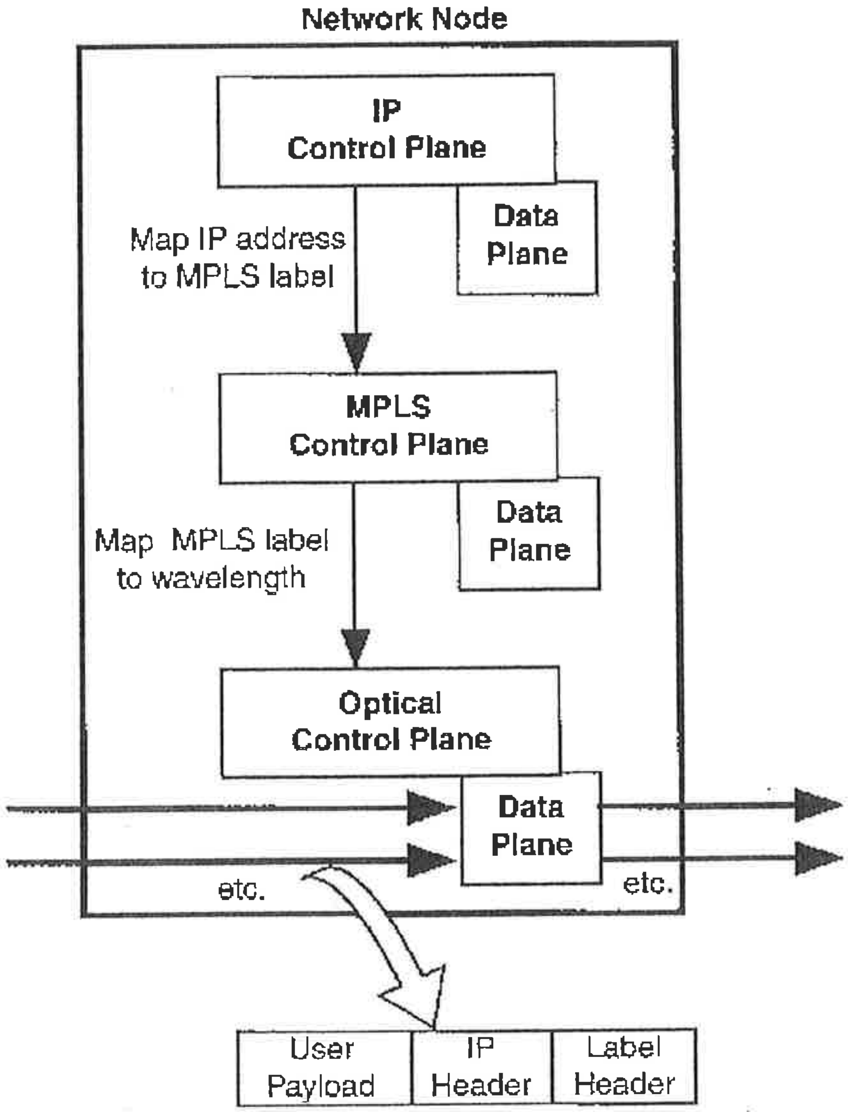


Fig.-7-.Interworking the three control planes. Adapted from figure references [2].

2.7.5. Interconnections for IP over Optical

Given that IP/MPLS over optical can use the Domain or Unified models, the transport of the IP datagrams over an optical network can occur through three kinds of interconnections:

- Peer: Under the peer model, the IP/MPLS layers act as peers of the optical transport network, so that a single control plane runs over both the IP/MPLS and the optical domains. When there is a single optical network involved, presumably a common routing protocol such OSPF or IS-IS, with appropriate extensions, can be used to distribute topology information over the integrated network.

- Overlay: Under the Overlay model, the IP/MPLS routing, topology distribution, and signaling protocols are independent of the routing, topology distribution and signaling protocols at the optical layer, those are defined for the optical domain. Interactions between routing and signaling are accomplished through UNI-defined procedures.

- Augmented: Under augmented model, there are actually separate routing instances in the IP and optical domain, but information from one routing instance is passed through the other routing instance.

2.8. Link Manager Protocol (LMP)

Future networks will consist of photonic switches (PXCs), optical crossconnects (OXCs), routers, switches, DWDM systems, and add-drop multiplexors (ADMs) that use the Generalized MPLS (GMPLS) control plane to dynamically provision resources and to provide network survivability using protection and restoration techniques. A pair of nodes (e.g., two PXCs) may be connected by thousands of fibers, and each fiber may be used to transmit multiple wavelengths if DWDM is used. Furthermore, multiple fibers and/or multiple wavelengths may be combined into a single traffic-engineering (TE) link for routing purposes. To enable communication between nodes for routing, signaling, and link management, a control channel must be established between the node pair. Following references [7] specifies a link management protocol (LMP) that runs between neighboring nodes and is used to manage TE links.

LMP can be used for any type of node, enhancing the functionality of traditional DXCs and routers, while enabling PXCs and DWDMs to intelligently interoperate in heterogeneous optical networks.

In GMPLS, the control channel between two adjacent nodes is no longer required to use the same physical medium as the data-bearing links between those nodes. For example, a control channel could use a separate wavelength or fiber links, and vice-versa. Therefore, a clean separation between the fate of the control channel and data-bearing links must be made. Furthermore, new mechanisms must be developed to manage the data-bearing links, both in terms of link provisioning and fault localization.

A data-bearing link may be either a "port" or a "component link" depending on its multiplexing capability; component links are multiplex capable, whereas ports are not multiplex capable. This distinction is important since the management of such links (including, for example, resource allocation, label assignment, and their physical verification) is different based on their multiplexing capability.

If multiple interfaces are grouped together into a single TE link using link bundling [8], then the link resources must be identified using three levels: TE link Id, component interface Id, and timeslot label.

Resource allocation happens at the lowest level (timeslots), but physical connectivity happens at the component link level. If multiple interfaces are once again grouped together into a single TE link, then link bundling [8] is not required and only two levels of identification are required: TE link Id and port Id. Both resource allocation and physical connectivity happen at the lowest level (i.e. port level). LMP is designed to support aggregation of one or more data-bearing links into a TE link (either ports into TE links, or component links into TE links).

2.8.1. LMP Overview

LMP runs between a pair of nodes and includes a core set of functions; two additional tools are defined in some references [7] to extend the functionality of LMP and are optional. The core function set includes control channel management and link property correlation.

Control channel management is used to establish and maintain control channel connectivity between neighboring nodes. This is done using lightweight Hello messages that act as a fast keep-alive mechanism between the nodes. Link property correlation consists of a LinkSummary message exchange that is used to synchronize the link properties (e.g., local/remote Interface ID mappings) between the adjacent nodes.

LMP requires that a pair of nodes have at least one active bi-directional control channel between them. This control channel may be implemented using two uni-directional control channels that are coupled together using the LMP Hello messages. All LMP messages are IP encoded (except, in some cases, the Test Message which may be limited by the transport mechanism for in-band messaging).

In LMP, multiple control channels may be active simultaneously between a pair of nodes. Each control channel must individually negotiate the control channel parameters, and each active control channel must exchange LMP hello packets to maintain LMP connectivity. If a group of control channels share a common node pair and support the same LMP capabilities, then LMP control messages may be transmitted over any of the active control channels of that group without coordination between the local and remote nodes.

LMP also allows secondary (or backup) control channels to be defined. For example, data-bearing may be used as backup control channels provided control channel traffic has preemptive priority over the data traffic on the link. Secondary control channels only become active control channels when the switchover is complete and they inherit the configuration properties of the primary control channel that is being switched over to it.

Two additional tools are defined that extend the functionality of LMP: link connectivity verification and fault management. These tools are particularly useful when the control channel is transmitted out-of-band from the data-bearing links.

Link connectivity verification is used to verify the physical connectivity between the nodes and exchange the Interface Ids, these Ids are used in GMPLS signaling. The procedure uses in-band Test messages that are sent over the data-bearing links and TestStatus messages that are transmitted over the control channel. The fault management scheme uses ChannelActive and ChannelFail message exchanges between a pair of nodes to localize failures in both opaque and transparent networks, independent of the encoding scheme used for the data. As a result, both local span and end-to-end path protection/restoration procedures can be initiated.

The LMP fault management procedure is based on two message exchanges: ChannelActive and ChannelFail. The ChannelActive message is used to indicate that one or more data-bearing channels are now carrying user data. This is particularly useful for detecting unidirectional channel failures in the transparent case.Receipt of a ChannelActive message must be acknowledged with a ChannelActiveAck message.

The ChannelFail message is used to indicate that one or more active data channels or an entire TE link have failed. Receipt of a ChannelFail message must be acknowledged with either a ChannelFailNack or ChannelFailAck message, depending on if the channel failure is clear or not in the adjacent node.

2.8.2. Control channel management

To initiate an LMP session between two nodes, a bi-directional control channel must be established. The control channel can be used to exchange MPLS control-plane information such as link provisioning and fault isolation information (implemented using a messaging protocol such as LMP), path management and label distribution information (implemented using a signaling protocol such as RSVP-TE or CR-LDP), and network topology and state distribution information (implemented using traffic engineering extensions of protocols such as OSPF and IS-IS). For the purposes of LMP, we do not specify the exact implementation of the control channel; it could be, for example, a separate wavelength or fiber. Our control channel is defined in following chapters.

Furthermore, the control channel messages should be defined to be IP encoded. This allows the control channel implementation to encompass both in-band and out-of-band mechanisms; including the case where the control channel messages are transmitted separately from the associated data link(s).

Note that for in-band signaling, a control channel could be allocated to a data-bearing link; however, this is not true when the control channel is transmitted separately from the data links.

The control channels may also be used for transmitting and receiving signaling and routing messages. Each LMP control channel must individually negotiate the control channel parameters, and each active control channel must exchange LMP Hello packets to maintain LMP connectivity. For LMP, it is essential that at least one control channel is always available.

CHAPTER 3:

Network Management for the STOLAS project: System Analysis and Architecture.

3. NETWORK MANAGEMENT FOR THE STOLAS PROJECT: SYSTEM ANALISYS AND ARCHITECTURE

Note: This chapter is an adaptation of the WINMAN project to the STOLAS project, any information included about WINMAN is EXTREMALY CONFIDENTIAL, request WINMAN Consortium to use any information.

3.1. Introduction

The management functions specifically needed for optical-label-switched networks will be defined. These include the elements management of the advanced optical network elements (i.c. the label swapping wavelength converter and phase modulator circuits, the fast widely tunable laser diodes, the 2R regenerators..), the optical signal monitoring, the interaction with the electrical IP routers, the control of the optical routing, and the control of the labeling functions. Also the requirements on a network management and control communication channel will be identified.

MPLS-controlled optical label switching (OLS) related management requirements are related to both element management and network management. The activities will focus on:

- The definition of the management architecture (logical and physical) for the Element Management System (EMS) and the Network Management System (NMS).
- The definition of the configuration, fault and performance requirements for EMS and NMS.
- The definition of the functional architecture and atomic functions necessary for EMS and NMS.
- Identification of the methodology useful to describe the information model for managing the network element through the EMS, if necessary.
- Description of the basic blocks of the OLS information models useful to satisfy the requirements defined for each functional management area.

The development of a network management system is not a project objective, within STOLAS it is mainly hardware-oriented and focuses on a new optical labeling concept. Nevertheless, to be able to deliver a proof-of-concept and the mapping of electrical labels on the optical stack in a field trial, a limited network management system is required, taking into consideration that without configuration management is difficult to set up paths between the commercial IP routers.

The main objective of the WINMAN-over-STOLAS solution is to provide an integrated network management system, which is capable of providing managed end-to-end IP connectivity services derived from Service Level Agreements (SLA's). In other words, WINMAN will test in field a Network Management System, capable of performing integrated provisioning of IP/MPLS Label Switched Paths over optical paths, as well as integrated multi-layer fault and performance management. WINMAN will capture the requirements, define and specify and open, distributed, and scalable management architecture. Will support multi-vendors, multi-technology environments and evolution scenarios. Therefore, it is a good manager for STOLAS project, moreover WINMAN is a project from IST in which Lucent is involved.

Note that the term domain in this paper, does not refer to administrative domains, but to technology domains , like IP and WDM.

3.2. The Control Plane vs. Management Plane approaches in the IP/WDM Integration

The term control plane is used in the literature to refer to the set of real-time mechanisms and algorithms needed for call or connection control. It deals mainly with the signaling to set-up, supervise and release calls and connections [9]. Although, a detailed de-composition of the control plane and a description of each component is not our purpose, we can safely assume that the signaling protocol for connection set-up and the routing protocols supporting network discovery are the most significant features of the control plane. In that respect, it is significantly easier to follow all the recent advances and proposals about the integration of the IP-electrical world and the WDM-optical world control planes.

Many standardization bodies as well as international for a have addressed the issue of integrating the control lane of the IP MPLS-capable and the WDM network elements (NE). The IETF has proposed the MPLambdaS framework [6], which extends the MPLS ideas to the optical domain, allowing the re-usability of the existing Internet protocols with the appropriate extensions. The OSPF as well as the IS-IS routing protocols have been enhanced to disseminate information relevant to the optical domain ([10], [11]). On the other hand, the OIF and he ODSI forums have made one step towards the definition of the appropriate signaling messages ([12], [13]), which will allow the dynamic set-up of end-to-end connections between IP routers spanning the optical network. Furthermore, [14] and [15] present the mapping between the signaling messages defined in [12] and existing IP/WDM signaling protocols, namely RSVP-TE and CR-LDP.

The latter is an important issue, since the automatic provisioning of end-to-end connection between two IP routers by means of signaling permits to Network Operators

(NO) and Network Service Providers (NSPs) to provide rapidly the requested IP connectivity to customers. This is the main requirement of the NO and NSP in order to cope with the changes in the service-chain and to be able to provide in a cost efficient way the requested services.

On the other hand, the term management plane is used in the literature to refer to the set of near real-time management mechanisms and algorithms related to the system as a whole and to the OAM [9]. It deals mainly with the procedures related to five functional areas, namely Configuration, Fault, Accounting, Performance and Security (FCAPS). The three main management functions of the FCAPS which have similar functionality with the control plane functions, thus competing each other, are path provisioning with routing and QoS support, in the Configuration Management Area, and automatic recovery of failures or performance degradations in the Fault and Performance Management Areas. Network Management functionality mainly exists independently for the IP-electrical world and the WDM-Optical world rather than for the integration of the two worlds. In this direction we will first deal with the advances and proposals for IP layer and WDM layer separately and then we will provide the limited efforts on their integration.

3.2.1. The Control & Management approach (best of two worlds)

It seems clear that comparing the two previous approaches we see that one of them is clearly better in the first phase and the other one is clearly better in the second phase.

What we try to do in this third approach is to define a mixed solution that combines the best of the previous approaches.

The design phase is performed by the management system. The management system has a clear picture of the network and it has access to information about network topology changes faster than any other network element. Because it is the management system that makes the design of the IP paths, it is the management system that decides what resources should be reserved and when. So the management system knows where and what resources are reserved on the network. The information about resources is always up to date on the management system.

By making the design in the management system we maintain the advantages of freeing the edge routers from this high CPU consuming process and having the best-informed entity making the design of the IP paths.

The implementation phase is performed by the control plane, but triggered by the management plane. After the management system has designed the path it triggers the signaling protocol.

The signalling protocol must support explicit routing and must be able of performing resource reservation. The information transmitted by the signalling protocols is given by the management system when it triggers the signaling.

This way we have paths designed by the management system implemented as fast as the ones designed by the control plane (and implemented much faster than the ones implemented by the management system).

The major problem in this approach is the same that we had in the implementation phase of the control plane approach. All the network elements must use compatible signalling protocol implementation. This is very hard to provide in multi-vendors networks nowadays.

We think that with this approach we are using the best of the previous two. We can have IP paths established by the best-informed entity (management system) and implemented by the fastest mechanism (signaling protocols of the control plane).

3.3. High level architecture

A two-layer management architecture has been conceived for the system (multi-layered network), the first layer contains technology dependent managers, that is one for the IP network and another manager for the WDM optical network. These two management system will interact with existing element management systems or even subnetworks called southbound WINMAN (Fig.1), can be deployed as a stand-alone management system capable of managing a WDM network by its own.

On top of this layer there is a second one, called INMS (Integrated Network Management System), whose purpose is the integration of the above mentioned management systems to make the best use of the underlying transport technologies.

Goals in WINMAN architecture:

Open, flexible, modular, scalable, distributed (transparent), independent, minimization of interfaces.

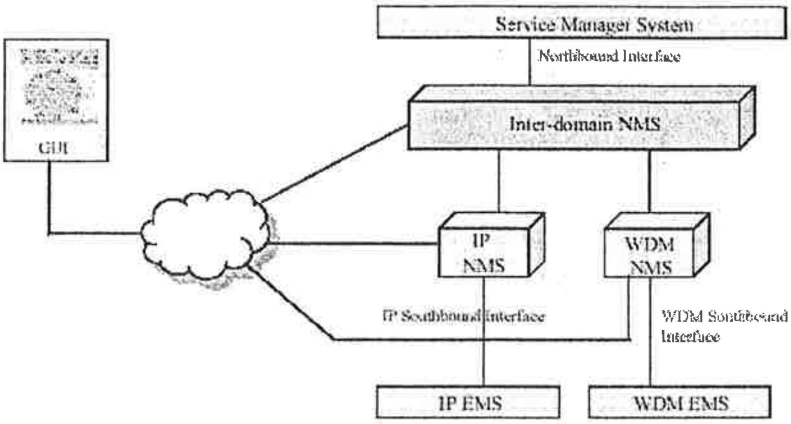


Fig 1. Overview of WINMAN high level architecture.

3.4. Systems description

3.4.1. Integrated Network Management System (IMNS)

Through a properly abstracted information model, the INMS is able to use the domain NMSs for the provisioning and monitoring of end-to-end connections without any knowledge about network details except for the layered structure.

The main functions that the IMNS performs are in the area of configuration management, fault management and performance management (end-to-end).

- Serving of requests from the Service Management Systems.
- Configuration of the end-to-end connections and services, designing the route as an ordered set of subnetwork connections.
- Presentation of an end-to-end view to the operator, resources.
- Maintenance of an inventory of resources.
- Co-ordination of the restoration process and Qos.
- Provides updates.
- Inter domain alarm correlation.
- Analysis of data and assessment of the QoS delivered.

3.4.2. WDM-NMS

The WDM Network Management System carries out the management of the WDM transport network layer .

The main functions are:

- Provision of optical paths, including protection.
- Wavelength routing, transmission and implementation of the optical path in the network elements.
- Discovery of network resources in terms of network elements (ports, fiber connections, wavelengths per fiber, transport capacity per wavelength…).
- Co-ordination.
- Gathering of data about the performance levels (power, BER, Eb/No….).
- Control of alarms.

3.4.3. IP-NMS

The IP policy-based Network Management System carries out the management of the IP network layer , adapted to the connection oriented approach on which WINMAN focuses.

The main functions are:

- Design of connections inside its own subnetworks.
- Establishment and maintenance of end-to-end QoS connectivity services, including the setting up of MPLS LSPs.
- Discovery of network topology.
- Gathering of performance information (link load, network congestion, effective throughput,…).

3.4.4. GUI (Graphic User Interface)

The graphical user interface interacts with the three previously described systems.It supports the access to information and functions of the INMS, WDM-NMS, IP-NMS. Should be as thin as possible in order to be WEB enabled.

3.5. Domain model and Network views

WINMAN will focus on a connection-oriented IP domain model, corresponding to MPLS technology. We identify 2 layers, MPLS and WDM, so we consider only MPLS entities and we replace everything related to IP with MPLS.

We suppose MPLS domain starts at the Customer Edge IP equipment, so all the traffic that enters in WINMAN is marked as MPLS. However, the MPLS domain should be further developed since this domain is not completed established and is still under updating.

3.5.1. Common Entities

- Connectivity Service (CS):
  Is the basic service delivered by the MPLS or WDM network and managed by the Management system. The service is delivered either by the MPLS network meeting some QoS goals resembling a connection-oriented circuit or by the WDM network.

- Element Management System (EMS):
  The EMS represents the abstraction of the sub-network(s) managed by the EMS managed domain and the element management system itself.

  - A termination point (TP) shall be a logical abstraction of an end –point (actual or potential) of a topological (physical) link, or a subnetwork connection. A TP is contained within a managed element.

- Physical Termination Point (PTP):
  A termination point that is an actual or potential endpoint of a topological (physical) link shall be abstracted as a physical termination point (PTP). Essentially, is the representation of a physical port.

- Connection termination Point;
  Actual or potential end point of a subnetwork connection.

- Topological Link:
  A Topological Link is a physical link between two PTPs. A topological link has a name and references to the two PTPs. A topological link reported bya n ERMS to NMS will be between two managed elements (MEs) managed by the same EMS.

- Link connection:
  Represents the transparent capacity of transfer information characterized by a given signal identification between two fixed points.

- Managed element:
  Is an abstract class used to represent Network Elements visible across the interfaces.

3.5.2. MPLS entities

Some of the most important are:

- MPLS Connectivity Service Connection (MPLS-CS): CS is the basic service delivered by the MPLS
- IP Element Management System (IP-EMS), manages both IP and MPLS protocols.
- MPLS Connection Termination Point (CTP), At the MPLS layer it corresponds to a specific label and port.
- MPLS Cross-Connect table, is a set of associations (FEC).
- MPLS Trail or Label Switched Path, transport entity.
- MPLS Physical Termination Point, representation of a physical port, belonging to a Label Switch Router (LSR).

3.5.3. WDM entities

- Optical Connectivity Service Connection (OCS), provides OCh-SNCs.
- WDM Element Manager System (WDM-EMS).
- Optical Topological Link.
- WDM Physical Termination point (WDM PTP), representation of a physical port, belonging to an Optical Managed Element.
- Optical Connection Point (OCP), in an optical network it is a representation of a specific wavelength of an optical port.
- Optical Managed Element (OME),used to represent network elements, we should include: OADM, OXC, Optical Amplifiers, Terminal Multiplexer.

3.5.4. External and Internal network view

(modules referred to next figure 2.)

(1) Module (INMS):This external interface should be as abstract as possible. The INMS need to show to SMS only the edge Physical Term. Points, which are the ones interconnecting the provider and the customer.
PTPs are always MPLS PTPs and not WDM. CTPs are inside PTPs.

(2) Module (IP NMS): PTP are edge MPLS-PTPs which were also shown towards the SMS and the MPLS over WDM PTPs, which interconnect to the WDM network. Between CTP exist MPLS subnet. connections (MPLS – SNC). Note that the MPLS CTP interconnecting to the WDM network are marked with ('), while the corresponding CTPs in WDM will be later marked with (").

3.5.5. Multiple views in a single picture

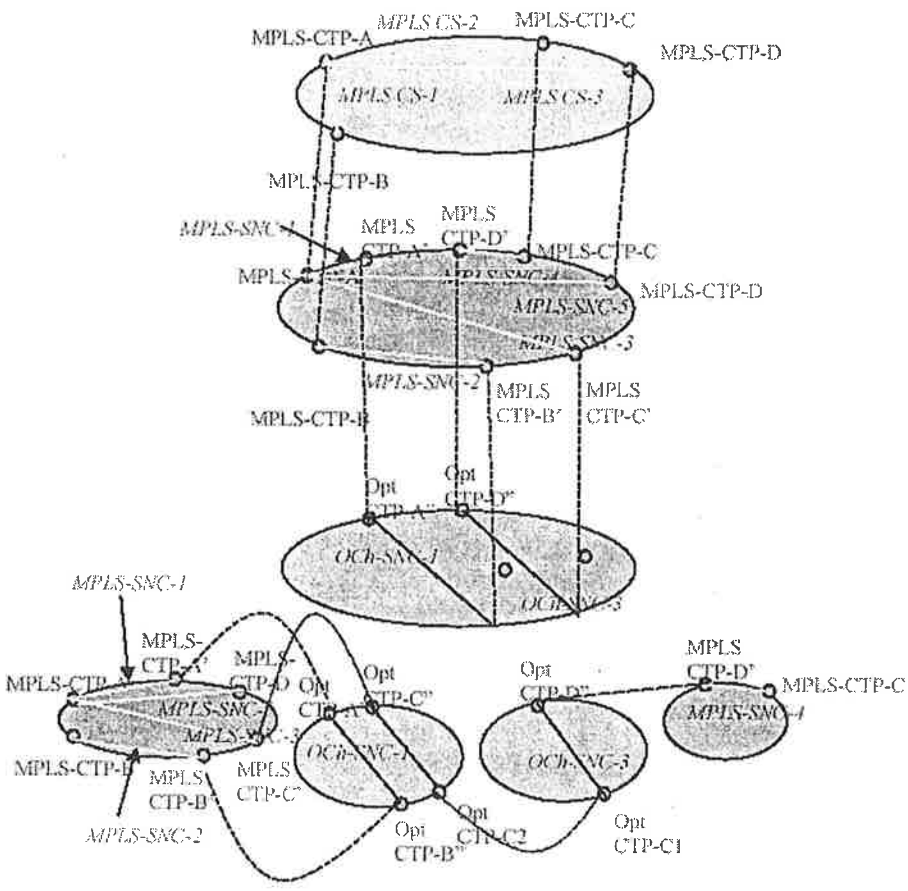


Fig 2. Multiple SMS-INMS, WDM NMS-INMS, and IP NMS-IMNS. Adapted from figure references [4].

The previous picture depicts the multiple views between SMS-INMS, WDM NMS-INMS, IP NMS-INMS and IP EMS-IP NMS, WDM EMS-WDM NMS in a single picture.

Analysing for example the MPLS-CS –1, it is composed by:

MPLS CS-1= MPLS-SNC-1+LC+Och-SNC-1+LC+MPLS-SNC-2

- The MPLS-SNC1 between MPLS CTP-A and CTP-A'
- The optical link connection between MPLS CTP-A' and optical CTP-A"
- The Och-SNC1 between MPLS CTP-A" and optical CTP-B"

- The optical link connection between optical-CTP-B" and MPLS-CTP-B' and finally
- The MPLS-SNC2 between the MPLS-CTB' and MPLS-CTP-B.

We can see also:

MPLS CS-2= MPLS-SNC-3+LC+Och-SNC-#+LC+MPLS-SNC-4
MPLS CS-3= MPLS-SNC-5 (pure MPLS).

In order to interconnect with a real example and test-bed configuration, next figure (3) shows the Element Management Layer of the MPLS and Optical Sub-networks together with the corresponding technology equipment.

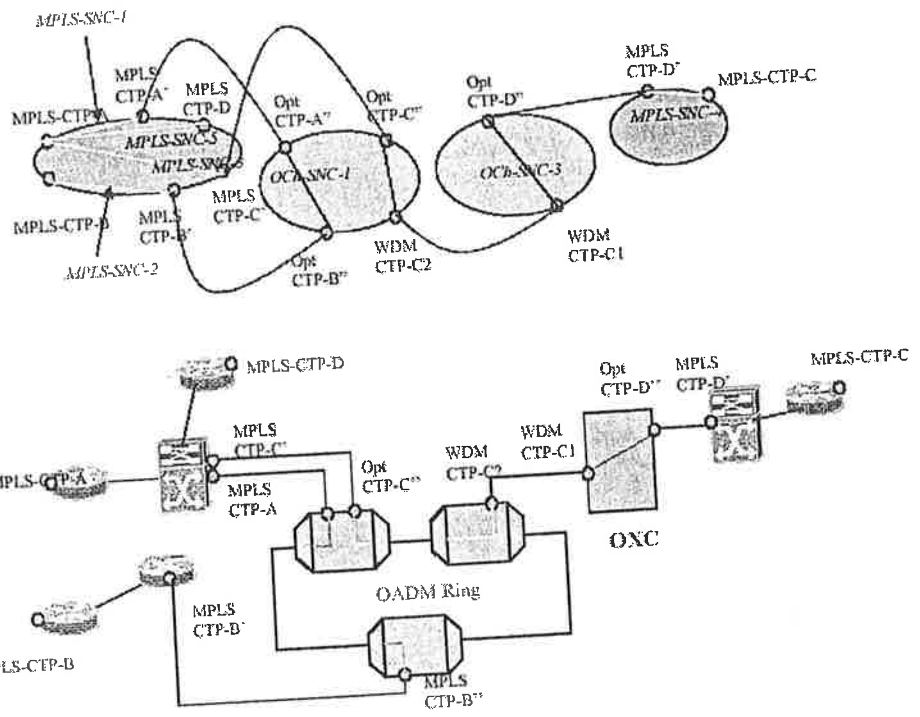


Fig. 3. Element Management Layer of the MPLS and Optical Sub-networks together with the corresponding technology equipment. Adapted from figure references [4].

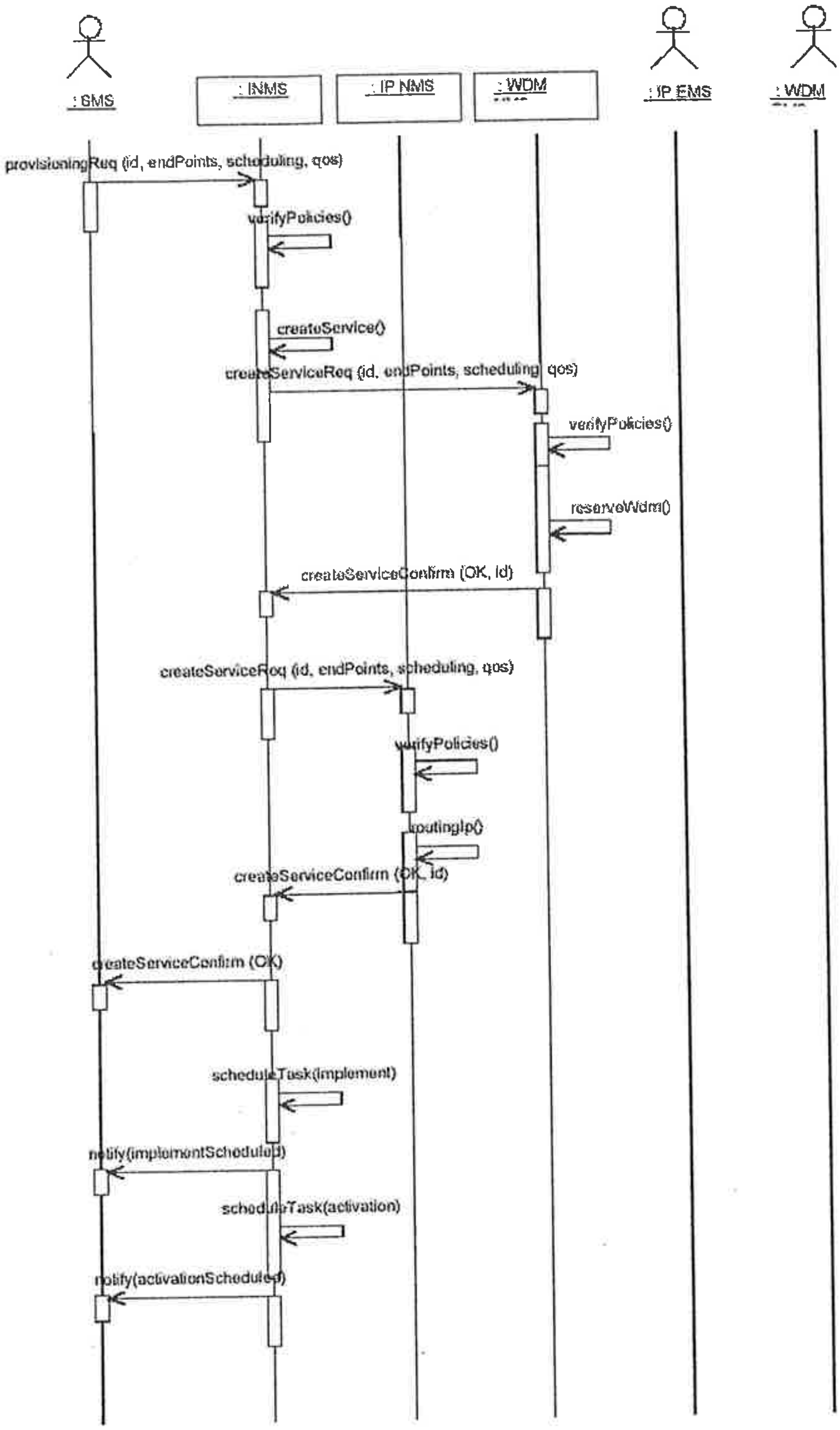
The network that is being displayed is composed by the 2 Optical Sub-networks and 2 MPLS sub-networks. The first optical sub-network is a ring of Optical Add-Drop-Multiplexers and the second one is an Optical Cross-Connect. In the first case, OChSNC are a concatenation of optical link connections (inside fibers) and cross-connects inside the OADM, while in the second case the OChSNC is generated only in the cross-connect inside the OXC. On the other hand the MPLS network is composed by multiple Label Switch Routers, either playing the role of ingress, core or egress LSRs.

Note that we have selected 3 MPLS CS connections, 2 of them spanning both domains, while 1 of them is a pure MPLS one (between MPLS-CTP-A and MPLS-CTP-D).

3.6. System scenarios

In this section is intended to show the responsibilities of each system and the main interactions among them.

One of these is, for example , provision of ICS involving creation of new lightpath. Corresponds with the request received from the Service Management System for the provisioning of an IP connectivity service with NLA parameters and scheduling data. Includes the following steps:


: SMS
: INMS
: IP NMS
: WDM
: IP EMS
: WDM
provisioningReq (id, endPoints, scheduling, qos)
verifyPolicies()
createService()
createServiceReq (id, endPoints, scheduling, qos)
verifyPolicies()
reserveWdm()
createServiceConfirm (OK, id)
createServiceReq (id, endPoints, scheduling, qos)
verifyPolicies()
routingIp()
createServiceConfirm (OK, id)
createServiceConfirm (OK)
scheduleTask(implement)
notify(implementScheduled)
scheduleTask(activation)
notify(activationScheduled)

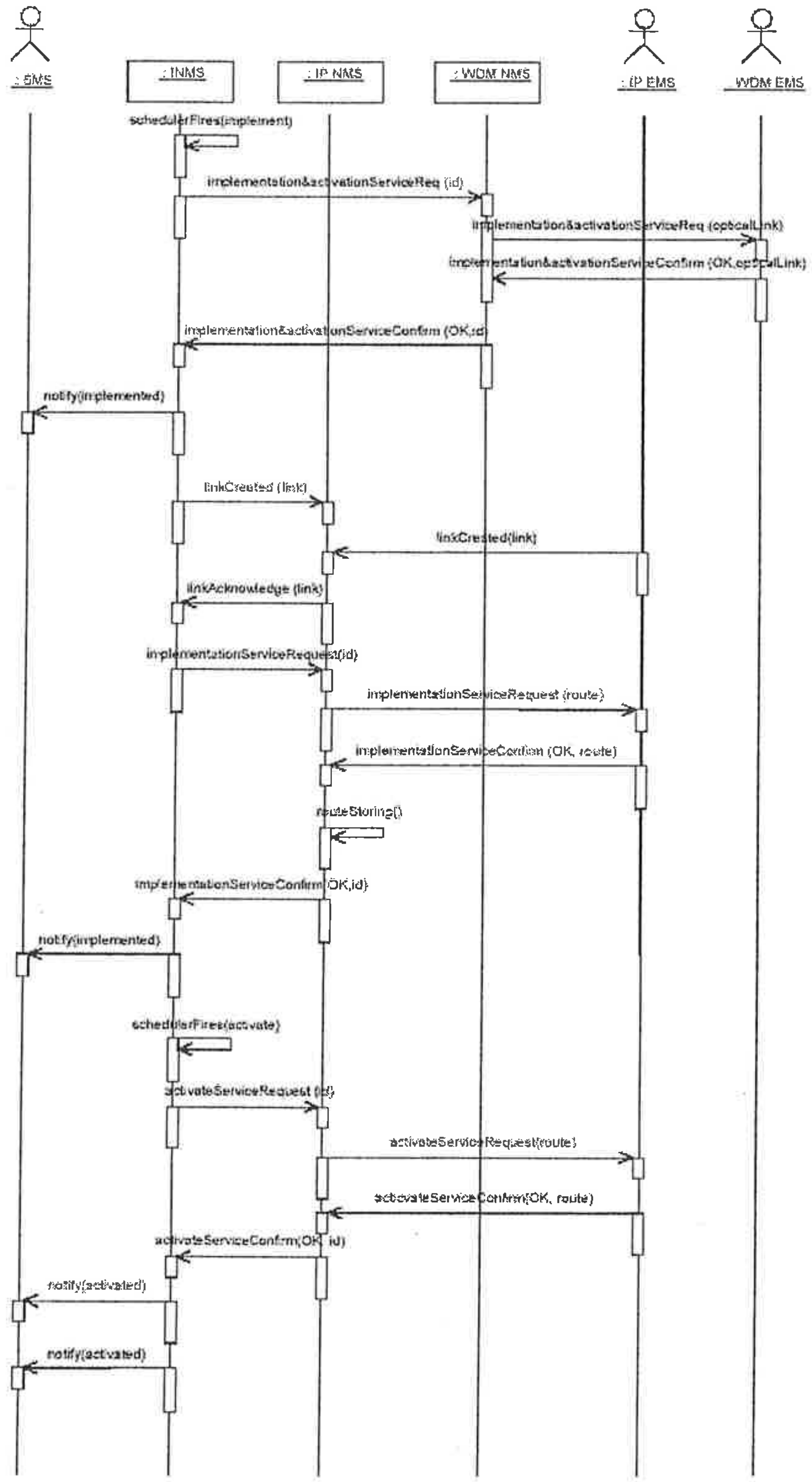


Fig.4. Provision of ICS involving creation of new lighpath. (WINMAN)

3.6.1. Description

- SMS sends a provisioning request to the INMS, specifying ICS identifier and new parameters for the service.
- INMS verifies the request and determines the validity of the requested service by applying policy rules. In the current step only INMS policy rules are verified.
- INMS makes the design of the network connections at INMS level, cause this system have enough info for that sort of decisions.
- Once INMS have made the design, it sends a request to the WDM NMS to design the WDM path in the WDM domain.
- WDM NMS proceeds to verify WDM policy rules and then to design and reserve the WDM path.
- WDM NMS sends response to INMS pointing success.
- INMS sends a request to the IP NMS to design the IP route.
- IP NMS verifies the request against policies and proceeds to the creation of the IP route.
- IP NMS sends a service created confirmation to the INMS.
- The implementation task is scheduled.
- A notification is sent after the schedule.
- The activation task is scheduled.
- A notification is sent after the schedule.
- At the scheduling time, INMS send a request to the WDM for the implementation and activation of the designed optical path.
- WDM NMS forwards that request to the correspondent EMSs, indicating the optical link to be implemented and activated.
- The EMSs responds to the WDM NMS indicating success in the operation.
- The WDM NMS forwards that confirmation to the INMS.
- INMS notifies to the IP NMS that a new link has been created, thus the IP topology has changed.
- The creation of a new link in the optical domain can be detected by the network elements, and notified upwards. In that case the EMS sends a notification to the WDM NMS informing about the new link created. Only the first notification of the new link is processed by the IP-NMS. If this notification from the EMS were not received, the IP NMS would send to the appropriate EMSs a new link notification.
- IP NMS sends notification to the INMS indicating that IP NMS has been informed about the new link.
- INMS sends an implementation request to the IP NMS, indicating the service identifier.
- IP NMS forwards the request to the appropriate EMSs.
- EMS sends a response to the IP NMS.
- IP NMS stores the route in the database.
- IP NMS sends a confirmation to the INMS about the service implementation.
- INMS notifies to the SMS that the service has been implemented.

- At the scheduled time the INMS sends to IP NMS an order to activate the given connection.
- IP NMS sends to EMS the order to activate the new route and receives its confirmation.
- IP NMS notifies the activation confirmation to the INMS (for WDM).
- IP NMS notifies the activation confirmation to the INMS (for IP).

3.7. STOLAS Network Management

The STOLAS Management System follows to a certain extend the outline of the general management architecture.

The OLS-EMS has knowledge about the OLS-NEs and is able to create/delete a communication through the OLS network. The OLS-GUI is the point where the operator has access to the OLS management network.

The interfaces between the management system are all CORBA (Common Object Request Broker Architecture) based with a possible exception for the IP-routers (Nes under IP_GUI). The interfaces have to be specified in IDL (interface definition language).

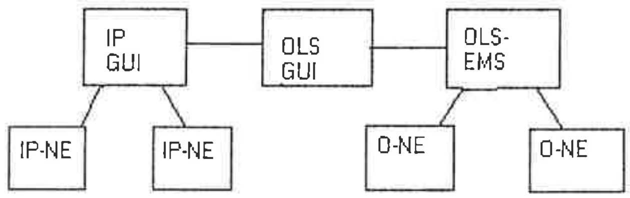


Fig.5. STOLAS network Management System.

3.8. STOLAS Network Elements

3.8.1. Optical Edge Router

The OER (optical edge router)has on his access side data interfaces like GigabitEthernet or SDH (input ports and output ports). On his network side the OER has optical interfaces (input ports and output ports) with the optical label added. For management purpose the edge router has an Ethernet interface. The data ports are PTP (physical termination points).

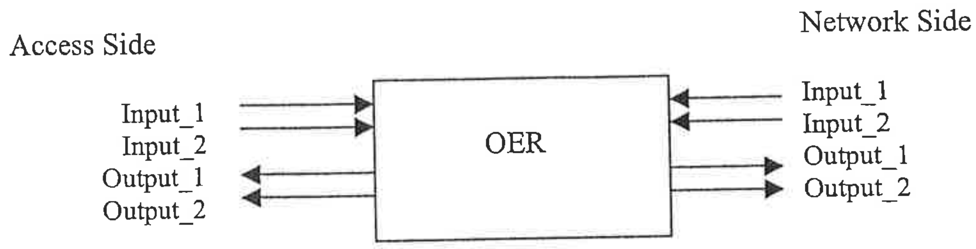


Fig.6 . Data ports of OER

3.8.2. Optical Cross-Connect

The OXC is placed in the optical network. It has inputports, outputports, addports and drpports. There is a possibility of muti-cast from inputports to dropports and/or outputports. For management purpose the OXC has an Ethernet interface. The add/drop ports can be connected to network side ports of the OER.

3.8.3. Optical Add-Drop Multiplexer

The OADM is placed in the optical network. An OADM consist of two modules that have each one inputport, outputport, addport and dropport. There is multi-cast within a module possible. For management purpose the OADM has an Ethernet interface.

Physically OADM-NE and OER-NE are in the same box. The addports/dropports can be connected to the network side of the OER.

3.8.4. STOLAS internal Structure of OLS-EMS

For the testbed the EMS is constituted by a number of classes.

The OLS-EMS will manage a number of network elements (4 in the testbed). The focus is on the configuration management (i.e. establishing and removing subnetwork connections). To achieve this the OLS-EMS should be able to instruct the optical NEs to set up or delete OLS-WDM lighpaths. The configuration of the network and its elements should be represented.

This configuration consists of managed elements, topological links, physical termination points and connection termination points, logically separated in a number of subnetworks. This structure will be presented to the EMS-operator using a GUI.

Fault management of the optical nodes (laser error, fibre cut, etc.), and performance parameters (like power measurements and S/N ratio) should be added and defined into the WDM-EMS. The structure of the WDM-EMS is shown in figure (7).

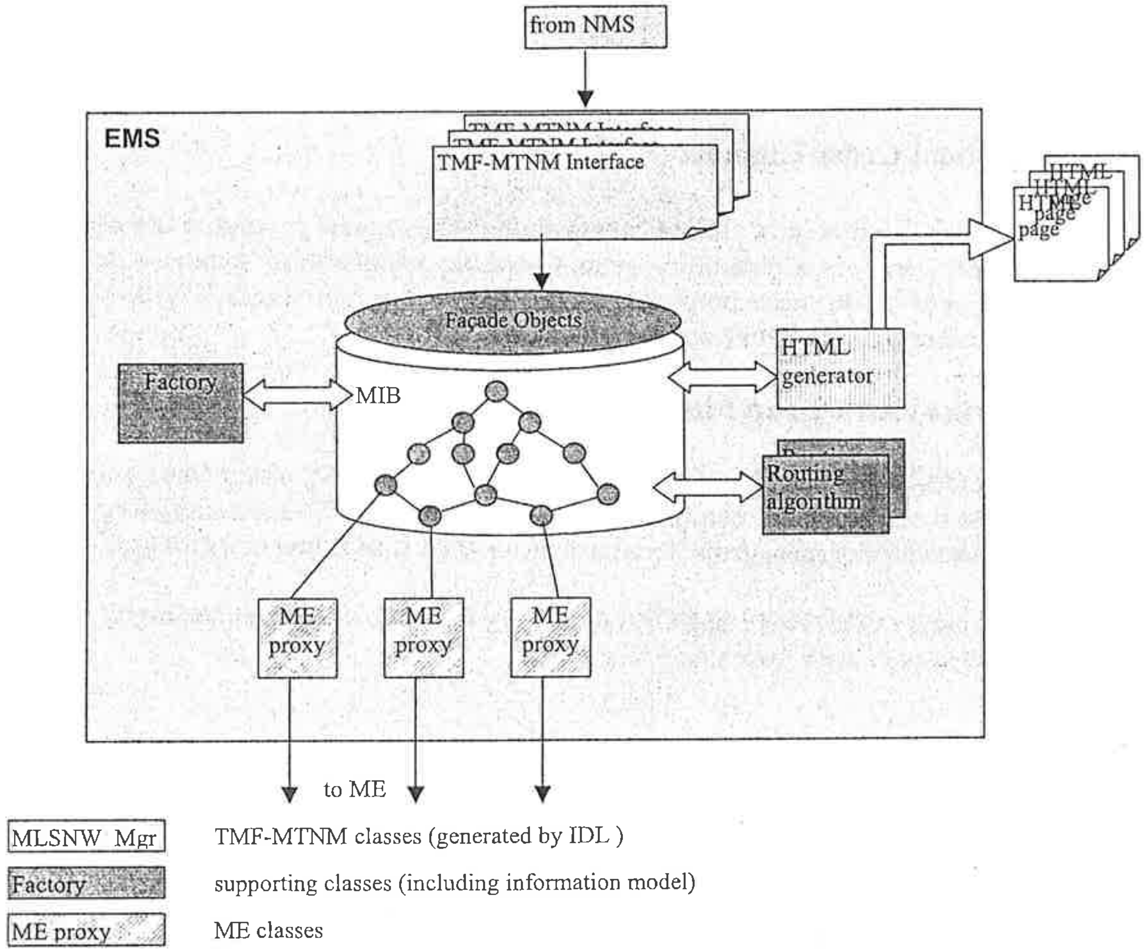


Fig.7 Structure of the OLS-EMS

3.9. Information Model

The basic and dynamic configuration of the subnetwork managed by the EMS is described by the Information Model (IM). This model gives the objects that can be present in the subnetwork, and their relationships. Instances of these objects are contained in the database of the EMS. This database has a static and a dynamic part. In the static part, the physical configuration is represented, like the network elements, their physical ports and the links between those ports. The dynamic configuration contains the connections that are present in the network elements, together with the connection termination points. If a new optical network element is added, the static configuration changes; if a new subnetwork connection is created, the dynamic configuration is altered. The instances of the information model will be contained in a List of Top objects, which is a data member of the Facade objects. These objects are the facade objects of the MIB, through which all of the operations on the MIB are initiated.

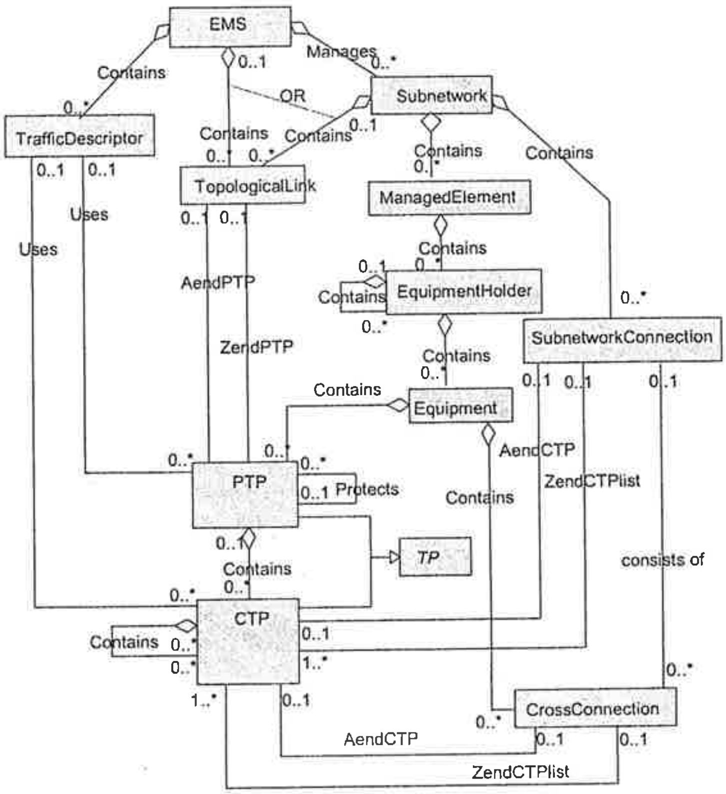


Fig.8 Object classes constituting the OLS-network information model (STOLAS-WINMAN)

3.9.1. TopObjects

Each object in the information model inherits from the Top object. Attributes will be added in the constructor of the object, which also gives the attributes an initial value. With the **set** operation, attributes can be assigned a value. If a set operation is invoked using a name that is not valid (i.e. for which no **addAttribute** operation is invoked), the set operation will fail. With the parentId and the containmentList the object instance tree can be created and searched.

Object Name	**TopObjects**	
Object Description	TopObjects is the base class for all STOLAS based information model classes	
Super Class	Top	
ParentId	none	
ChildIdList	All managedObject, which represents EMS System	
Attributes	globalDefs::NamingAttributes_T	name
	String	UserLabel
	String	NativeEMSName
	String	Owner
	globalDefs::NVSList_T	additionalInfo
Operations	set(TattributeId aId, const Tval& tval)	

Object Name	**EMS**	
Object Description	This object represents the aggregation of the networks as managed by the EMS.	
Super Class	TopObjects and emsMgr::EMS_T	
ParentId	none	
ChildIdList	Subnetwork, TopologicalLink, TrafficDescriptor	
Attributes	String	EmsVersion
	String	type
Operations	getAllTopLevelSubnetworks	
	getAllTopLevelTopologicalLinks	
	getAllManagedElements	
	createTrafficDescriptor	
	deleteTrafficDescriptor	
	getAllTrafficDescriptors	

Object Name	**Subnetwork**

Object Description	This object represents a logical grouping or partitioning of NetworkElements in a way that is entirely determined by the EMS	
Super Class	TopObjects and multiLayerSubnetwork::MultiLayerSubnetwork_T	
ParentId	EMS object	
ChildIdList	TopologicalLink, SubnetworkConnection,ManagedElement	
Attributes	Topology_T	subnetworkType
	TransmmissionParameters::LayerRateList_T	supportedRates
Operations	activateSNC	
	createAndActivateSNC	
	createSNC	
	deactivateAndDeleteSNC	
	deactivateSNC	
	deleteSNC	
	getAllEdgePoints	
	getAllManagedElements	
	getAllSubnetworkConnections	
	getAllTopologicalLinks	

Object Name	ManagedElement	
Object Description	The managedElement represents an abstraction of a set of co-located physical resources (i.e. OADM, Switch) managed as a single entity by the an EMS	
Super Class	TopObjects and managedElement::ManagedElement_T	
ParentId	EMS	
ChildIdList	EquipmentHolder	
Attributes	String	Location
	String	Version
	String	productName
	CommunicationState_T	communicationState
	TransmissionParameters::LayerRate List_T	SupportedRates
Operations	createSNC (This is considering as CC)	
	getAllEquipment	
	getAllPTPs	
	getAllCrossConnections	
	setTPData	

Object Name	CTP

Object Description	a CTP(Connection Termination Point) is an actual or potential end point of a subnetwork connection. For a DWDM Och CTPs, it contains an attribute specifying the wavelength it occupies.
Super Class	TP
ParentId	PTP or CTP
ChildIdList	TPIdList(containing lower order CTP's)
Attributes	none
Operations	getContainingTPs

Object Name	**EquipmentHolder**	
Object Description	This object class represents resources of the NE that are capable of holding other physical components. Specific resources that are represented by instances of this entity are racks(bays), shelves, and slots.	
Super Class	TopObjects and equipment::EquipmentHolder_T	
ParentId	ManagedElement	
ChildIdList	Equipment , EquipmentHolder(with lower order)	
Attributes	Boolean	alarmReportingIndicator
	EquipmentHolderType_T	holderType
	NVSList	expectedOrInstalledEquipment
	EquipmentObjectTypeList_T	acceptableEquipmentTypeList
	HolderState_T	holderState
Operations	getAllEquipment	

Object Name	**EquipmentHolder**	
Object Description	This object class represents resources of the NE that are capable of holding other physical components. Specific resources that are represented by instances of this entity are racks(bays), shelves, and slots.	
Super Class	TopObjects and equipment::EquipmentHolder_T	
ParentId	ManagedElement	
ChildIdList	Equipment , EquipmentHolder(with lower order)	
Attributes	Boolean	alarmReportingIndicator
	EquipmentHolderType_T	holderType
	NVSList	expectedOrInstalledEquipment
	EquipmentObjectTypeList_T	acceptableEquipmentTypeList
	HolderState_T	holderState

Operations	getAllEquipment

Object Name	CC	
Object Description	A cross-Connection represents a physical connection within a managed element. A cross-Connection is atomic and is identified, similarly to an SNC in a singleton subnetwork, based on its external shape(A end(s), Z end(s), "SNC" type, and directionality).	
Super Class	TopObject and subnetworkConnection::CrossConnect_T	
ParentId	Equipment	
ChildIdList	none	
Attributes	Boolean	active
	ConnectionDirection_T	direction
	SNCType_T	ccType
	NVSList_T	aEndNameList
	NVSList_T	zEndNameList
Operations	none	

Object Name	TrafficDescriptor	
Object Description	Represents a collection of attributes that are used to define bandwidth and QoS characteristics on a CTP.	
Super Class	TopObject and trafficDescriptor::TrafficDecriptor_T	
ParentId	EMS	
ChildIdList	none	
Attributes	ServiceCategory_T	serviceCategory
	TrafficParameterList_T	trafficParameters
	String	conformanceDefinition
Operations	getAssociatedCTPs	

Note: These object classes are from IST-WINMAN project and is confidential information, where says TopObjects before was written LAMPIONTop, but was changed to adapt it to the STOLAS project . There are more classes but they are not in the scope of this report. For more information refer to WINMAN.

As we are going to consider each node as a Network Element, each node should be characterized. All the components of the node should be described to have a perfect performance, that means characterize the following components:

- Tunable transmitters
- 2R regenerators
- Wavelength converters
- Amplitude modulators
- Label detection modules
- Interfaces
- Optical switch controllers for the delay lines (FDLs)
- Label swappers
- Optical supervisory channel

Note: It is possible that many other components should be added or deleted to/from the list since the STOLAS project is still under development.

3.10. Optical Supervisory Channel

The OSC of the STOLAS project will use a 1310 nm wavelength that will be transported over the same fibre. The OSC is terminated electrically on every node in the ring. The OSC provides communication capabilities between Network Element (NE) and the EMS. The interface between the NE and the EMS is based on CORBA IIOP, which uses IP as transport protocol. All those interfaces and characteristics should be programmed on a C++ or Java base.

Note: under discussion within the STOLAS project.

3.11. Data Bases

The STOLAS/WINMAN solution requirements should consist of 3 types of data bases used for network configuration purposes and for fault and performance data:

Network inventory or Physical database:

The network inventory database maintains the physical inventory of the network resources, represents the equipment in the network. At least the following attributes should be found:

-Cards
-Ports
-Fiber conduits
-Location

Logical data base:

The logical data base maintains the logical inventory of the network resources, and consists of logical paths that are supported by physical resources of the network, and has at least the following attributes:

- Protection of the path
- Available bandwidth
- Network quality: Delay and Delay variation

Connectivity database:

All active or the ones that are going to be activated connections in the network are stored with their attributes like:

- Connection status
- Logical route
- Connection bandwidth
- Connection quality and delay

Fault topology database:

Should have at least the following attributes:

- Fault case
- Fault severity
- Location

3.12. Conclusions

The focus of this chapter has been done under the possibility of a general application of the STOLAS Management System for the demonstrator that is going to be used to probe the STOLAS scopes, but to be readapted to a large network, where many other components and issues could be necessary.

As an adaptation to the demonstrator where there are only three basic nodes: OADM, OXC and PhaSar, many blocks could be avoided. In that case, the use of data bases is not necessary, protection systems are not specified since we have not decided what kind of protection is needed for STOLAS, only monitoring of the components' performance, for that issue signaling protocols are needed, also signaling allows to discover failures, but between nodes. For a short testbed signaling protocols should be readapted. An adaptation of RSVP-TE signaling protocol has been decided for STOLAS since the project scope is not on resource reservation and it will be used for path discovery ,so an adaptation of that protocol has to be done.

Further work has to be done to characterize a large network within STOLAS Management System concept, that is, in case that a large network would be implemented, connection between internal node control and distributed network management system should be discussed.

CONCLUSIONS and FURTHER WORK

4. Conclusions

This report has described the STOLAS node architecture, signaling and routing concepts of MPLS and MPλS and the Manager System for the STOLAS project.

Summary conclusions are presented regarding the performance of the concepts.

4.1. STOLAS node architecture

After studying the STOLAS node architecture we conclude that the main points to take in account are:

Considering slotted operation, we know which input ports are active at the start time of every slot. It has to be said that the wavelength assignment algorithm should be done for each output fiber independently from each other. This assumes that each packet/burst has a unique output fiber, determined by a lookup table (case of GMPLS), and excludes the possibility of a routing mechanism.

The key component for the switch control unit is the scheduler, for our STOLAS node and after some comparisons and simulations made by IMEC we will use LAUC-VF as a scheduling algorithm.

Due to specific novel orthogonal modulation scheme to label signals in STOLAS, then the problem arises what to do when no data is preceding the burst, if there is nothing to FSK on it is impossible to send the information, one option is to use "dummy" packets or bursts with no information, but our option is not to use offset between header and payload and to use an input buffer made of FDLs. The dimensioning of the FDL buffer influences the level of congestion in the control part of the router, so a key issue is the correct dimensioning of the FDL buffer. Some implications are related to that decision, use of input FDLs for optical buffering that adds complexity to the control issues, appropriate FDLs lengths, performance of these FDLs, use of fixed or switched ones.

Since we are going to use BCH channel orthogonal to payload channel: link utilization increases in comparison with other out-of-band signaling systems like SCM. Orthogonal signaling can scale better than SCM signaling. Due to SCM has problems with dispersion and RF processing of subcarriers at higher speeds, this system is not as scalable as STOLAS system. However, at high rates at the STOLAS label swapper the appearance of chirp could be a problem. More study has to be done to verify this issue.

Link management and protection should be emphasize due to the high level of link utilization.

4.2. MPLS signaling, basis for STOLAS

After studying MPλS signaling protocols we conclude that:

Both CR–LDP and TE–RSVP provide very similar functionality for establishing traffic-engineered, labeled switched paths. Each has its strengths and weaknesses. While LDP is the younger of the two protocols, RSVP has been previously deployed and has operational experience. It is true that there have been extensive enhancements to RSVP in order to support the needs of MPLS. As both CR–LDP and TE–RSVP evolve they will offer more and more similar functionality.

Eventually, MPLS traffic engineering should evolve into a single entity that combines the best-of-breed attributes from both TE–RSVP and CR–LDP. In the meantime, any MPLS implementation by original equipment manufacturers (OEMs) developing LER or LSR platforms should consider supporting both TE–RSVP and CR–LDP to ensure interoperability.

For the STOLAS project RSVP-TE signaling protocol has been chosen, but since we are not going to make resource reservations, this protocol will be used only for path discovery, so some adaptations should be done to limit its functions.

The control channels may also be used for transmitting and receiving signaling and routing messages. Each LMP control channel must individually negotiate the control channel parameters, and each active control channel must exchange LMP Hello packets to maintain LMP connectivity. For LMP, it is essential that at least one control channel is always available.

4.3. STOLAS Management System

The focus of the management system has been done under the possibility of a general application of the STOLAS Management System for the demonstrator that is going to be used to probe the STOLAS scopes, but to be readapted to a large network, where many other components and issues could be necessary.

As an adaptation to the demonstrator where there are only three basic nodes: OADM, OXC and PhaSar, many blocks could be avoided. In that case, the use of data bases is not necessary, protection systems are not specified since we have not decided what kind of protection is needed for STOLAS, only monitoring of the components' performance, for that issue signaling protocols are needed, also signaling allows to discover failures, but between nodes. For a short testbed signaling protocols should be readapted. An adaptation of RSVP-TE signaling protocol has been decided for STOLAS since the project scope is not on resource reservation and it will be used for path discovery ,so an adaptation of that protocol has to be done.

Further work has to be done to characterize a large network within STOLAS Management System concept, that is, in case that a large network would be implemented, connection between internal node control and distributed network management system should be discussed.

4.4. Recommendations and further work

- Performance simulations between "classic" STOLAS node architecture, the multistage architecture and loop-back configurations in order to describe pros and cons of each architecture.
- Accurate design for the scheduling algorithm.
- Improving of header processing times; detection, process and switching.
- Readapting of RSVP-TE signaling protocol.
- Implementation of the data bases for the Management System (programming).

Abbreviations and References

Abbreviations

STOLAS	Switching Technologies for Optically Labeled Signals
WDM	Wavelength Division Multiplexing
IP	Internet protocol
RSVP	Resource Reservation Protocol
CR-LDP	Constrain route Label Distribution Protocol
MPLambdaS	Multiprotocol Lambda Switching
OBS	Optical Burst Switching
CoS	Class of Service
QoS	Quality of Service
ATM	Asynchronous transfer Mode
LSP	Labeled Switched Path
XC	Cross-Connect
LSR	Labeled Switched Route
IETF	Internet Engineering Task Force
IGP	Internal Gateway Protocol
BGP	Border Gateway Protocol
SCU	Switch Control Unit
FDL	Fiber Delay Line
BHP	Burst Header Packet
DCG	Data Channel Group
CCG	Control Channel Group
FIFO	First in first out
AWG	Array Waveguide Gratings
TWC	Tunable Wavelength Converter
LAUC-VF	Latest available unused channel with void filling
OXC	Optical Cross Connect
OTN	Optical Transport Network
LMP	Link Management Protocol
ADM	Add-Drop Multiplexer
TE	Traffic Engineering
OSPF	Open shorter path first
EMS	Element Management System
OLS	Optical Label Switching
NMS	Network Management System
SLA	Service Level Agreement
IST	Information Society Technologies
NO	Network Operator
NSP	Network service Provider
INMS	Integrated Network Management System
GUI	Graphic User Interface

CS	Connectivity Service
TP	Termination point
PTP	Physical termination point
OCh	Optical Channel
FEC	Forwarding Equivalence Classes
OME	Optical Management Element
OADM	Optical Add-Drop Multiplexer
OCP	Optical Connection point
SNC	Sub-network Connection
CTP	Connection Termination Point
OER	Optical Edge Router
CORBA	Common Object Request Broker Architecture
SDH	Synchronous Digital Hierarchy
IDL	Interface Definition Language
IM	Information Model
MIB	Management Information Base
OSC	Optical Supervisory Channel

References:

[1] J. Turner, "Terabit burst switching". Journal High Speed Networks, vol 8, pp. 3-16. 1999.

[2] P. Gambini, et al., "Transparent Optical Packet Switching: Network architectura and demonstrators in the KEOPS project", IEEE J. Selec. Areas Commun., vol.16, pp. 1245-1259, Sept. 1998.

[3] M.K. Smith, C. van Dam. "Phasar-based WDM devices: principles, design and applications, IEEE J. Selec. Topics Quant. Electron, Vol. 2, 236-250, 1996.

[4] F. Callegati, H. Cankaya, Y. Xiong, and M. Vanderhoute, "Dessign of optical IP routers for Internet backbone applications". IEEE Commun. Magazine, pp. 124-128, Dec. 1999.

[5] A. Raboso, Graduation Thesis. "Performance Analysis of combined DPSK and ASK modulation Formats for Labeling of optical signals". Sept. 2001.pp.3-12.

[6] D. Awduche, et al., "Multiprotocol Lambda Switching: Combining MPLS Traffic Engineering Control with Optical Cross-Connects", draft-awduche-mpls-te-optical-03, April 2001.

[7] draft-ietf-mpls-lmp-02.txt

[8] draft-ietf-mpls-bundle-03.txt

[9] ITU-T Recommendation I.321 (1991)
B-ISDN Protocol reference model and its applications.

[10] K.Kompella et al., "OSPF Extensions to support MPLambdaS".
Draft-Kompella-ospf-mpls-extensions-00.txt.

[11] K.Kompella et al., "IS-IS Extensions to support MPLambdaS".
Draft-Kompella-isis-mpls-extensions-00.txt.

[12] OIF Contribution #2000.125.2 "User network interface 1.0 Proposal", 31/10/2000.

[13] Y.Cao, Y.Xu, "ODSI Signaling specifications_optical Gateway Protocol, version 1.0", July 2000.

[14] OIF Contribution #2000.171.0 "LDP extension for optical UNI signaling", November 2000.

[15] OIF Contribution #2000.140.1 "RSVP extension for optical UNI signaling", August 2000.

[16] Chunming Qiao, "Labeled Optical Burst Switching for IP-over-WDM Integration", IEEE Communication Magazine, September 2000, pp.104-114.

[17] D.J. Blumenthal, et al., "All-optical label swapping networks and technologies", IEEE J. of Lightwave Technol., Dec. 2000, pp. 2058-2075.

Figure references and other references:

[1] figures

- Jan Cheyns, Chris Develder, Erik v. Breusegem, Piet Demeester, "Routing in an AWG Based Optical Packet Switch". Submission for Photonic Network Communications. March 2002.
- www.netplane.com
- www.dataconnection.com

[2] figures

- Uyless Black. "Optical Networks, third generation transport systems". Prentice Hall,2002.

[3] figures

- Y. Xiong, Marc Vanderhoute. "Control architecture in Optical Burst-Switched WDM Network". J. Selected Areas Comm. Vol. 18. Oct 2000.

- "IP over DWDM Networks supported by GMPLS-based LOBS deploying Combined modulation Format".
Sulur, Ton Koonen, Idelfonso Tafur Monroy, Huug de Waardt, Jean Jennen, Opticomm 2001, August 2001, Denver, Colorado (USA)

- "Optical packet routing in IP over WDM networks deploying two-level optical labeling".
Ton Koonen, Geert Morthier, Jean Jennen, Huug de Waardt, Piet Demeester, ECOC 2001, October 2001, Amsterdam, The Netherlands.

[4] figures

- IST-WINMAN project, Deliverables (internal documents).