NTNU

Norwegian University of
Science and Technology

Employing Ethernet Multiple Spanning Tree Protocol in an OpMiGua network

Raimena Veisllari

Master in Security and Mobile Computing

Submission date: June 2010

Supervisor: Steinar Bjørnstad, ITEM

Co-supervisor: Peter Sjodin, Royal Institute of Technology, KTH Sweden

Norwegian University of Science and Technology
Department of Telematics

Problem Description

The Optical Migration Capable Networks with Service Guarantees (OpMiGua) concept has the main objective of combining the best properties from both circuit and packet switched networks into a hybrid solution. The main objective of this project is to employ the spanning tree protocol from Ethernet, which is a pure packet switched network protocol, in an OpMiGua hybrid network. The student will find how the topology of the network can be controlled and how two very different paths in the OpMiGua network can be set up by using MSTP. During the project, the student will build experience and competence of Ethernet networks specifically, as well as on packet switched networks in general.
The thesis will propose schemes of combining OpMiGua and spanning tree protocols in order to gain the benefits of the hybrid approach. Furthermore, the student will investigate and quantify the performance of the proposed schemes based on results obtained from simulation.

Assignment given: 22. January 2010
Supervisor: Steinar Bjørnstad, ITEM

Abstract

Hybrid optical packet/circuit switched networking architectures are increasingly becoming an interesting research field. They integrate and combine the high resource utilization of statistically multiplexed packet switched networks with the low processing requirements and guaranteed quality of service provided by circuit switched networks. The aim of this thesis is to integrate the OpMiGua hybrid optical network with Ethernet. Specifically, the work is focused on the compatibility of the Ethernet's loop-free topology protocols with the redundant multiple traffic service paths of OpMiGua.

We analyse the problems and limitations imposed on the network architecture and propose our topology solution called the SM chain-connectivity. The analysis and the proposed schemes are verified based on results obtained from simulations. Furthermore, we design an integrated logical OpMiGua node that relies on an Ethernet switch instead of the Optical Packet Switch for the Statistically Multiplexed traffic. To date, to our knowledge there are no studies analysing the compatibility of Ethernet and its protection mechanisms in a hybrid optical network. This is the first work addressing the use of Ethernet in OpMiGua.

Acknowledgments

I would like to express my gratitude to a number of people for without them the completion of this thesis would have been impossible.

My supervisor Steinar Bjornstad has been extremely patient and I would like to thank him for his guidance, support and motivation especially during times when my confidence seemed to fail me. I am thankful for the opportunity of having many interesting discussions leading me to further expand and deepen my knowledge. Furthermore, to him and my KTH co-supervisor, Peter Sjodin, for their feedbacks and comments and hopefully this is a thorough, better organized and well explained report because of them.

To the NordSecMob Consortium for without their support this Nordic adventure would not have been possible. Thanks are also due to Eija Kujanpää, May-Britt Eklund Larsson and Mona Nordaune for their administrative assistance.

To my family for their support and to May who always encourages me in everything that I do.

To my father.

Contents

List of Figures

List of Abbreviations

ATM	Asynchronous Transfer Mode
AWG	Arrayed Waveguide Grating
BE	Best Effort
BID	Bridge Identifier
BPDU	Bridge Protocol Data Unit
CA	Critical Applications
CAM	Content Addressable Memory
CIST	Common and Internal Spanning Tree
CLI	Command Line Interface
CoS	Class of Service
COST	Cross-Over Spanning Tree
CSI	Canonical Form Indicator
CSMA/CD	Carrier Sense Multiple Access with Collision Detection
DMUX	DeMultiplexer
DWRON	Dynamic Wavelength Routed Optical Network
EAPS	Ethernet Automatic Protection Switching
EE	Excellent Effort
ELPS	Ethernet Linear Protection Switching
FDL	Fiber Delay Line
FIFO	First In First Out
FTTP	Fiber To The Premises
GbE	Gigabit Ethernet
GMPLS	Generalize Multi Protocol Label Switching
GST	Guaranteed Service Transport
HCT	High Class Transport
IEEE	Institute of Electrical and Electronic Engineers
IC	Internetwork Control
IP	Internet Protocol
LACP	Link Aggregation Control Protocol
LAN	Local Area Network
MAC	Media Access Control
MAN	Metropolitan Area Networks
MEN	Metro Ethernet Network
MISO	Multiple Input Single Output
MPEG	Moving Pictures Experts Group
MPLS	Multi Protocol Label Switching
MRP	Metro Ring Protocol
MSTP	Multiple Spanning Tree Protocol
MUX	Multiplexer
NC	Network Control

NCT	Normal Class Transport
NGN	Next Generation Networks
OAM	Operation, Administration and Management
OBS	Optical Burst Switching
ODL	Optical Delay Lines
OPS	Optical Packet Switching
OCS	Optical Circuit Switching
OXC	Optical Cross-connect
OpMiGua	Optical Migration Capable Networks with Service Guarantees
PBB	Provider Backbone Bridging
PBB-TE	Provide Backbone Bridging with Traffic Engineering
PBS	Polarization Beam Splitter
PLR	Packet Loss Ratio
PLS	Physical Layer Signalling
QoS	Quality of Service
RRSTP	Rapid Ring Spanning Tree Protocol
RPR	Resilient Packet Ring
RSTP	Rapid Spanning Tree Protocol
SDH	Synchronous Digital Hierarchy
SONET	Synchronous Optical NETworking
SP	Service Provider
STA	Spanning Tree Algorithm
STEP	Spanning Tree Elevation Protocol
STI	Spanning Tree Instances
STP	Spanning Tree Protocol
SWRON	Static Wavelength Routed Optical Network
SM	Statistically Multiplexing
TBTP	Tree-Based Turn Prohibition protocol
TCA	Topology Change Acknowledgment
TCI	Tag Control Information
TCN	Topology Change Notification
TPID	Tag Protocol Identifier
TWC	Tunable Wavelength Converter
VID	Vlan IDentifier
VLAN	Virtual Local Area Network
VLP	Variable Length Packets
WRON	Wavelength Routed Optical Network

Chapter 1

Introduction

1.1 Motivation and current work

Hybrid optical packet/circuit switched networking architectures [1], [3], [11], [14], [19] are increasingly becoming an interesting research field. They integrate and combine the high resource utilization of packet switched networks with the low processing requirements and guaranteed quality of service provided by circuit switched networks. The aim is to improve the overall network performance by obtaining the advantages of both switching technologies while trying to minimize or avoid their disadvantages. The Optical Migration Capable Networks with Service Guarantees (OpMiGua) [2][3] is a hybrid architecture that introduces the ability of dividing the traffic into two service classes while using the capacity of the same wavelength in a wavelength routed optical network (WRON). The traffic is distinctively divided into:

1. Guaranteed Service Transport (GST) service class for the circuit-switched traffic;
2. Statistically Multiplexed (SM) service class for the best-effort packet-switched traffic.

Thus, this network model achieves a high throughput and guaranteed service with no packet loss and constant delay [3].

Ethernet is the most widely deployed Data Link layer technology with more than 85 percent of all installed network connections and more than 95 percent of all Local Area Networks [4]. Its plug-and-play deployment simplicity, low-cost and optimal characteristics

for carrying IP traffic have appealed to the networking industry. In addition, it is evolving to meet the increasing bandwidth and functionality demands required from networking technologies nowadays. The efforts have resulted in the usage of 10Gbit/s Ethernet in enterprise and carrier networks while continuing the expansion to 40Gbit/s and 100Gbit/s Ethernet. The major large-scale trend shows that Ethernet is to dominate in the access and metro network for the future [5].

The aim of this thesis is to integrate the hybrid OpMiGua network and Ethernet. Specifically, the work is focused on the compatibility of the Ethernet's spanning tree based protocols with the redundant multiple traffic service paths of OpMiGua. Much research work [23, 24, 31-34] has been directed toward enhancing the spanning tree protocols' recovery time after a failure. Especially when employing Ethernet in metro and carrier networks, the down time is crucial. The research community and standardization bodies have also focused on creating and standardizing new loop-free protocols [45, 46, 51] when using Ethernet in the optical domain. However, the spanning tree protocols (xSTP) are still widely deployed in legacy Ethernet switches and proprietary implementations claim recovery times comparable with those of the new protocols [24]. Furthermore, other research work has focused on hybrid optical packet/circuit switched networks [1-3, 6, 8, 10, 11-14, 18, 19]. We chose OpMiGua as the hybrid optical network since NTNU has been part of its creation and further work is being carried on its architecture. However, to our knowledge there are no studies analysing the compatibility of Ethernet and its protection mechanisms in a hybrid optical network.

1.2 Problem definition

In a generalized scenario, Ethernet switches are connected to an OpMiGua network and are aggregating traffic while assigning it to the GST or SM classes based on the Quality of Service (QoS) policies in the switches, as shown in figure 1.1.

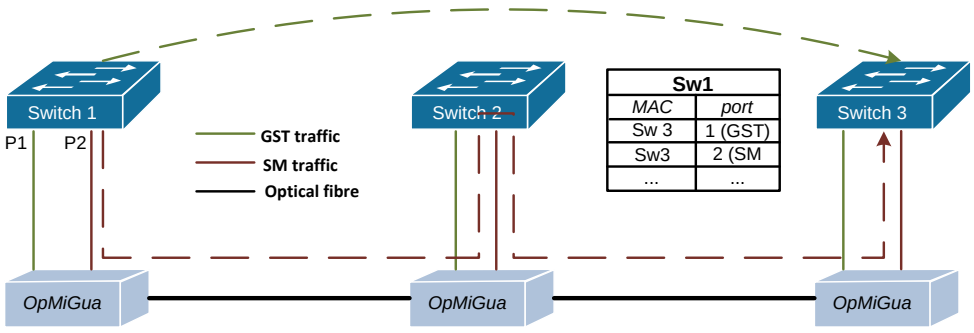

Figure 1.1 A simplified OpMiGua network topology

The GST traffic is circuit-switched by the OpMiGua nodes in direct physical connections bypassing the intermediate nodes. The SM traffic follows the same physical connections between nodes and is processed at each hop because it is packet-switched based on the information carried on the header of each packet. The former presents no transitional processing for the delay sensitive traffic, while the latter allows a higher throughput of the network through the statistical multiplexing of the traffic. Thus, both paths are needed to obtain the benefits of a hybrid optical network. However, Ethernet does not allow loops and employs protocols based on the Spanning Tree Algorithm to create a loop free topology. It is implemented on all the nodes of a network turning off interfaces and considering the two aforementioned paths as redundant. As a consequence, the GST or SM ports might be blocked on the Ethernet nodes.

1.3 Goals

The main goal of this thesis is to analyse and propose solutions for the interoperability of the implemented xSTP in Ethernet switches with the packet-switched and circuit-switched traffic of the hybrid approach. The problems derived in the analysis are closely related to the way the physical connections in the optical domain are logically perceived by the Ethernet nodes. The dedicated point-to-point lightpaths in an OpMiGua network are recognized as a shared medium by the Ethernet. The intermediate nodes responsible for switching the SM traffic do not forward the received BPDUs thus leading to a failure in the logical spanning tree convergence process. Different network connectivity scenarios are considered for this analysis and we propose the *SM chain-connectivity* topology solution to the interoperability problems. We derive the network physical limitations when implementing a provider's network that integrates both domains with respect to xSTP. This architecture allows for a full spanning tree convergence.

Furthermore, we address the problem introduced by xSTP which blocks the redundant paths. Our solution is to logically differentiate the packets by using the VLAN tag as a label. Different ways of differentiating the traffic through the VLAN tags hierarchy and QoS mappings are proposed. This is achieved by employing the Multiple Spanning Tree Protocol (MSTP) which has the ability to build several logical topologies using the Virtual Local Area Network (VLAN) tags. We assign different VLANs for packet/circuit switched traffic at Layer 2 based on QoS allowing the normal function of the two distinct traffic paths.

Additionally, we scrutinize the possibility of avoiding the use of MSTP because of its difficulties in configuration, management and scalability. We investigate the possibility of using the simplified spanning tree protocols in order to avoid loops without blocking the functional ports on the Ethernet switches.

Another important part of the thesis is the verification of the analysis and proposed schemes based on results obtained from simulations. We also propose an integrated OpMiGua node design that relies on an Ethernet switch instead of its Optical Packet Switch (OPS) block. The purpose is to be able to implement a viable OpMiGua node because the OPS [16] is still commercially not available other than in research.

1.4 Methodology and Outline

The work is based on an empirical approach. The analysis and the proposed schemes are validated in a Data Link Layer level by the emulation of Cisco switches using the open source emulation software dynamips [47].

The report is structured as follows. First we introduce OpMiGua and its main characteristics in Chapter 2. Then in Chapter 3 we give a short overview of Ethernet with the focus on carrier Ethernet technologies and the Spanning Tree based protocols. In Chapter 4 we analyse the interoperability problems considering the underlying OpMiGua network as a transparent Service Provider. Furthermore in Chapter 5 are given the network topology solutions and the description of the simulation model, results and discussions from the simulation runs. Later, in Chapter 6 we propose two node designs replacing the optical packet switch of the OpMiGua node with an Ethernet one. The achieved work with its advantages and limitations is discussed in Chapter 7. The thesis report is finished with our conclusions in Chapter 8 and the listing of the proposed further work in Chapter 9.

Chapter 2

OpMiGua

Future optical networks should be able to serve a client layer that includes packet-based networks [11], [16], [20]. The aim is to provide to the Internet and the IP layer a high-capacity transmitting technology. Nowadays the switching solutions are mostly performed through the electronic fabric, which is why the bandwidth utilization is limited by the capacity and the conversion speed of these circuits. Furthermore, the utilization of Optical Circuit Switched (OCS) networks for traffic with a bursty nature as IP is bandwidth inefficient [1]. This is mostly because of the coarse granularity which is the wavelength. The intermediate nodes do not have the ability to apply the full capacity on those connections by means of statistical multiplexing. Furthermore, the over-dimensioning of the number of connections and the bandwidth reservation for each connection is needed to avoid delays and extensive buffering at the ingress nodes [20]. Optical Packet Switching (OPS) and Optical Burst Switching (OBS) overcome these problems by introducing statistical multiplexing (SM) at the optical layer [12]. However, they lack the beneficial guaranteed-service characteristics of a circuit-switched network.

Hybrid optical packet/circuit switched networking architectures integrate and combine the high resource utilization of packet switched networks with the low processing requirements and guaranteed quality of service provided by circuit switched networks. In [1] are categorized and listed the most inquired hybrid optical network architectures based on the level of the interaction and integration of the two domains. In this thesis we are focused on the integrated hybrid networks, more specifically the OpMiGua network architecture, where

the different technologies share the bandwidth of the same wavelength resources simultaneously on a packet-per-packet basis [10].

2.1 Introduction

The Optical Migration Capable Networks with Service Guarantees (OpMiGua) [3], [9] is a hybrid architecture that introduces the ability of dividing the traffic into two separate service classes. This achieved while using the capacity of the same wavelength in a wavelength routed optical network (WRON). The traffic is distinctively divided into:

1. Guaranteed Service Transport (GST) service class for the circuit-switched traffic;
2. Statistically Multiplexed (SM) service class for the best-effort packet-switched traffic.

The network model attains the advantages of high throughput efficiency and guaranteed service with no packet loss and constant delay [3].

2.2 The OpMiGua hybrid network concept

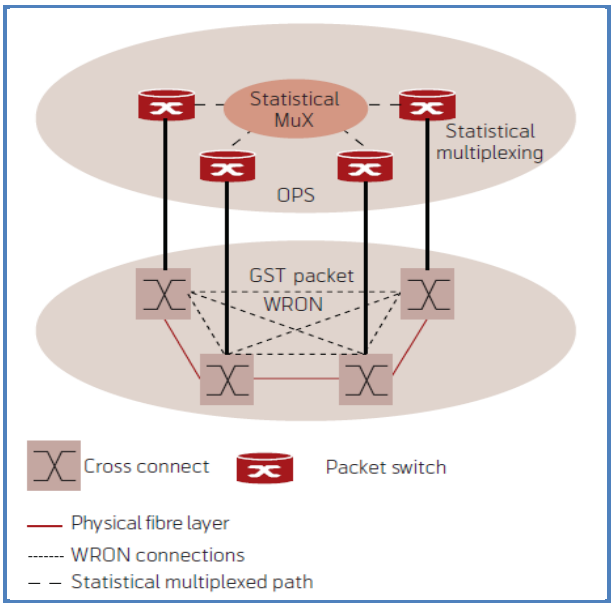

Figure 2.1 A hybrid network model illustrating the sharing of the physical fibre layer. The optical cross connects and optical packet switches are co-located, either as separate units or as one integrated unit. The WRON can be a Static or a Dynamic-WRON, taken from [3].

In figure 2.1 is presented a simplified network model of the hybrid architecture. The GST traffic follows pre-assigned lightpaths from the source to the destination through either a static or dynamic WRON. These packets are served with the benefits of the circuit-switched

paths that offer a fixed delay, no jitter and no packet loss. The lightpaths are created by the interconnection of fibres and wavelengths through one or many, static or dynamic optical cross connects.

The use of the optical packet switches employs a hybrid network where the SM packets are switched based on their header information. The Packet Loss Ratio (PLR) of this traffic is improved compared with pure OPS/OBS networks. This is achieved by the bypassing of the packet switches from the GST traffic, thus reducing the processing overhead and overload of these nodes.

On the other hand, the strict priority of the GST packets is achieved based on two design principles:

1. The GST packets of a traffic flow do not contend with other GST flows since there is at least one assigned wavelength for a given source-destination combination. A GST circuit in our thesis is considered as a pre-assigned wavelength in a SWRON. The use of the SWRON architecture avoids the lightpath setup delay. However, a lightpath might not need to preserve the wavelength continuity constraint if a DWRON with wavelength conversion is used. Moreover, a GST path in a synchronous system can be a timeslot and there can be multiple paths within the same wavelength.
2. The contention of GST with SM traffic is avoided by implementing a reservation technique as presented in [9] and [10].

2.2.1 Hybrid Asynchronous Node Design

The hybrid node design is illustrated in figure 2.2. The GST and the SM packets use the same input/output ports but are separated by employing two different states of polarization (POS). At the input interface, a polarization beam splitter (PBS) is assigned for each wavelength. This means that the capacity of a given wavelength channel is not doubled as in traditional polarization multiplexing where the two polarizations are transmitted simultaneously. In OpMiGua the two different polarizations states are used to label the two traffic classes all optically [3]. The packet header allows the separation of the traffic into GST and SM path instead of the orthogonal state of polarization. Other optional optical label techniques may use a sub-carrier modulation method [14]. Furthermore, the switching fabric can use an opto-electronic converter to allow the use of an electronic label. In chapter 6 we propose two integrated node designs considering both the optical and electronic labelling

scenarios. The details of these label processing techniques are out of the scope of this thesis. However, despite the optical labelling used, the packet separator forces the GST traffic into the OXC while the SM traffic into the OPS.

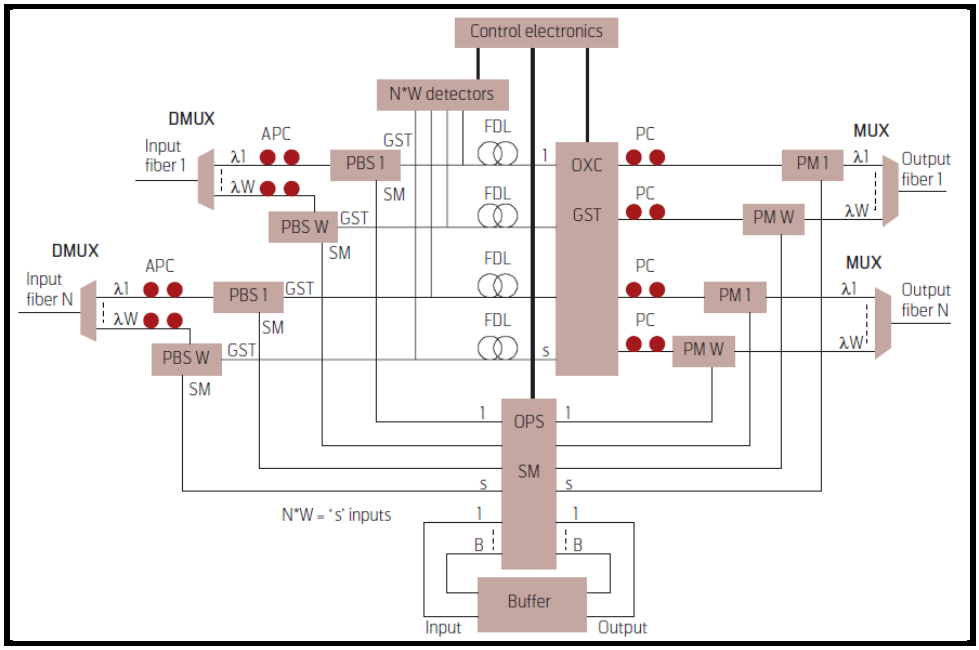

Figure 2.2 A functional illustration of a hybrid node with *w* wavelengths from one fiber, taken from [3].

The GST packets are delayed in the fibre delay lines (FDL) as a part of the proactive time-window reservation technique which avoids the contention between GST and SM packets. In the case of switching variable length packets (VLP), the FDLs will delay GST packets for a time corresponding to the longest SM packet. However, when employing Ethernet on top of an OpMiGua node, typically the maximal transmission unit would be as specified in 802.3. The priority reservation techniques and their performance evaluation are given in details in [10].

An electronic header for low speed networks should not suffer from the opto-electronic conversion, however for networks with 100Gbps throughput and optical label is preferred. In [3] are given the three main advantages of using polarization to optically label GST and SM packets:

1. No fast switches operating on a per packet basis;
2. No separate header is required, meaning no fast electronics for header processing in the GST case;
3. No guard band is required because there is no processing and insertion of headers.

The GST packets are switched to the correct output port based on the configuration of the OXCs, while the SM packets can use any of the idle output wavelengths. The arrival of GST packets is signalled by the packet combiner to the control unit of the OPS. In this case the SM packets may need to be re-labelled and delayed by the optical buffers. Thus, by inserting SM packets in-between the gaps created by subsequent GST packets, the resource utilization is increased. Bjørnstad et al. [8] demonstrate a three-node OpMiGua hybrid network with up to 98% utilization of the bandwidth and SM packet loss of less than 10^{-6}. The experiments also confirm zero packet loss and jitter for the GST traffic regardless of the SM traffic.

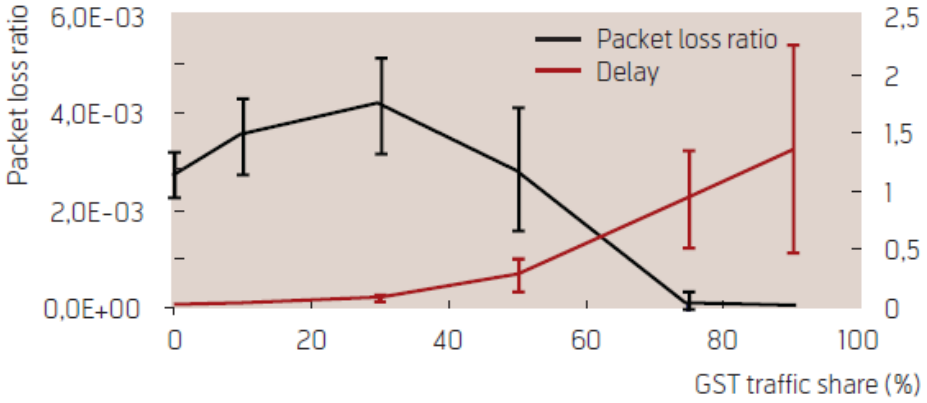

Figure 2.3 PLR and buffered packets delay of SM traffic as a function of GST traffic share, taken from [20].

Furthermore, in [20] is evaluated the performance of the system when the GST packet size is fixed to 100 times the mean Internet 256 B SM packet length in order to avoid the overhead of the reservation technique. The results in figure 2.3 show the effect of two counteracting factors for the Packet Loss Ratio:

1. The wavelength availability decreases with the increase of the GST share up to 30%, resulting in higher PLR for the SM packets.
2. The increase of the GST share and of PLR because of 1) increases the available buffer resources share for the remaining SM traffic in the OPS. This results in a lower PLR because of a longer mean waiting time for SM packet insertion and the packets are waiting in the buffer.

Furthermore, the SM packet delay increases with the increase of the GST share that causes SM traffic contention. The buffered packets will wait longer for a free wavelength, resulting in a longer delay for the SM traffic with the increase of the GST share.

2.3 Quality of Service

[3] presents three traffic classes named GST bearer service, high class transport (HCT) bearer service and normal class transport (NCT) bearer service. The HCT and NCT classes are sub-classes of the SM class. The differentiation between the HCT and NCT classes is performed in the electronic buffer in the OpMiGua node presented previously. However, as discussed later in chapter 4 and 5, in our case it is the Ethernet switch aggregating traffic which can dynamically assign these service classes. The HCT class is given absolute priority when a wavelength to the destination becomes vacant, that is why the HCT class experiences lower delay than the NCT class. This scheme is called the buffer priority (BP) scheme. Regarding the packet loss differentiation, the HCT class has access to all the inputs of the buffer, while the NCT class has limited access. The number of inputs which can be accessed by the NCT class are also being shared with the HCT-class. This means that a given number of inputs on the buffer will be reserved for HCT class [3], which is why the HCT class has a higher probability to be buffered compared to the NCT class.

The GST class experiences constant switching delay and will not suffer any packet jitter. Furthermore, there is no re-sequencing of packets and no packet loss is caused by contention. For the HCT-class the delay and jitter is kept at a minimum and the packet loss rate should be 10^{-6} or better when considering the class carrying MPEG2 and MPEG4 traffic.

When employing an Ethernet switch instead of OPS in the node, it can dynamically assign high priority (HCT for example) SM traffic into a GST path in the case of low provisioning of the available wavelengths. Furthermore, the issues of mapping Ethernet QoS classes with the hybrid node traffic classes are discussed in chapter 5.

Chapter 3

Ethernet and Spanning Tree Algorithm

3.1 Introduction

Ethernet was originally designed to allow simple data sharing over a local area network (LAN) in campuses or enterprises. At the present time the standard technology used to manage data transmission on carrier networks is SONET/SDH. It is a circuit-based system which is mainly intended for the transport of voice traffic. In the last years new technologies are being developed to replace it. This is mainly because of the rise of new requirements from the carriers' customers [17]. The residential triple play market (data, television, voice) requires high peak data bandwidths approaching Gigabits per second, priority for voice traffic and high definition broadcast/on-demand video services. Residential access networks are evolving to fiber to the premises (FTTP) technologies to support these bandwidth and QoS requirements. Furthermore, metro core networks are being driven to a converged IP/Ethernet architecture which is capable of offering prioritized services and handling several Gbps of traffic.

Carrier-grade Ethernet is a term for a number of industrial and academic initiatives that aim to equip Ethernet with the transport features it is missing [35]. There is a lot of research and evaluative work being done related with carrier Ethernet nowadays [35-39, 41-43]. One of the most important reasons behind the development of carrier Ethernet is the growing demand for high-bandwidth applications at increasingly lower costs. However, the introduction of Ethernet as a packet carrier technology introduces many challenges which have to be addressed in order to be able to replace the circuit-switched SONET/SDH technology.

Many are putting their efforts into transforming Carrier Ethernet to fulfil the Next Generation Network (NGN) service and transport requirements. The NGN has been developed by telecom carriers for more than 10 years and its concept is to allow simultaneous delivery of packet-based and circuit-based services. Metro Ethernet improves operational efficiency and can be a launch pad for newer services; from the carriers' point of view, it gives service providers the ability to offer higher revenue services. Moreover, Ethernet has got "sanitized" ("SONETized and ATMized") to acquire some of the proven carrier grade characteristics from SONET/SDH and ATM technologies [19].

In the first part of this chapter we will discuss the native Ethernet technology and its physical characteristics. Furthermore, the new standards developed with the aim of evolving to carrier-grade Ethernet will be discussed later on while focusing on the Virtual LANs and Provider Bridge Backbone technologies which are important for our proposed solution in chapter 5. The second part of the chapter gives a detailed overview of the loop-free protocols based on the Spanning Tree Algorithm (STA).

3.2 Native Ethernet

The Ethernet local area, access and metropolitan networks are specified by the IEEE 802.3 standard [38]. It employs the Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Medium Access Control (MAC) protocol for all the specified speeds of operation. This characteristic has allowed the adaption of the protocol for new high-speed technologies as optical fibre. However, there are only two modes of operation over the shared medium: *half-duplex* and *full-duplex*. If two or more stations share the common transmission medium in a *half-duplex* mode, the stations will implement the original CSMA/CD. It specifies that a node will wait for an idle period on the medium (carrier sense) and initiate the transmission while still listening for message collisions (collision detection). In case of a collision, the stations will continue transmitting for a predefined period of time in order to ensure the propagation of the collision throughout the system. Afterwards, there will be no transmissions on the medium while each station waits for a random *backoff* time before attempting to retransmit. The *full-duplex* mode of operation allows the simultaneous communication between two stations using a point-to-point media or a dedicated channel. In this case it is implied that the CSMA/CD is not required because the nodes do not need to monitor and react to the activity on the medium as there would be no contention. These

modes of operation are important when implementing the spanning tree protocols, as we describe in section 3.4, because they affect the convergence time of the network topology. Furthermore, new Ethernet standards, as 10GbE, implement only the *full-duplex* mode of operation.

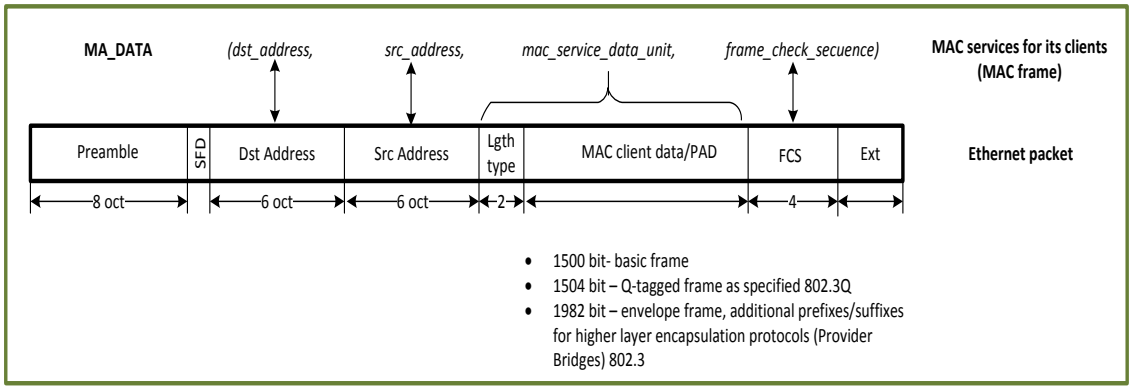

Figure 3.1 Ethernet packet format and MAC service mapping.

The MAC sub-layer provides services to the MAC clients, which can be the Logical Link Control sub layer, Bridge Relay Entities or other LAN services in the Data Link Layer [21], [38]. Furthermore, these services (fig. 3.1) are mapped into three different MAC frames: a basic frame, a Q-tagged frame and an envelope frame. However, the frames use the same Ethernet packet format as shown in figure 3.1:

- The 7 octets Preamble field is used to allow the physical layer signalling (PLS) circuitry to reach its steady-state synchronization with the received packet's timing. It is followed by the sequence 10101011 (SFD) to specify that the MAC frames starts immediately after that.
- The Destination Address field specifies the destination address of the MAC frame. The Source Address field identifies the station from which the MAC frame was initiated. Each address field is 48 bits in length and may be a unicast, broadcast or multicast address.
- The Length/Type field is 2 octets and depending of its value it is associated with the mutually exclusive MAC client data Length (≤1500 bytes) or the MAC client protocol (≥1536).
- The correct operation of the CSMA/CD protocol needs a minimum MAC frame size in order to sense collisions [21]. This is achieved by padding the client's data through the PAD field.

- The Frame Check Sequence is 4 octets and is based on the Cyclic Redundancy Check (CRC) encoding polynomial. The calculation is based only on the MAC frame bits.
- The carrier extension field is used in half-duplex mode only and is needed to successfully achieve the contention resolution when operating at high speeds (e.g. 1Gbps).

The data link layer was designed with the assumption that the communicating nodes are connected to a common link [21]. This characteristic implies that a data link protocol should be designed to carry a packet of information across a single hop. Ethernet and the 802.3 MAC protocol are specifically designed for this, which is why they lack the header fields for a connection-oriented and multihop infrastructure. Functionalities such as fragmentation, hop count, congestion feedback and next-hop are delegated to upper layers. This not only simplifies the design of the protocol and the management of layer 2 devices, but also enhances their performance. Furthermore, the flat 802 addressing scheme requires that the stations should be served by the layer 2 devices independently of their address as opposed to layer 3 addresses which have a topological meaning. Ethernet switches/bridges provide this plug-and-play capability through their forwarding logic implementation [21], [22], [25]. The basic functionality of a switch is identical to that of a transparent bridge on a per-VLAN basis. A transparent bridge has these characteristics [25]:

- It learns addresses by "listening" on a port for the source address of a device. When a source MAC address is read in frames coming into a specific port, the bridge assumes that the frames destined for that MAC address can be sent out of that port. The bridge builds its forwarding table (Content Addressable Memory) that records which source addresses are seen on which port. A bridge is always listening and learning MAC addresses through this process.
- It must forward all broadcast packets out of all its ports, except for the port that initially received the broadcast.
- If the bridge does not have information on the destination address, it forwards the frame out of all ports, except for the port that initially received the frame. This is called a unicast flooding.

The lack of a hop-count field in the layer 2 header makes the network prone to broadcast and unicast flooding storms. In addition, as with traditional shared Ethernet, transparent bridges inherently lack the capability to provide redundancy because of the possibility of

creating bridging loops. A bridging loop occurs when there is no Layer 2 mechanism, such as the time-to-live, to manage the redundant paths and stop the frame from circulating endlessly. This circulation overloads the nodes and might bring down the network. The most important method of implementing and managing redundancy in a layer 2 network is the spanning tree algorithm and its related protocols as we will discuss in the section 3.4.

3.3 VLAN and Carrier Ethernet technologies

IEEE has developed a number of standards providing enhancements to the original Ethernet standards and aiming toward a carrier-grade Ethernet technology. These standards include:

• 802.1Q: Virtual LAN
• 802.1ad: Provider Bridging
• 802.1ah: Provider Backbone Bridging
• 802.3ah: Ethernet in the First Mile (with OAM)
• 802.1ag: Connectivity Fault Management (OAM)

3.3.1 VLAN tagging and QoS

Basically, a virtual LAN is really no different from a LAN. It is the part of the network over which a broadcast or multicast packet is delivered, known as a *broadcast domain*. The difference between a VLAN and a LAN is in the encapsulation. Virtual LANs allows us to have separate LANs among ports on the same switch, which would act as two separate bridges. As Ethernet switches have always aimed at switching IP traffic, it is because of some of the problems of IP routing that VLANs were created and aimed to address. Such problems are:

1. The IP broadcast traffic within a LAN can cause congestion and single node misbehaviour may lead to broadcast storms.
2. Routing IP traffic compared to switching Ethernet frames is rather slow and expensive as the diameter of the LAN grows in size and geographical coverage.
3. The management of the IP addressing scheme as all the nodes in a LAN share the same range is made easier by employing DHCP and VLANs.

However, as specified in the 802.1Q standard [38] the usage of VLANs aim to offer the following benefits:

1. VLANs facilitate easy administration of logical groups of stations that can communicate as if they were on the same LAN. They also facilitate easier administration of adding, removing and changing the members of these groups.

2. Traffic between VLANs is restricted. Bridges forward unicast, multicast, and broadcast traffic only on individual LANs that serve the VLAN to which the traffic belongs. In our case, it translates in the need of maintaining an independent native VLAN spanning tree that would allow the interconnection of all the nodes in an OpMiGua infrastructure.

3. VLANs maintain compatibility with existing bridges and end stations because of the implementation of an untagged version of the frame.

4. If all Bridge Ports are configured to transmit and receive untagged frames, bridges will work in plug-and-play IEEE802.1D mode allowing all end stations to communicate throughout the network.

A VLAN tag is shown in figure 3.2 and includes these elements [28]:

- **Tag Protocol Identifier (TPID)**. It is two octets in length and includes an Ethernet Type value that is used to identify the frame as a tagged frame and to select the correct tag decoding functions.
- **Tag Control Information (TCI).** It is two octets and is used to identify the traffic circulating on the VLAN; it basically indicates the origin and destination of the frame transmission. The first three bits of the VLAN tag indicate the priority of the traffic that is included in the packet. This allows for some basic QoS assurance, which ensures that critical data can pass through the network quickly with as little delays as possible. The value of this field can be generated at the end station and updated on every switch (VLAN-aware) along the way as well. The fourth bit is a canonical format indicator (CFI), which is used mainly for 802.3 source routing information. The last 12 bits comprise the VLAN identifier (VID), which enables the creation of 4094 operational VLANs.

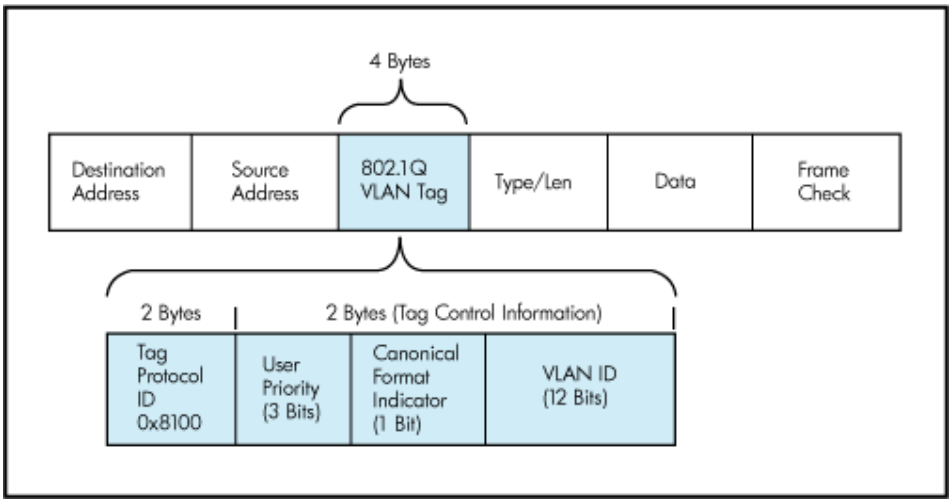

Figure 3.2 802.1Q frame format with VLAN tagging.

The user priority bits as specified in the IEE 802.1p standard and later enhanced in [38] provide QoS-aware switching at the MAC layer. The use of three bits limits the traffic classification in eight classes:

1. Network Control (NC) is characterized by a guaranteed delivery requirement to support configuration and maintenance of the network infrastructure.
2. Internetwork Control (IC) in large networks comprising separate administrative domains there is typically a requirement to distinguish traffic supporting the network as a concatenation of those domains from the Network Control of the immediate domain.
3. Voice characterized by less than 10 ms delay.
4. Video characterized by less than 100 ms delay or other applications with low latency as the primary QoS requirement.
5. Critical Applications characterized by having a guaranteed minimum bandwidth as their primary QoS requirement.
6. Excellent Effort or "CEO's best effort" is the best-effort type services delivered to the most important customers.
7. Best Effort for default use by not prioritized applications with fairness only regulated by the effects of TCP's dynamic windowing and retransmission strategy.
8. Background bulk transfers and other activities that are permitted on the network but that should not impact the use of the network by other users and applications.

The standard allows the use of different numbers of queues at each node allowing an ongoing user-traffic to user-priority classes mapping on the network. Table 3.1 and 3.2 show

the mapping of the traffic type to traffic classes and assigning them to the queues available in a node.

Table 3.1 Traffic type and user-priority

Priority	Acronym	Traffic type
1	BK	Background
0 (default)	BE	Best Effort
2	EE	Excellent Effort
3	CA	Critical Applications
4	VI	Video < 100ms latency and jitter
5	VO	Voice < 10ms latency and jitter
6	IC	Internetwork Control
7	NC	Network Control

Table 3.2 Defining traffic types

Number of queues	Defining traffic type							
1	BE							
2	VO				BE			
3	NC		VO		BE			
4	NC		VO		CA		BE	
5	NC	IC	VO		CA		BE	
6	NC	IC	VO		CA		BE	BK
7	NC	IC	VO		CA	EE	BE	BK
8	NC	IC	VO	VI	CA	EE	BE	BK

3.3.2 Evolution of Ethernet hierarchy

In figure 3.3 is shown the evolution of the Ethernet frame hierarchy based on the IEE standardized frame formats.

3.3.2.1 Provider Bridging 802.1ad

This method is usually referred to as Q-in-Q and added an additional service provider VLAN ID (S-tag) to the customer's Ethernet frame. The customer's VLAN ID (C-tag) is not modified while the S-tag identifies the service in the provider's network. The use of the tag as a service identification means that each service instance will need a different S-Tag. Furthermore, since the S-Tag consists of a 12-bit tag, provider bridges have the same scalability issue that allows the creation of a maximum of 4094 services instances. Also, the standard specifies the creation of different spanning trees for each instance. However, even if these spanning trees fall under the same common one, it still is not scalable as we will discuss

in the STP part of this chapter. An interesting approach is using the S-Tag as an MPLS label for creating connection-oriented paths through VLAN Cross-connect [56].

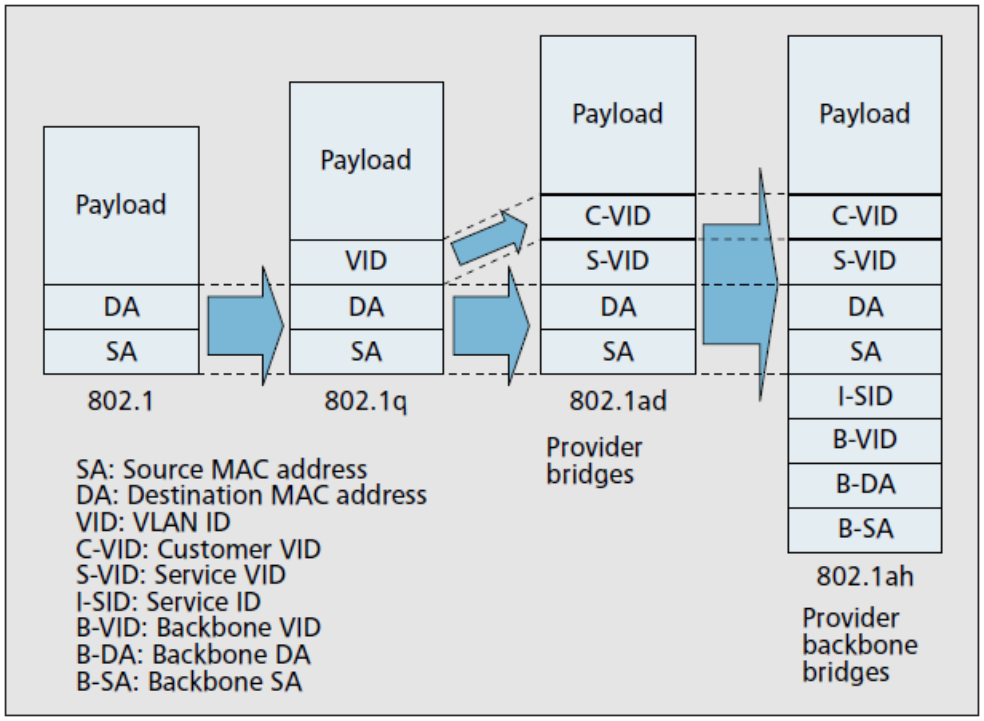

Figure 3.3 Evolution of Ethernet hierarchies, taken from [39].

3.3.2.2 Provider Backbone Bridges and PBB-TE

This method is usually referred to as MAC-in-MAC and was standardized in 2008 by IEEE 802.1ah [40]. The 802.1ah frame adds a second MAC encapsulation to any 802.1 frame type which is the customer's payload. This approach allows a level of hierarchy that is not provided by the Q-in-Q tagging. Now the provider's network is completely isolated from the customer's and it is a significant step toward making Ethernet suitable for carriers [39].

802.1ah also introduces a new I-SID service instance identifier of 24 bits. This tag field is proposed as a solution to the scalability limitations encountered with the 12 bit S-VID defined in Provider Bridges. The bridges operate the same way as the traditional Ethernet bridges: service is still connectionless and flooding is used when destination MAC addresses are not recognized. Furthermore, what is most important for us, the spanning tree protocols are still used to prevent loops. VLAN tags are reserved on a network, rather than a per-port basis, by means of proprietary VLAN trunking protocols.

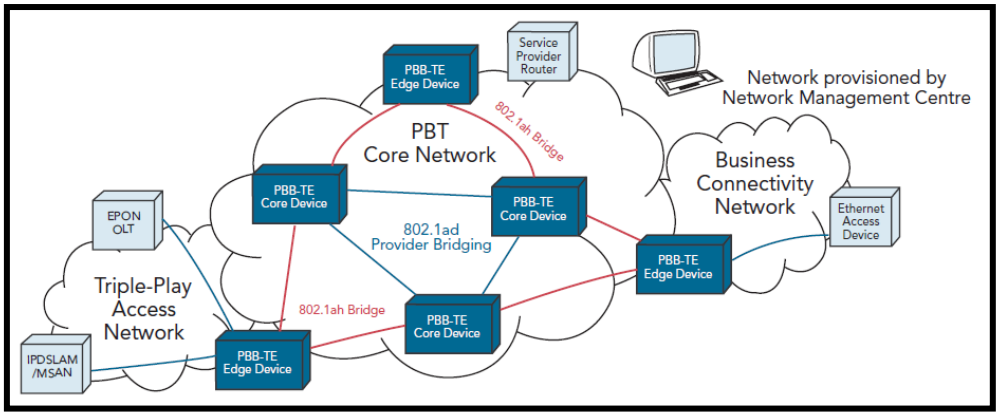

Figure 3.4 Example carrier network applications, taken from [42].

PBB-TE builds upon these standards to provide a network solution designed specifically for transport applications [43]. It creates an independent connection-oriented packet-switched transport layer (see Figure 3.4). This allows various services to be transported transparently through the network. However, what is an important characteristic for our work, it turns off some of the native Ethernet features to realise its MAC addresses management through the control plane. In this case the Spanning Tree Protocols are not used and it differs from the main objective of our thesis.

3.4 Spanning Tree Algorithm and Protocols

The Spanning Tree Protocol (STP) as conceived by Perlman [22] is based on the graph theory. A spanning tree of a graph *G* is the sub graph of *G* that is a tree and contains all the vertices of *G* (spanning). In the graph G we use *n* to indicate the number of vertices and *e* for the number of edges. Based on the Prufer theorem [26] the number of spanning trees in G_n is n^{n-2}. Furthermore, the number of nonisomorphic spanning trees in a general graph is computed by the recursive formula $\gamma(G)=\gamma(G-e)+\gamma(G/e)$, where G/e is the resulting graph after removing edge e. The STP is a self-stabilizing distributed algorithm based on the minimum spanning tree of a weighted graph and the protocol uses the links' cost as its primary weight. It has a deterministic behaviour that provides the desirable reproducibility, configurability and predictability properties for the network topology [22], [33].

3.4.1 Spanning Tree Protocol

In a network it is always beneficial to accomplish dependability based on the physical redundancy of the network nodes and interconnections. STP is the mechanism employed in

Ethernet switches to configure, set-up and manage a loop-free active layer 2 path across the network and provide redundancy in case of failure. The distributed spanning tree algorithm (STA) runs on each switch to activate or block redundant links. To categorize these links, the STA chooses a reference point (*the root switch*) in the network and determines the paths to that reference point from each node of the network. In case that there are multiple redundant paths to the root, it decides which path forwards data frames and which paths are blocked. This effectively finds and blocks the redundant links within the network in order to create the loop-free topology. Spanning tree standards often refer to a "bridge" but to be consistent throughout the thesis, we will use the term switch for all the devices exchanging spanning tree information at layer 2.

3.4.1.1 Switches and ports' roles

The IEEE 802.1D STP standard [27] specifies the encoding and the structure of the information exchanged between the switches through Bridge Protocol Data Units (BPDU).

Table 3.3 Bridge Protocol Data Unit

Byte	**Field**
2	Protocol ID
1	Version
1	Message type
1	Flags
8	Root ID
4	Cost of path of all the links from the transmitting switch to the root
8	Bridge ID the lowest bridge ID in the topology (*priority 4+vlan 12+MAC48*)
2	Port ID
2	Message age
2	Max age
2	Hello Time
2	Forward delay

BPDUs contain the required information for the STP establishment, management and configuration. The Type field for the BPDU message in the Ethernet packet is 0x00 and it uses the multicast MAC address 01-80-C2-00-00-00.

There are three roles for switches and ports in a spanning tree:

1. **Root**. The root is the switch with the smallest ID and is elected dynamically. Every switch starts the algorithm assuming that it is the root until it receives BPDUs with

lower switch IDs. When a topology change occurs the root sends messages throughout the tree so that the content addressable memory (CAM) table of every switch in the network is flushed in order to learn and provide a new path for the end host devices. The ports of the root are always forwarding data and BPDUs.

2. **Designated**. The switch in a LAN segment that provides the best path toward the root is the designated switch for that segment. The port of the switch which is providing this path is the root port of the switch while the other ports that provide connectivity for the other switches are designated ports.
3. **Blocking**. The port is not active in the network topology.

The algorithm takes the input from the information carried in the BPDU and follows these steps:

1. Elect a single switch, among all the switches on all the LANs, to be the *Root Switch.*
2. Each switch computes the best path from itself to the root.
3. Elect the *Designated Switch* based on step 2; this switch will forward packets from that LAN toward the *Root Switch*.
4. Choose a port (*root port*) that gives the best path from the switch to the *Root Switch*.
5. Select ports to be included in the spanning tree. The ports selected will be the *root port* and any other ports connected to the segment on which the switch has been elected as *Designated Switch*.

3.4.1.2 Port states

The ports of the root are always forwarding BPDUs and data while for a non-root switch; the spanning tree determines four port states [21], [25], as shown in figure 3.5:

1. **Blocking**: The non-designated port is not part of the active spanning tree topology and does not forward either BPDUs or data frames. However, it receives BPDUs to determine the location and root ID of the root switch and which port roles (root, designated, or non-designated) each switch port should assume in the final active STP topology in case of failure. The port waits 20 seconds in this state (*max age*).

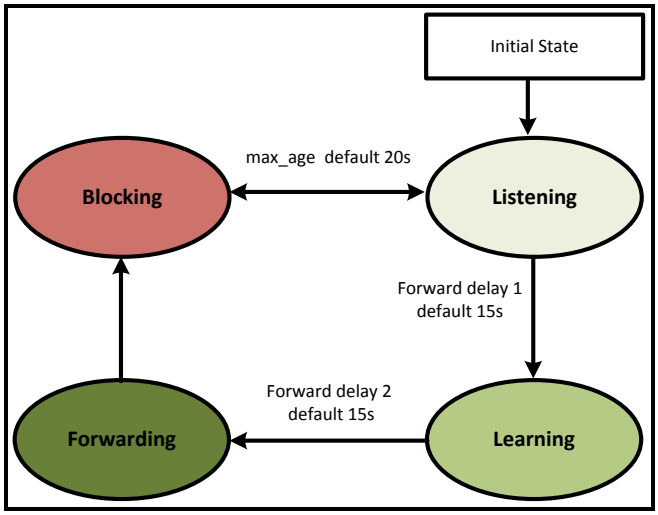

Figure 3.5 STP Port state transitions.

2. **Listening**: Spanning tree has determined that the port can participate in the frame forwarding according to the BPDUs that the switch has received. The switch port is receiving BPDUs, transmitting its own BPDUs and informing adjacent switches that the switch port is preparing to participate in the active topology. By default, the port spends 15 seconds in this state (*forward delay*). These timers in between states are used to prevent transition loops while the network topology is converging.
3. **Learning**: The Layer 2 port prepares to participate in the data frame forwarding and begins to populate the CAM table. The port is still sending and receiving BPDUs, while staying in this state for 15 seconds (*forward delay*).
4. **Forwarding**: The Layer 2 port is considered part of the active topology. It forwards frames and also sends and receives BPDUs.

The timers carried in BPDUs are very important for STP because they are used to determine the transitional period in-between states, determine the availability of neighbouring switches and caching time of MAC addresses in the forwarding table:

- **Hello timer:** determines how often the root switch sends configuration BPDUs to inform the nodes about the liveliness of the spanning tree.
- **Maximum Age (Max Age):** Indicates to the switch how long to keep ports in the blocking state before starting the transition to become part of the active topology.
- **Forward Delay (Fwd Delay):** Is a tuneable parameter needed to prevent transient loops and to transition port states in accordance with the network convergence.

The root bridge informs the non-root bridges of the time intervals to use and the STP timers can be tuned based on the network size. Non-root bridges place various ports in their proper roles by listening to BPDUs as they come in on all ports and may trigger the re-computation of the spanning tree. Receiving BPDUs on multiple ports indicates a redundant path toward the root bridge. The switch looks at the following components in the BPDU in order to decide the state of the ports:

1. Lowest path cost
2. Lowest sender BID
3. Lowest sender port ID

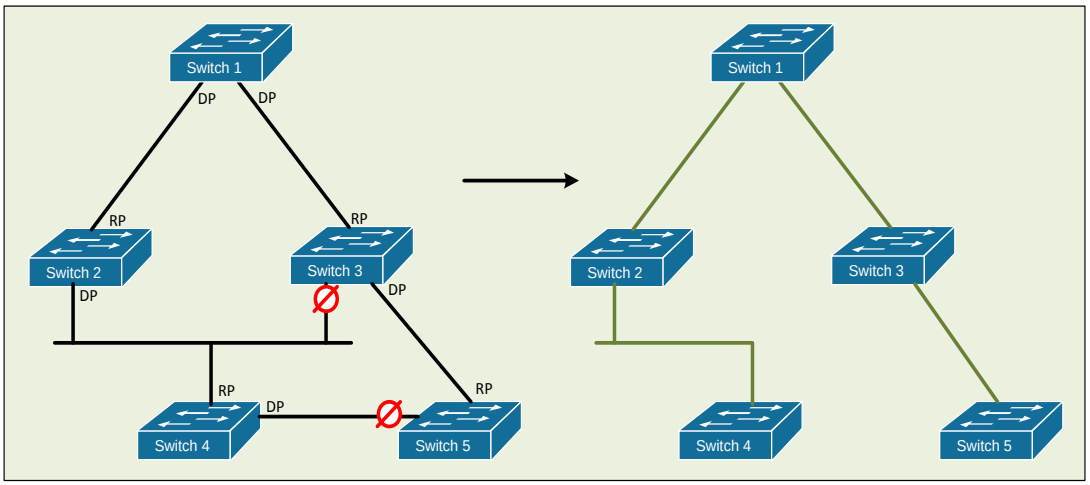

Figure 3.6 An example of Spanning Tree Protocol convergence.

The path cost is calculated on the basis of the link speed as defined in [27] and the number of links the BPDU has traversed. Ports with the lowest cost are eligible to be placed in forwarding mode while the other ports that are receiving BPDUs will continue to stay in a blocking state. If the path cost and sender BID are equal, as with parallel links between two switches, the switch uses the port ID. In this case, the port with the lowest port ID forwards data frames, and all other ports continue to block data frames. Each bridge advertises the spanning tree path cost in the BPDU. This spanning tree path cost is the cumulative cost of all the links from the root bridge to the switch sending the BPDU. The receiving switch uses this cost to determine the best path to the root bridge.

Figure 3.6 illustrates an example of a spanning tree topology with port roles based on the STP decision process. The links throughout the network have the same cost, which implies that the algorithm translates in a shortest spanning tree. Switch 5 receives BPDUs from switch 3 and 4. As a result of the shortest path computation, the lowest cost value will be

received from switch 3. The port connecting to this segment will be the root port in a forwarding state while the other will transit into the blocking state. STP selects one designated switch per segment to forward data traffic (switch 2); while the other switch ports on the segment become non-designated ports (switch 3). They continue receiving BPDUs while discarding the data traffic to prevent loops. The BPDU exchange in a generalized scenario yields the following results:

- Election of a root bridge as a Layer 2 topology point of reference.
- Determination of the best path to the root bridge from each switch.
- Election of a designated switch and corresponding designated port for every switched segment.
- Removal of loops in the switched network by placing some switch ports to a blocked state (link pruning).
- Determination of the "active topology" for each instance or VLAN running STP.

3.4.2 Link failure

The active topology is the final set of communication paths that are created by the switch ports that forward frames. In case of a link failure, after the active topology has been established, the network must reconfigure the active topology using Topology Change Notifications (TCNs). The TCN BPDU is generated when a bridge discovers a change in topology, usually because of a link failure, switch failure, or a port transitioning to the forwarding state. The TCN BPDU is set to 0x80 in the Type field and is forwarded on the root port toward the root switch. The upstream switch acknowledges the received BPDU through a Topology Change Acknowledgment (TCA) and sends the message to its designated switch. In the Flag field (Table 3.3), the least significant bit is for the TCN while the most significant bit is for the TCA. This process repeats until the root bridge receives the notification and sets the TCN flag in its BPDU. This upstream step-by-step approach minimizes the protocol overhead as compared to broadcasting the change throughout the network. However, it is the main problem for the slow convergence time of STP compared to RSTP as we will discuss in section 3.4.3.

The 802.1D STP standard was developed long before VLANs were introduced and its implementation would create a different spanning tree instance for each VLAN. This would result in an increased network bandwidth overhead. Also the root switch becomes a possible single point of failure in the network because of the increased switch memory usage and

processing overhead. Another important drawback of STP is that its convergence time in case of failure is approximately 30-60 seconds [24]. The introduction of a high-speed physical medium, such as the optical fibre, would result in a critical amount of data loss. To overcome these STP bottlenecks, two new IEEE standards were introduced, RSTP (802.1w) [27] and later MSTP (802.1s) [28]. Rapid Spanning Tree Protocol (RSTP) provides much faster convergence, while Multiple Spanning Tree Protocol (MSTP) allows the creation of multiple instances of spanning tree making use of the redundant resources and efficiently managing VLANs.

3.4.3 Rapid Spanning Tree Protocol

RSTP (802.1w) supersedes 802.1D, while still retaining backward on a per-port basis [27]. It requires a *full-duplex* point-to-point connection between adjacent switches to achieve fast convergence. As a result, RSTP cannot achieve fast convergence in half-duplex mode and employs STP in such cases. The spanning tree algorithm is essentially the same as described in the previous section, while the main differences are the port states and additional port roles [30]. RSTP divides the blocking port role of STP in alternate and backup port roles. This differentiates between, respectively, the redundant connection through another LAN segment and the redundant connection to the same designated switch on the segment. Furthermore, it combines the blocking and learning states of a port in a single discarding state which allows the faster transition of the ports to the forwarding state.

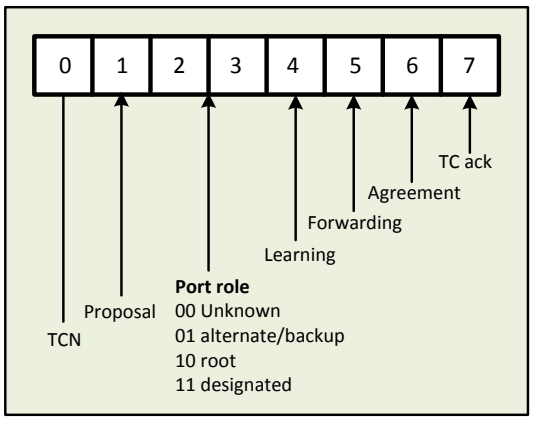

Figure 3.7 RSTP BPDU flag usage.

The RSTP BPDU format is the same as the IEEE 802.1D BPDU format, except that the Version field is set to 2 to indicate RSTP, and the Flags field makes use of all 8 bits, as shown in figure 3.7.

RSTP is proactive and therefore it provides rapid convergence following a failure or during the re-establishment of a switch, switch port, or link. The topology changes trigger the transition process through explicit handshakes between adjacent switches also called the proposal/agreement synchronization process. The BPDUs are sent regardless of the root BPDUs which allows for faster and localized failure detection in the network. The enhanced reaction speed to the topology changes is based on the convergence on a link-by-link basis and is not relying on timers, as in STP, for transitioning between port states. Figure 3.8 illustrates how rapid transition is achieved through the proposal/agreement protocol, as follows:

1. Switch 3 has a path to the root via switch 4 and switch 2. A new link is then created between the root (switch 1) and switch 3 and both ports are in blocking state until they receive a BPDU from their counterpart. When a designated port is in a discarding or learning state it sets the proposal bit on the BPDUs it sends out. This is what happens for port P0 of the root bridge.

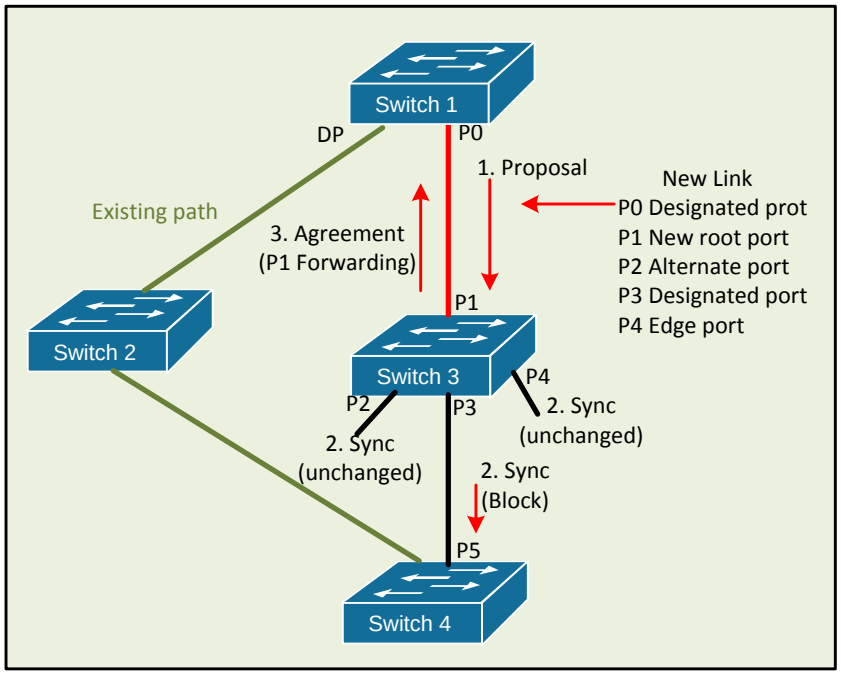

Figure 3.8 RSTP transition examples.

2. Switch 3 observes that the proposal BPDU has a superior path cost. It blocks all non-edge designated ports other than the ones over which the proposal-agreement process are occurring. This synchronization operation prevents switches downstream from causing a loop during the proposal-agreement process.

3. Switch 3 sends an agreement that allows the root bridge to place the root port P0 in the forwarding state. Port P1 becomes the root port for switch 3.

After switch 3 and the root are synchronized, the proposal/agreement process continues on switch 3 out of all of its downstream-designated non-edge ports, in our case with switch 4 and so on. The handshake propagates throughout the network and quickly restores connectivity because the TCN BPDU is flooded from the upstream node. It is not necessary to wait for the message to reach and be flooded by the root. Furthermore, the CAM tables are flushed when receiving the notification on each switch. In the case of using STP, it is needed to maintain the network topology for a *max_age+forwarding delay* amount of time. In general, RSTP needs only 1-3 seconds for a global network topology re-establishment [24]. However, RSTP still computes spanning tree instances for each VLAN leading to the scalability issue we previously discussed with STP. Furthermore it still shares other drawbacks with STP, such as network underutilization, congestion near the root, and no load balancing.

3.4.4 Multiple Spanning Tree Protocol

The most recent enhancement to STP is the Multiple Spanning Tree Protocol (MSTP), as defined in IEEE 802.1s [28], [29]. MSTP partitions the topology into different regions that are connected together by a common Spanning Tree, called the Common and Internal Spanning Tree (CIST). The regions in MSTP are instances of RSTP each with their own regional spanning tree which forwards the traffic of one or more VLANs. Recalling the BPDU frame format (Table 3.1), the VLAN ID (VID) is part of the bridge ID. The regional roots are connected to the common root from the CIST, as shown in figure 3.9.

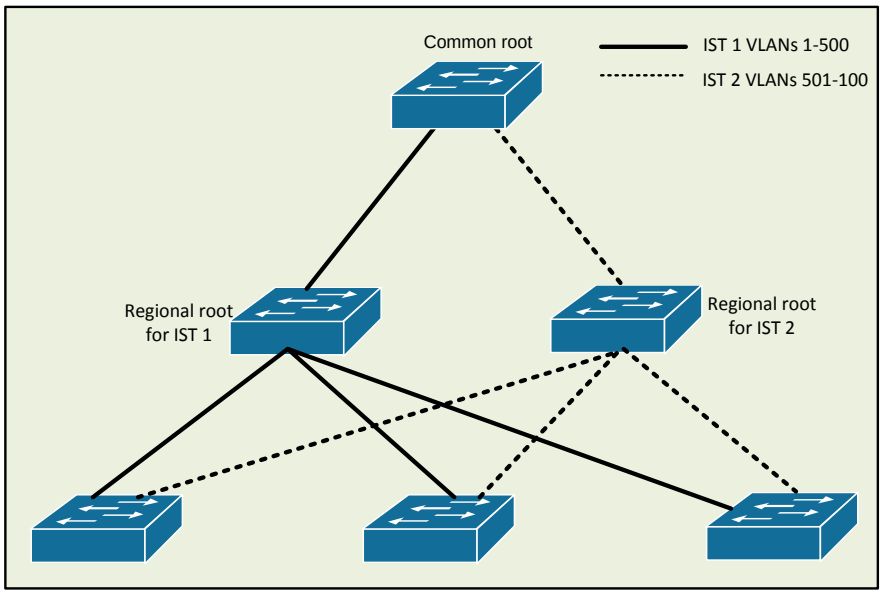

Figure 3.9 An example of an MSTP configuration.

The main advantages of MSTP are:

1. It reduces the total number of spanning tree instances in order to match the physical topology and simplifies the management of the per-VLAN STI introduced by RSTP.
2. By distributing and directing the traffic over different VLANs, it is possible to achieve a more balanced load across the network and achieve higher utilization of the redundant interconnections. The ST instances will be equal to the number of total redundant paths available. For our work it may translate in assigning different VLANs for the GST and SM traffic.
3. MSTP lessens the processing load of the root node and avoids the root bottleneck problem.

All these properties are important in an OpMiGua network since MSTP supports the main benefits of our hybrid architecture. To provide the logical assignment of VLANs to spanning trees, each switch running MSTP has a single configuration that consists of three attributes:

- An alphanumeric configuration name (32 bytes);
- A configuration revision number (two bytes);
- A 4096-element table that associates each of the potential 4096 VLANs with a given instance. However, the table is sent as a digest to prevent network overhead and to allow switches on the same region to learn about their neighbouring switches.

The group of switches part of a common MSTP region share the same configuration attributes. The proper propagation of this configuration through the region is possible only by

using the command-line interface (CLI) or Simple Network Management Protocol (SNMP). Each region is treated by the CIST root as a single switch and to facilitate the virtualization, MSTP distinguishes between the internal connectivity MSTI BPDUs and external STI connectivity BPDUs. Furthermore, the edge regional nodes are connected to the CIST root through ports that have a new Master role. Otherwise, the spanning tree algorithm, port states and port roles are the same as for RSTP. The extended MST BPDU parameters and format is shown in figure 3.10.

	Octet
Protocol Identifier	1–2
Protocol Version Identifier	3
BPDU Type	4
CIST Flags	5
CIST Root Identifier	6–13
CIST External Path Cost	14–17
CIST Regional Root Identifier	18–25
CIST Port Identifier	26–27
Message Age	28–29
Max Age	30–31
Hello Time	32–33
Forward Delay	34–35
Version 1 Length = 0	36
Version 3 Length	37–38
MST Configuration Identifier	39–89
CIST Internal Root Path Cost	90–93
CIST Bridge Identifier	94–101
CIST Remaining Hops	102
MSTI Configuration Messages (may be absent)	103–*39 + Version 3 Length*

Figure 3.10 MST BPDU parameters and format, taken from [29].

The protocol identifier value specifies if the BPDU is that of an STP, RSTP or MSTP instance. The extended fields related to MIST are processed only by MST entities and are encapsulated only within a region. Otherwise the CIST information fields are treated the same as RSTP information outside the region.

The main drawbacks of employing MSTP are derived from its complexity. The protocol requires additional human interaction and also interaction with legacy switches can be a challenge [29]. In chapter 5 we propose the network architecture solution when

employing spanning-tree protocols with support for VLANs after analysing in chapter 4 the problems xSTP introduce in an OpMigua architecture.

3.5 Current work on Ethernet loop-free protocols

A lot of research work has been focused on the enhancement of the current spanning tree protocols [23, 24, 31-33, 42] in order to meet the Metro Ethernet Networks (MENs) resiliency, load management and QoS requirements. Huynh et al. [32] introduces the Cross-Over Spanning Trees (COST) protocol which is based on the MSTP logic, but allows the traffic flow to switch between different spanning tree instances while *en-route* to its destination. The simulation results show that this method achieves considerably higher load balance of the total throughput and also has a slightly advantageous convergence time. Furthermore, Huynh et al. [31] presents another modification to the ST protocols, called the spanning tree elevation protocol (STEP) that has a similar design with COST but includes support for QoS by means of traffic policing and service differentiation.

Another interesting approach for cycle-breaking algorithms is presented in [34]. The method is not based on the spanning tree algorithms but on the theory of turn prohibition. The tree-based turn prohibition protocol (TBTP) is compatible with the 802.1D standard. The simulations on a variety of graph topologies show that the TBTP algorithm can lead to one order of magnitude improvement over the spanning-tree protocol with respect to throughput and delay metrics.

In [24] and [41] is given an overview of the current resilience technologies in Ethernet, implemented by proprietary solutions or standardization bodies. The Ethernet Automatic Protection Switching (EAPS), Metro Ring Protocol (MRP) and Rapid Ring Spanning Tree Protocol (RRSTP) are of interest for our work when looking into the OpMiGua Ring network resiliency.

The Resilient Packet Ring (RPR) [45] architecture is another IEEE standard that aims at improving the xSTP convergence time in the case of optical transmission medium. The RPR recovery times are comparable to SONET/SDH. However, since the topology is specifically a dual ring and also is not compatible with most of legacy Ethernet switches, it is not in the focus of this thesis. The Ethernet Operation, Administration and Management standard OAM [46] is focused on the connectivity failure notification protocols, but still relies on xSTP for recovery. The latest ITU-T standards for protection switching are G.8032 Ring Protection Switching and G.8031 Ethernet Linear Protection Switching. These standards are not

supported by the current Ethernet switches in the market. In addition the G.8031 is still undergoing modifications but G.8032 is an option to xSTP for optical ring networks. Their implementations are proprietary and not compatible with native Ethernet switches, that is why we leave the evaluation of these protocols for further future work and focus on xSTP.

Chapter 4

Problem Analysis

As discussed in chapter 3, Ethernet employs protocols based on the Spanning Tree algorithm to create the loop free network connectivity. The distributed algorithm runs on all the nodes of the network blocking the available redundant paths and creating a spanned tree. This approach not only affects the network performance in terms of load-balancing and resource utilization, but what is more important for us, it influences the GST and SM traffic in OpMiGua. In this chapter we will analyse the problems that arise when combining Ethernet and OpMiGua in terms of the spanning tree protocols. We consider different network topology connectivity of the Ethernet-OpMiGua node combination. First we examine the general scenario with single dedicated connections for GST/SM and the respective xSTP protocol problems. Furthermore, we extend the analysis in a full-mesh scenario for both GST/SM connections. The proposed architectural solution is given in chapter 5 together with a discussion of its advantages and disadvantages. Based on the required connectivity rules derived from these two chapters, we design an integrated OpMiGua node that employs an Ethernet switch instead of its OPS part in chapter 6.

4.1 Single Dedicated-port for GST and SM traffic

In this scenario, Ethernet switches are connected to an OpMiGua network and are aggregating traffic while assigning it to the GST or SM classes based on the QoS policies in the switches, as shown in figure 4.1. The switches are connected to the OpMiGua nodes with two gigabit interfaces each assigned to a coarse-grained QoS class; one is GST for the high-

priority traffic while the other is SM traffic. Additionally, the switch has many input interfaces (e.g. 100 Mbps interfaces) and aggregates the traffic in a finer-grained QoS matching the mentioned aggregated outputs. If switch 1 is switching both GST and SM traffic towards switch 3, there would be two existent paths between the two nodes (figure 4.1). The GST traffic is circuit-switched by the OpMiGua nodes in a direct switch1-switch3 logical connection, by-passing the intermediate switch. In a Static Wavelength Routed Optical Network (SWRON) each connection is assigned a different wavelength.

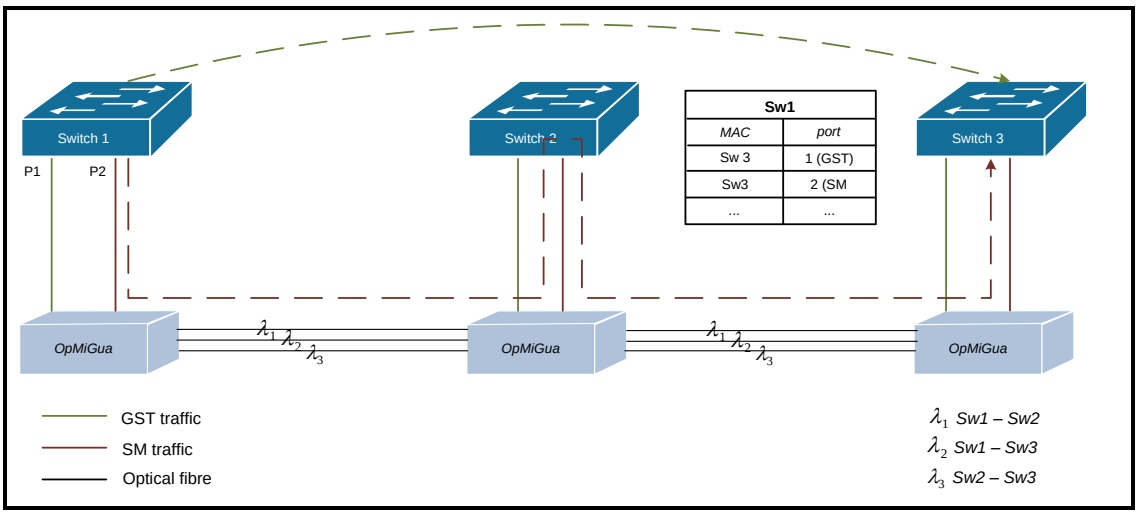

Figure 4.1 A simplified OpMiGua network topology.

The SM traffic is packet-switched in all the switches based on the information carried on the header of each packet. Currently the traffic is switched into two paths: the direct path for the guaranteed GST and a path going through intermediate switches also going to switch 3 for SM traffic. However, the spanning tree algorithm does not allow loops and will block one of the ports.

A detailed example of how the spanning tree converges is important to understand the problems introduced by the way OpMiGua presents the connections to the Ethernet layer. In order to build the spanning tree of the network topology shown in figure 4.2, we make the following assumptions:

1. The root switch will be the one with the highest processing power (arbitrarily sw1). The root will be manually configured using the priority bits avoiding the randomness of the STP selection which is based on the lowest MAC (Root ID).

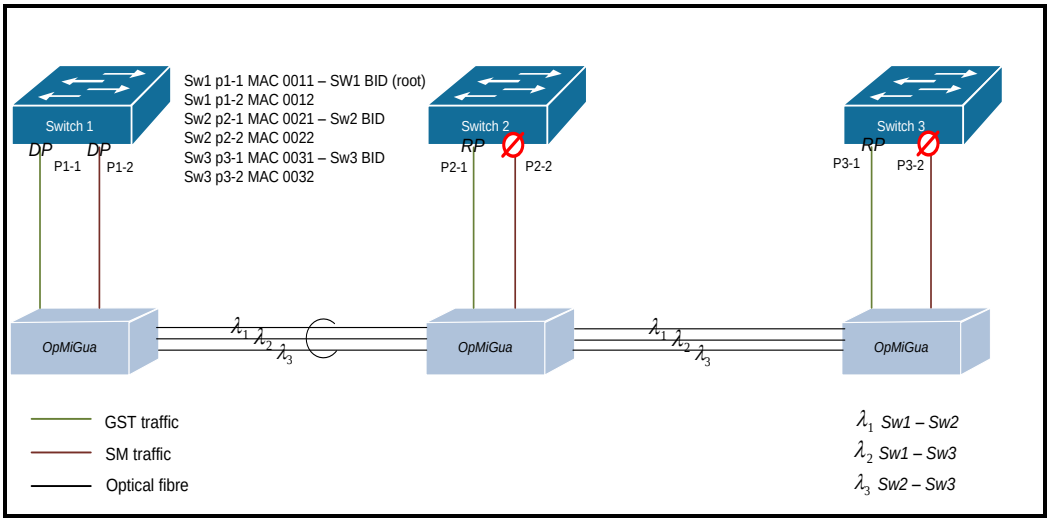

Figure 4.2 Three node network topology with dedicated ports for GST/SM traffic. The Ethernet switches consider the underlying OpMiGua network as transparent and are logically connected directly to each-other.

2. The port ID of the interfaces is important in the *Decision Making* process of the algorithm. This is the reason why there is a factor of randomness on which traffic (GST or SM) will be blocked by STP. In our scenario we have selected an incremental value for the port ID (MAC address of the switch port) to derive the spanning tree.

3. The path costs are the same on all the switches (20000) as all the ports are 1Gbps.

4.1.1 STP

The traditional STP protocol creates only one common spanning tree based on these common steps:

a. Every switch sends out BPDUs on both interfaces as multicast assuming at first that each of them is the root. *Switch 1 is the root in our scenario with the lowest switch ID.*

b. Every receiving switch will decide upon the root switch and decide the port roles/states based on:

   1. Path cost to the root node
   2. Lowest sender BID
   3. Lowest sender port ID

c. The Decision Making process on the nodes:

1. Sw1 cost to itself is 0. Both ports will be designated and active. GST and SM traffic is not blocked on the root switch.

2. Sw2 receives BPDUs on both ports with the same path cost and the same BID. Since these factors are still the same in a generalized scenario of connecting two nodes with two redundant links, the process of deciding the port roles degrades to the randomness of port ID. In our scenario *p2-1* (GST) will be assigned as a root port while *p2-2* (SM) will be blocked. One problem that arises with STP is that switches will not send out BPDUs on ports that they received them from. Sw2 is responsible for switching the SM packets from Sw1 to Sw3. However it will not send out the BPDU on *p2-2* since it received it on that port form the root and afterwards the port is in the blocked state.

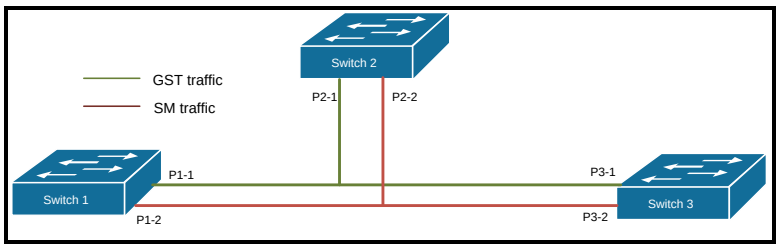

Figure 4.3 How the Ethernet switches sense the physical connectivity because of the OpMiGua transparency.

3. Sw2 should receive BPDUs on both ports from both switches. This creates the impression that both switches are connected on both ports in a shared medium, as shown in figure 4.3. For the GST traffic, this assumption is true as the ports on the switches will share the same fibre but will be assigned different wavelengths. However serious problems arise for the SM traffic:

   i. As BPDUs received from one switch are not sent out on that interface, Sw2 will not forward BPDUs from Sw1 to Sw3 and vice-versa.

   ii. Sw2 is responsible for switching all the traffic between sw1 and sw3. However, its SM port is seen as virtually connected to a shared medium receiving traffic from both Sw1 and Sw3. This is why it will not forward the BPDUs, the data traffic and prevent the convergence of the spanning tree.

4. Sw3 will receive BPDUs from the root on the GST port and from Sw2 on the SM port. The GST port will be assigned the root port role while the SM traffic port will be blocked due to the higher path cost value.

The problem with STP is thus clearly defined. The SM ports will not be able to take part in the active topology. Furthermore the spanned tree will not converge correctly because all the intermediate nodes in a hybrid network are responsible for packet switching to the next node in the chain. The BPDUs in the SM ports will not be forwarded down the network which fails the correct computation of the active topology.

4.1.2 RSTP

The protocol, as described in section 3.4, is based on the same algorithm as STP. Thus, the network convergence time when computing the active topology is the same as in STP. However, the convergence in case of a failure differs because of RSTP's usage of the handshake/proposal algorithm:

1. The convergence time in STP is approximately 30-60 seconds while RSTP uses a node by node synchronization protocol that allows the convergence of the whole network in 2-3 seconds. This convergence time is very important when employing Ethernet with OpMiGua as the data loss in case of failure is extremely high for high-throughput networks. Furthermore, if it is the GST port the one that fails, the guaranteed-service property is no longer offered.

2. The problem with the virtual shared-medium that we discussed for STP is very important for RSTP because the proposal can only be used on point-to-point links. In case of the shared medium, RSTP employs the STP legacy version thus not offering its enhanced properties. In the *single dedicated-port* OpMiGua network topology we showed that both SM and GST are perceived as shared-connections for all the switches. This implies that RSTP will always degrade to the STP convergence times during a failure. The blocked ports in the intermediate nodes (sw2 in figure 4.2) will have the *alternate port* role since both communicate directly with the root. However for sw3 the SM port problem still persists.

RSTP implementations support the creation of different spanning trees for each VLAN in the network [30]. The extension of BPDU version 2 includes the VLAN ID as part of the BID

[4 bits priority|12 bits VID|48 bit MAC]. It allows the creation of different active topologies for different VLANs. The solution to the problem of the shared medium discussed above, could be to separate the SM and GST traffic in different VLANs. However, in order for the GST/SM traffic to be able to use the redundant links, the ports need to support a trunking protocol. To our knowledge there is no standardized trunking protocol so far, but specific proprietary VLAN management protocols [29], [30]. VLANs are assigned either protocol-based or port-based. However, this assignment will affect the incoming traffic that will be aggregated from the input ports of the switch. The links connecting Ethernet nodes with the OpMiGua infrastructure should allow both traffic flows but we can have different STP instances for each one of them.

The main disadvantage with this network solution is the underutilization of the redundancy capability from the Ethernet switches. It is not possible for the SM port to be redundant to the GST port as it relies on the physical GST path for transmission. In addition, the separation into port-based VLANs excludes the possibility to use the GST ports for redundancy of the SM traffic. Thus the protection and restoration has to rely on the mechanisms offered in the optical domain. The Ethernet might support them by using multiple physical ports as trunks where the failure of a single link in a bundle will not interrupt the connectivity.

4.1.3 MSTP

The benefit of using MSTP is mapping different VLANs to a single STP instance allowing load-balance and easier management in case of an increased number of VLANs. In order to implement MST in a network, we need to determine:

1. The number of STP instances needed to support the desired topologies;
2. Mapping a set of VLANs to each instance. The way MSTP is employed derives from the way the VLANs are assigned:
    - **QoS traffic related**. In the OpMiGua network it is logical to differentiate two instances for the GST and SM classes. This classification allows the assignment of multiple VLANs to each class differentiating between the traffic of different customers.

- **Network topology and connectivity.** The approach is closely related with the physical topology and the instances would contain both GST and SM traffic. This case is more compatible with the essential concept of MSTP and allows a flexible management of the network from an administrative point of view. However, this implies that a generalized scenario needs to take into account the topology of the network. The network designer and administrator have to manually configure the instances through the network, which increases the possibility of human error.

From the studies in chapter 3 of the VLAN tagging, we recommend the use of a PBB tag. This would allow the service provider to use both of the mentioned VLAN assignments. The I-SID (service ID) could be used to differ between the GST/SM traffic classes, while the S-VID (Service VID) would allow the employment of different spanning trees exploiting the MSTP abilities.

4.2 Dedicated per-switch ports for GST and SM

In figure 4.4 is shown a scenario where the switches are connected in a full mesh topology. Furthermore, the most important aspect in this case is that each of the switches has a dedicated port to each-other. This topology is one possible solution to the virtual shared-link connectivity problem discussed in the previous section. The Ethernet switches will have point-to-point connections which enable the employment of RSTP and avoid the intermediate node processing problem. Sw2 will be able to switch the traffic from sw1 to sw3 and forward the BPDUs as well.

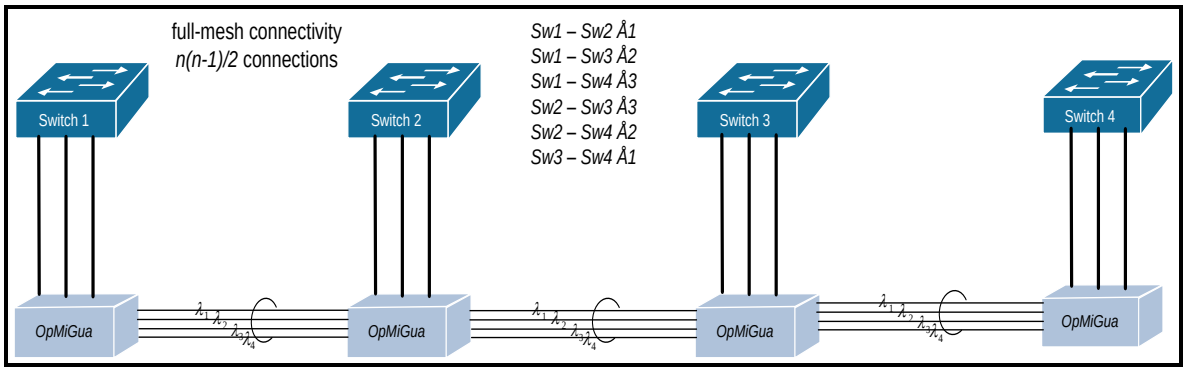

Figure 4.4 An example of full mesh connectivity for the GST traffic with static wavelengths.

The spanned-tree of the network topology after the convergence is shown in figure 4.5. The tree is the same for STP and RSTP. Since the network is a full-mesh, the number of links is

equal to *n(n-1)/2* where *n* is the number of nodes in the network. The converged topology will always be a star topology with the root switch in the centre.

This architecture creates a full-mesh topology for the SM paths as well. However, the packets will still be processed by all the intermediate nodes in order to be transmitted throughout the network. Thus, full-mesh connectivity at the Ethernet layer does not translate into full-mesh connectivity at the OpMiGua layer. Such abstraction incompatibility results in the same problem of processing BPDUs as shown above. Furthermore, multiple SM ports are an illogical waste of resources in this case. However, we could use multiple ports as an Ethernet Channel bundle for higher throughput and for redundancy.

This architecture allows the full convergence of the spanning tree protocols for the GST paths, but there are a number of issues related with it:

1. The number of ports in case of taking the approach one dedicated port per traffic class (one per GST and one per SM for each switch) would mean the double number of ports of a full mesh network. The number of ports required on the Ethernet switches is increasingly proportional with the network diameter [2 x n], while the total number of links is [2 x n(n-1)/2].

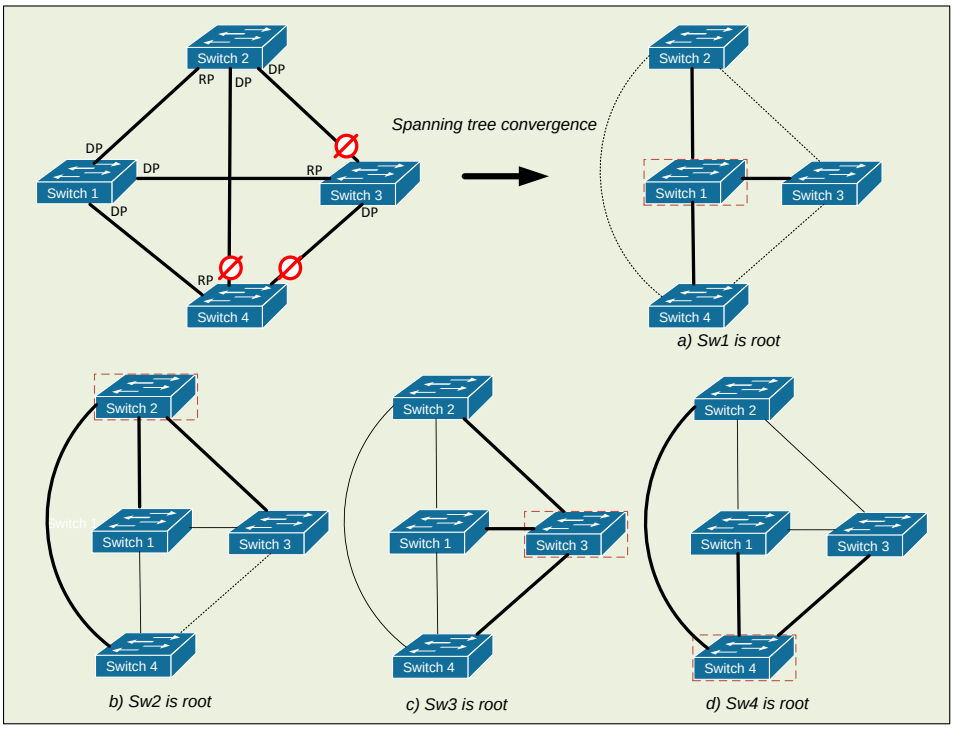

Figure 4.5 The spanned network topology after STP/RSTP convergence (GST or SM connectivity).

2. The SM traffic in an OpMiGua network follows a chain procession by the intermediate nodes. This scheme allows the full convergence at the initial state but makes no use of the resources afterwards and all the traffic will be processed by the root. The connectivity once again creates virtual network connectivity for the intermediate non-designated switches. For example in fig. 4.5 a) the traffic from sw4 to sw2 will need to be sent out to the root sw1 first. However, this traffic will need to be processed by sw2 on its way to the root (see figure 4.4).

3. The GST traffic, in the case of statically assigned wavelengths for each port, will make use only of one wavelength on each node, the one that connects to the root switch. Referring to figure 4.4, if sw3 will communicate with sw4, the physical port is assigned the wavelength λ1. However, this port is in a disabled state and the traffic will be sent out the port connecting to the root, λ2. The root not only will need to convert the traffic to λ3 directing it to sw4, but it needs to know that the traffic should be sent to node 4. There should be an optical packet labelling to allow this transition. Furthermore, we lose the main benefit of the OpMiGua network of bypassing the intermediate nodes for the GST traffic and degrading the network/root performance.

The problems discussed in this chapter are addressed by the network architecture in the next chapter.

Chapter 5

Proposed architecture

5.1 SM ports chain connectivity

The problems considered with the above architectures lead us to a network architecture solution based on these derived characteristics:

1. The SM traffic needs to be processed by all the nodes on the way. Additionally, for the spanning tree protocols to work, it needs only one distinctly assigned physical port for one incoming and one outgoing interface. These are interfaces connecting each intermediate node with its two neighbours. The links give the notion of point-to-point links to the Ethernet switches. Moreover, we refer to such topology as chain connectivity.

2. The GST traffic relies on the WRON architecture and requires minimally only one physical port on each Ethernet switch. The number of Ethernet interfaces is relatively with low cost. This allows the use of full-mesh GST connectivity between the Ethernet nodes. Each interface of the switch will be connected in a virtual point-to-point link with all the other nodes in the network, as shown in figure 5.1. In a SWRON underlying architecture, each GST port would be assigned to a specific wavelength simplifying the OpMiGua node. Furthermore, as it uses the orthogonal state of polarization to differentiate between the GST/SM flows, the use of different physical ports allows a simple assignment of the optical label.

3. The OpMiGua node is being extended to allow electronic labelling. In order to support this labelling method, an optical-electronic conversion of the packets will be needed for the header of each GST/SM traffic flow. We will discuss these options in the next section when assigning VLANs and QoS class of service to the Ethernet frames.

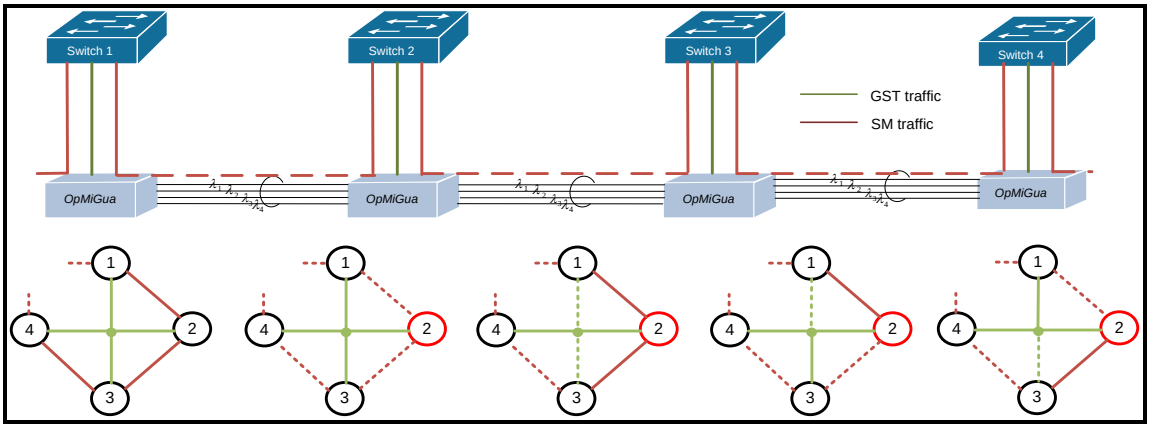

Figure 5.1 Ethernet/OpMiGua network architecture for chain SM ports connectivity. Active topologies are shown when sw2 is the root.

5.1.1 STP and RSTP

The convergence of the network at start-up allows all BPDUs to be processed by all the nodes on all the ports. However, the redundant GST/SM paths problem (see figure 5.1b) still remains unsolved till we apply a VLAN division scheme. It is important for this purpose to find a pattern for the selection of the active ports on switches distant from the root switch:

1. The root switch will always have all ports active in a designated role.

2. The path cost of the GST links will always be equal the cost of one link because the intermediate nodes are bypassed. The path cost of the SM links is incremental when going down on each leaf level of the tree. This implies that the GST ports can only be blocked only on the nodes connecting directly to the root switch. All other switches will have a higher path cost to the root for SM compared with GST. The SM ports will always be blocked on the nodes two hops from the root. This rule applies for all the possible network topologies, independently of the network diameter, as shown in figure 5.2.

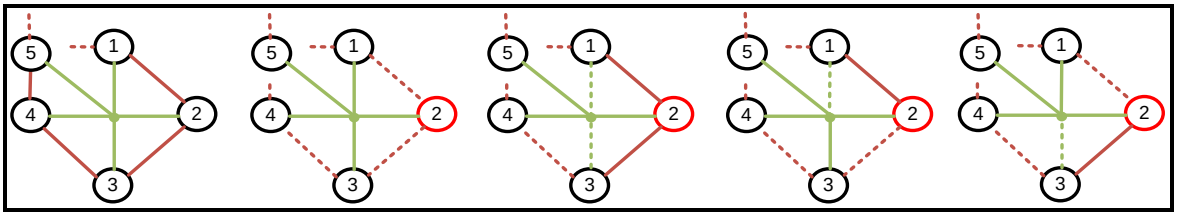

Figure 5.2 the possible spanning trees in a 5-node network.

5.1.2 Assignment of VLANs and MST instances

The assignment of VLANs is closely related with the Spanning Tree instances that will allow the interoperation of Ethernet switches with the OpMiGua infrastructure without blocking the traffic. The classification is based on:

1. *The traffic classes*. In this case, we would assign one VLAN to the GST traffic and one VLAN to the SM traffic. The 802.1Q standard includes 12 bit in the VLAN tag field for the VLAN ID. This value is included in the extended BPDU frame as part of the bridge ID. However, the VLAN assignment can be transparent when using IEEE 802.1Q-in-Q VLAN Tag Termination. Furthermore we propose the use of IEEE 802.1ad Provider Bridges MAC-in-MAC which allows a hierarchical approach of managing the VLANs. The VLAN assignment will make use of two different service provider assigned VID for GST and SM traffic, as shown in figure 5.3

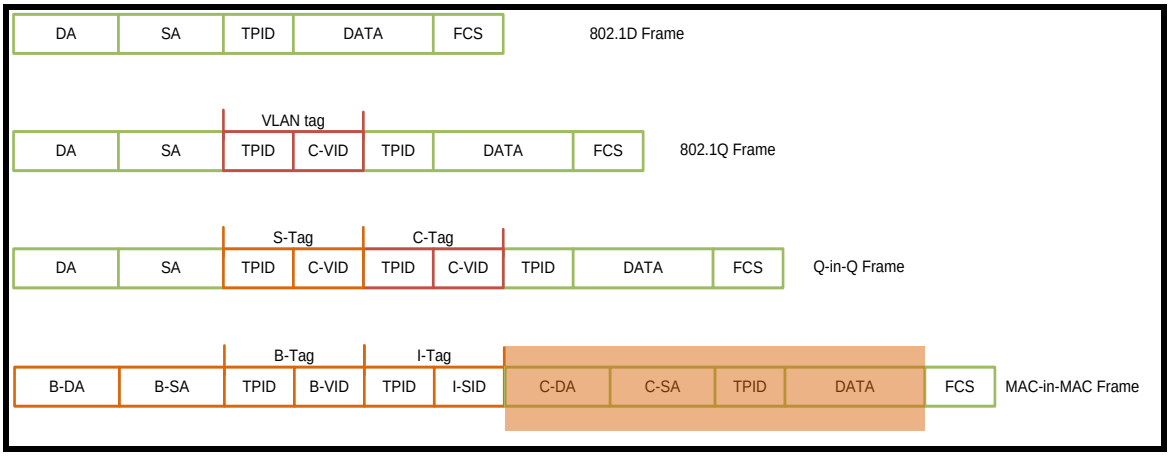

Figure 5.3 VLAN tagging formats

2. *The available ST in the network*. In figures 5.1 and 5.2 we illustrated the available ST instances in a network with equal link costs. However, the number of useful trees is still limited to one because we distinctly divide the GST and SM ports.

The GST/SM traffic must always pass through the GST/SM ports to the OpMiGua node to be optically tagged. One of the problems this architecture does not address is how could the traffic be transmitted through different ports if one of them fails. For example, using SM ports to send out the GST traffic and still be able to tag/route it correctly in OpMiGua. This is an issue closely related to the underlying OpMiGua node design and the protection mechanisms employed in the circuit-switched and packet-switched domains.

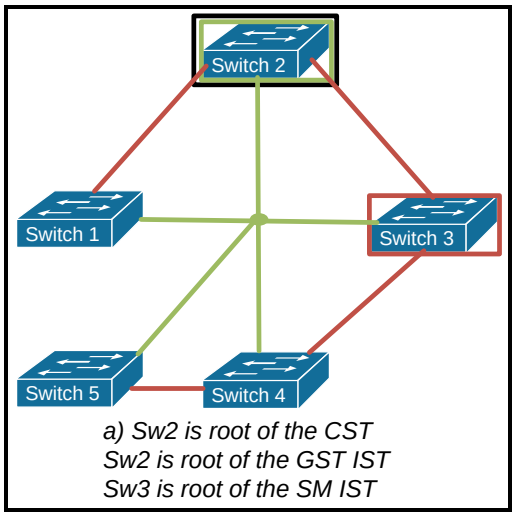

Figure 5.4 Distinct STIs in a single MSTP region.

The ST instances in MSTP are independent of each other and there is a one-to-one relation with the VLANs. This implies that if we proceed as in step1, there must be only two ST instances, see figure 5.4. MSTP allows up to 16 instances in one region, which is why within OpMiGua one region is enough. The Common ST of the region can use any ST while the GST and SM ones must be correctly configured to match the physical paths.

The physical ports, when assigned on per port-based VLANs, will not be able to allow redundancy for each other. Additionally, load balancing is achieved only through the separation of the classes into different ports. However, the load-balancing implementation requires that the traffic shares both ports; while forcing the creation of trunking ports and Ethernet channels. The redundant paths problem in this case disappears because the ports will be encapsulated into one virtual port. Nevertheless, there should be a mechanism of differentiating between the GST and SM traffic on the OpMiGua node.

5.2 Assigning VLANs and MAC QoS

In order to be able to distinguish between the traffic classes of service (GST/SM), it is needed to have support of the VLAN tagging. This statement was verified when analysing the problems arising with the xSTP protocols and OpMiGua in chapter 4.

In case the optical label is being used, the use of VLAN tagging as shown in figure 5.3 is relevant only for the switches. It allows the nodes to create and manage different STI for the different classes of service.

In the case an electronic label is being used, the OpMiGua nodes are able to identify and process the traffic from an electronic packet header. We propose to use:

1- The TPID in the Service Provider outer tag with two custom values for respectively GST and SM. These values should be different from the well-known Ethernet TPID values but would be configured on the entire Service Providers network. When a frame is processed first, the outer tag would specify the type of service that it is assigned to (see figure 5.3).

2- MAC QoS, called in terminology Class-of-Service, can be used to differentiate between the GST and SM traffic. As previously discussed in chapter 2, OpMiGua presents three traffic classes named GST bearer service, high class transport (HCT) bearer service and normal class transport (NCT) bearer service. The HCT and NCT classes are sub-classes of the SM class. The differentiation between the HCT and NCT classes is performed in the electronic buffer. The HCT class is given absolute priority when a wavelength to the destination becomes vacant, that is why the HCT class experiences lower delay than the NCT class. Recalling the traffic type- queue numbers mapping in 802.1Q, (table 3.2 and table 3.1)

Number of queues	**Defining traffic type**							
1	BE							
2	VO				BE			
3	NC		VO		BE			
4	**NC**		**VO**		**CA**		**BE**	
5	NC	IC	VO		CA		BE	
6	NC	IC	VO		CA		BE	BK
7	NC	IC	VO		CA	EE	BE	BK
8	NC	IC	VO	VI	CA	EE	BE	BK

The default mapping with three queues is incoherent with the OpMiGua QoS. The approach is to assign the GST for the high priority traffic and the two other classes available assigned to less sensitive traffic type. It is important to notice that both Network Control and Video/Voice traffic type should be delivered through the GST paths. Furthermore, the use of a four queue recommended mapping is possible with OpMiGua. The Normal Class Transport NCT can be mapped and used for Best Effort and Background traffic, while the other traffic type (Critical Applications, Excellent Effort) under HCT.

5.3 Verification

The verification of the analysed and proposed architectures for the operation of Ethernet switches in an underlying OpMiGua network was carried on the dynamips emulator [47]. It is an open source Cisco IOS emulator that works together with a GUI and a text-based front-end for dynamips. For the simulations has been used a 16 ports switch module NS-16ESW in the Cisco router 3640 with IOS version c3640-jk9s-mz.124-16a that supports xSTP protocols and Cisco proprietary Per-Vlan STP.

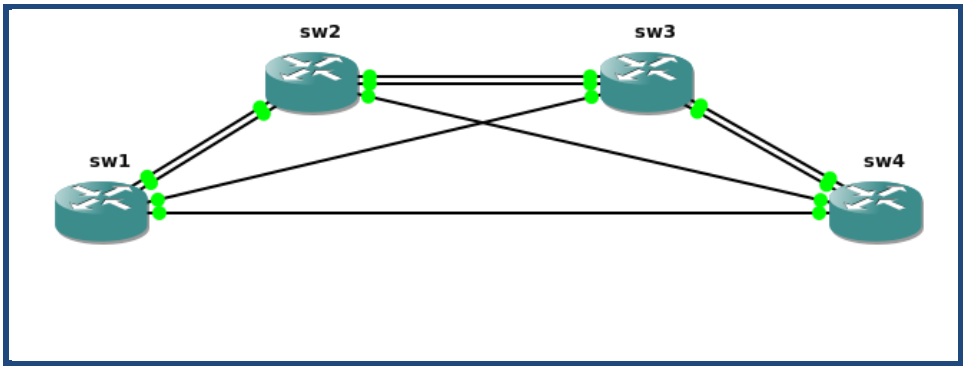

Figure 5.5 The network topology used when simulating SM chain-port connectivity.

A network of four nodes has been used throughout the simulations. All the nodes run the same IOS version and have the same physical specifications of the switching module. The network is connected as shown in the screenshot in figure 5.5. We have made the following analogies with an underlying OpMiGua network:

1- The GST paths are logically translated from wavelengths carried within the OCS to simple point-to-point FE links. The port speeds are not important in our simulation as we are not evaluating the recovery time after failure. Such measurements are out of the scope of this

thesis because there is already a lot of research work, as we discussed in chapter 3, focused on performance issues of the xSTP protocols. Thus the recovery times are well-known for xSTP.

2- The GST paths are connected in a full-mesh network as previously discussed.

3- The SM traffic will be processed by all the nodes that is why it is represented in a chained connectivity. The point-to-point single connections create a ring network.

4- The root switch is selected randomly and then changed throughout the simulations. The cost of the links is the same as not to affect the generalized scenario.

In Appendix A we have enclosed the console output of the switches in both cases explained further down.

Case 1 STP without virtual separation of GST/SM

The spanning tree is built based on PVST+ which is the Cisco implementation of RSTP supporting VLANs. The loop-free topology blocks all other ports except the designated ones. In figure 5.6 are shown the loop-free topologies acquired when changing the root switch manually through the priority value. It shows that if no VLAN division scheme is used, the GST and SM paths are intertwined by the switch and it is impossible to switch separately the traffic. The aggregated traffic from the switch will change ports and it will not obey the OpMiGua path rules. The results fit also the analysis made in chapter 4.

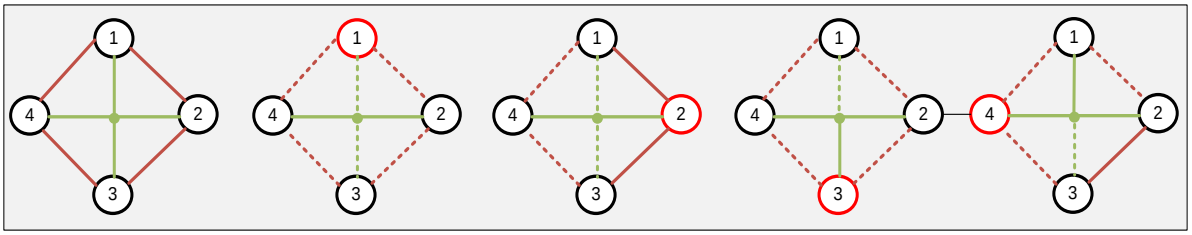

Figure 5.6 All spanned trees without virtual separation of GST/SM.

Case 2 STP with VLAN separation of GST/SM

Different VLAN values are used in order to logically separate the traffic. This logical division is additionally configured in the physical ports of the switches with dedicated ports for GST and SM. In figure 5.6 a) the GST links are marked with green while SM links with red lines. Two different spanning trees are built, one for each VLAN plus the default VLAN 0 which carries the management frames. The root and designated ports are dependable on the set priorities. We have changed these priorities and disabled different of the GST ports on

different switches. The results show that apart from the dependence of the root position, the redundancy within the GST paths is possible.

The topologies created by the per-port VLAN separation create completely different networks. Thus, there are no redundant links for the SM paths. This is one of the main limitations of this design. We cannot overcome it by a physical solution as we already have shown that only two different and specific in/out SM ports should be assigned to one switching node.

A possible solution is to use the Ethernet switch management protocols to change the tag of the HCT traffic and switching it within an available GST circuit. This could be done by VLAN management protocols. It is outside the possibility of this thesis to test such systems as they are proprietary solutions. The only open source solution is the open VLAN Management Policy Server. It requires a Server and EAP-based protocols to communicate with the switch.

The simulations are a proof of concept for our analysis and the proposed network design solution when using the OpMiGua infrastructure as a transparent Service Provider network.

Chapter 6

The integrated Ethernet/OpMiGua node

The proposed network solutions discussed so far have focused on the problems faced when the optical domain has been divided from the all packet switched Ethernet infrastructure. The layering technologies have been distinct from each other and the OpMiGua infrastructure has been considered as a provider backbone. After analysing the problems introduced by xSTP protocols and the proposed solutions, in this part we design the hybrid node which relies on Ethernet switches for its packet switching functionality. The motivation for this part of our work is that currently there are no commercially available optical packet switches other than in research [11], [16]. Moreover, the main problem with OPS is that the buffering system relies on either Fiber Delay Lines or an electronic buffer. As Ethernet switches are a relatively cheap technology offering the desired buffering capacity, we look into the possibility of designing a simple OpMiGua node based on Ethernet for its OPS part and evaluate how this adds to the limitations introduced by xSTP and vice versa. The previous OpMiGua demonstration test-bed [8] relies also on an Ethernet switch; however the limitations induced by the xSTP protocols are not foreseen on the nodes.

Thus, the scope of this design is to offer a node that integrates the Ethernet switch considering:

1- the physical connectivity solutions for a functional xSTP topology as in chapter 5;
2- offering the simplest design with as few optical elements as possible;
3- give an empirical analysis of the limitations of the node.

6.1 Optical packet header using PBS

Let us consider in a structured manner the functionalities expected when integrating an Ethernet switch instead of the optical packet switch inside the OpMiGua node, see figure 6.1:

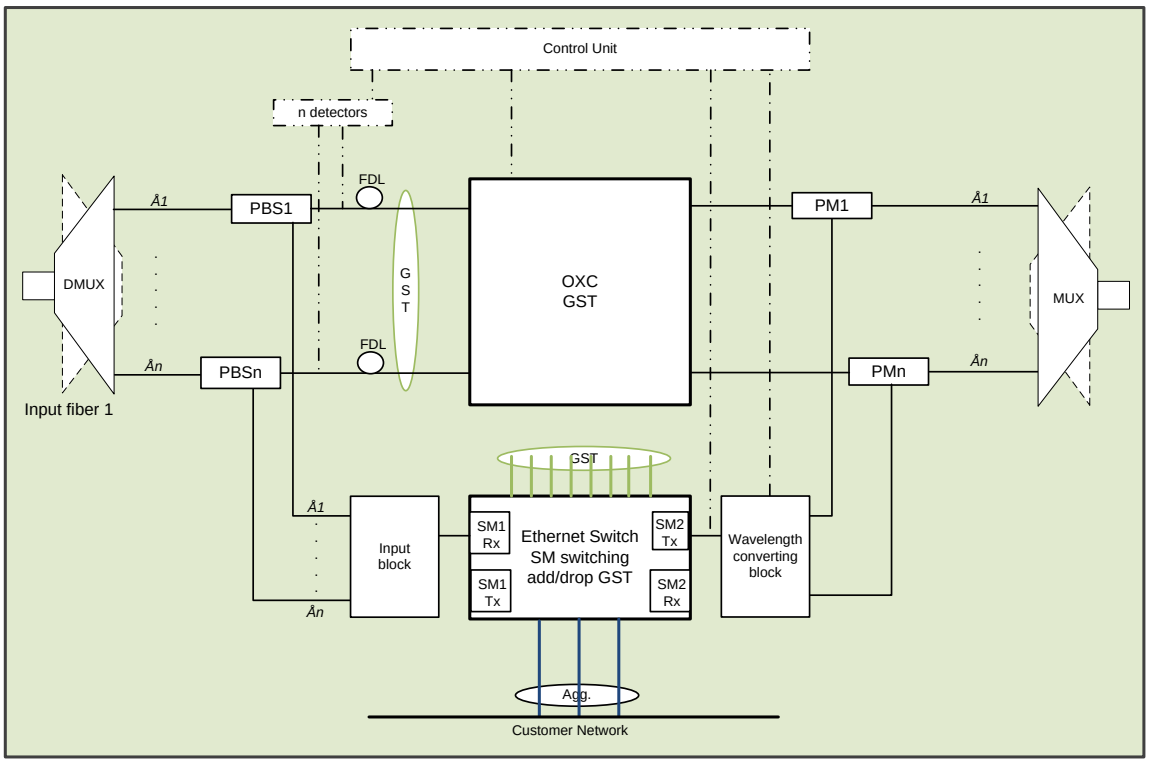

Figure 6.1 Functional integrated node design. The control signals are represented in dotted lines. The TWC is used to convert the SM port signal into the available wavelength. The PBS detects the traffic type and directs it to the appropriate switching module. The aggregated traffic is inserted as GST/SM based on the VID or QoS. GST traffic is inputted at the OXC through a coupler since the OXC is responsible for the circuit-switching.

1. In the node, the forwarding and switching functionalities are divided. The GST packets should be forwarded all optically while the SM packets will be switched electronically by the Ethernet switches. The Polarization Beam Splitter used in [3] allows the all optical differentiation of the traffic. In our scenarios, as discussed previously, we rely on the VLAN ID only to allow the switch to make the switching mode decision for the traffic added at that node. However, if the VID and QoS will be used to differentiate the traffic, the header needs to be processed for both types of traffic. This is done when the Ethernet switch is aggregating traffic from the customer's network.

2. The integrated node should support a high granularity. In the static WRON that we are looking into, the granularity is wavelengths. In OpMiGua the same wavelengths are used for transporting both packet-switched and circuit-switched traffic.

a. SM traffic. The Ethernet switches need to have at least two different physical ports connecting to their neighbour switches in order to build the spanning tree (refer to chapter 5).
b. Aggregation. The Ethernet switches are aggregating or transmitting traffic out of the network. The most important functionality is that the node should be able to aggregate and insert GST/SM traffic in the network. However, these interfaces connecting to layer 3 devices or other customer infrastructure are of no interest for the node architecture.
c. GST traffic. The Ethernet switches are connected in a full-mesh topology. Since the underlying connections are wavelengths in a SWRON, we assign a different wavelength to the distinct GST ports of the switch. These ports are considered as input interfaces to the Optical Cross-Connect (OXC). They should connect after the PBS/header processing block.

3. The SWRON, in a network with *n* nodes (Ethernet switches), needs *n(n-1)* connections. The connectivity when using spatial wavelength reuse can be achieved with n(n-1)/2 wavelengths per fiber [52]. The underlying GST lightpaths can be realised as an Optical Circuit Switched (OCS) network with transfer guarantees as in [3]. On the Ethernet switch are needed only *(n-1)* ports for a full GST mesh connectivity as proposed previously. For example, for a network with 8 nodes are needed 7 ports on each switch and 56 wavelengths. In our node for simplicity reasons we are considering only one bidirectional fiber between each pair of connected nodes. This represents the worst case scenario for the wavelength requirement in a SWRON without wavelength conversion. In the case of more than one fiber per connection, the number of wavelengths is reduced by the availability of a larger number of alternative physical links. The receiver part may accept any wavelength which the OXC drops depending on its configuration.

The transmitter part will need to be tuned at the specific wavelength based on the OCS configuration. A Tunable Wavelength Converter (TWC) controlled by the control block might be connected to the Tx of the Ethernet port before connecting to the GST circuit through an optical coupler, figure 6.2. The OXC is built by Static Optical Add/Drop Multiplexers. In order for the switch to be independent of the knowledge of the wavelengths

used, our proposal is to use TWC and Arrayed Waveguide Grating (AWG) to assign the specific wavelengths by the control block.

Most of the switching technologies nowadays employ electronic switches that introduce the opto-electronic conversion of the signal. It is desirable for the GST part to match the high-bandwidth of the optical fiber, thus avoiding the lower switching speed of electronics. The AWG is a fixed wavelength router that allows the signals to simultaneously be fed into different inputs without interfering with each-other. Moreover, it is a low-cost integrated device and gives more simultaneous routing options as compared to a passive star coupler [49]. The main disadvantage of AWG is their lack of flexibility because the switching is fixed based on the wavelength of the signal at the output. Our choice overcomes the flexibility problem by combining the TWC before the AWG which will be controlled by the control block and allow a flexible switching at the output block.

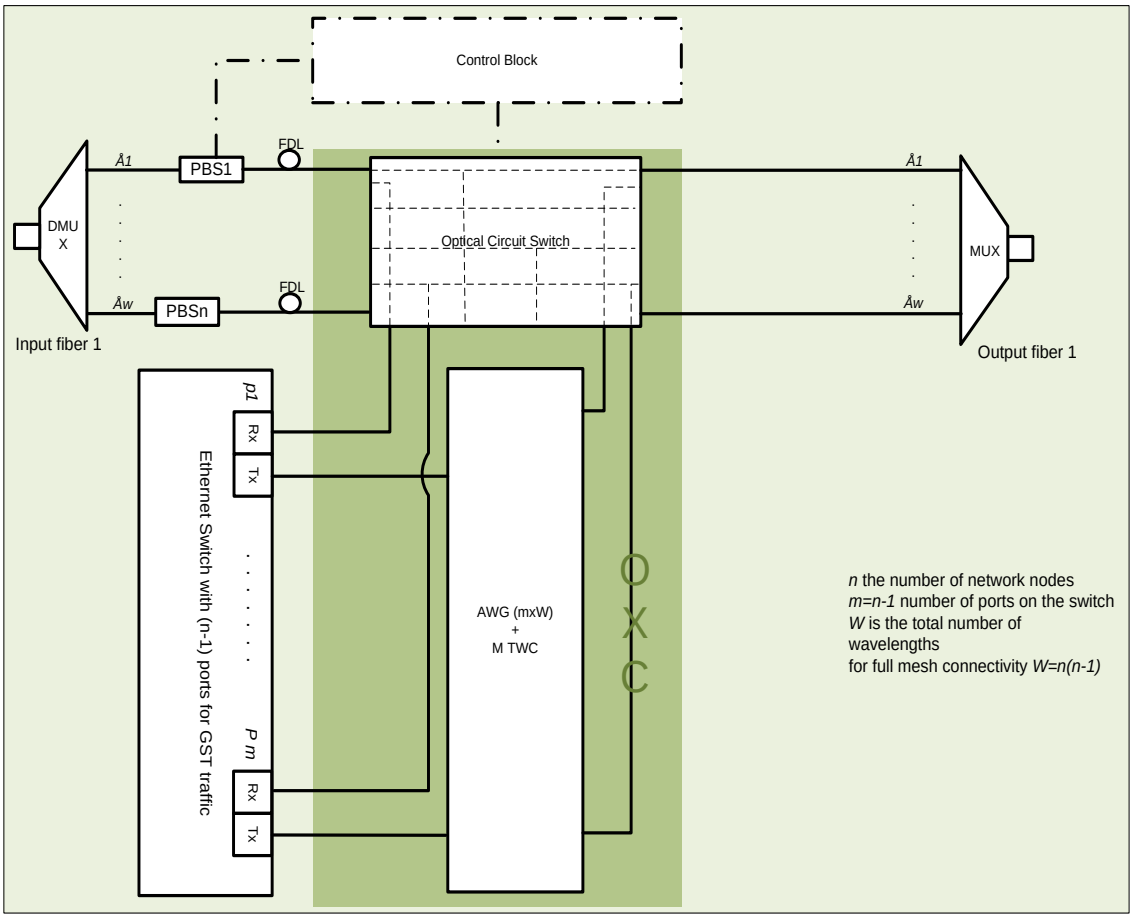

Figure 6.2 An example of the Optical Cross-Connect or the GST traffic in a configurable S-WRON.

4. Considering the OPS part of the OpMiGua node and making the analogy with the Ethernet switch :

    a. We are using only two ports for the packet switched traffic. The traffic entering the OPS in [3] is divided into different wavelengths/ports while we are using only two ports in the Ethernet switch.

    b. Actually from the input part of the node we only connect with one port. The receive part Rx of the Ethernet port is connected to this one-way transmission scenario through an input block. This block should accept any of the *w* wavelengths from the PBS-es at the input and deliver only one specific wavelength at the input of the Ethernet port. The wavelength choice depends on the physical Ethernet interface.

    c. It is important that we take into consideration that at least two different SM packets carried on different wavelengths might arrive at the same time at the input block. The block should be able to buffer the optical packets. Our proposed solution is to use the multiple-input single-output FIFO optical buffer presented in [5] with fixed length optical delay lines and with the characteristics of a RAM, figure 6.3. In our case, an *nxn* space switch could be used. . There are other possible solutions of using optical buffers [49], [50], but the main advantage we achieve from this selection is that the delay time of the signal and the amount of needed fiber for buffering is at least ½ less [5]. The need of a wavelength converter at its output would be avoided by the use of wavelength independent photodiode receiver at the GbE port. This choice minimizes the cost of an added element and the delay of converting every signal before receiving it in the switch. However, another possibility is to use an optoelectronic wavelength conversion. In this case though we would lose the benefits achieved by using the polarization as a packet label in an all optical manner.

    d. There should be a wavelength converting module connecting to the transmitting side Tx of the port. This block is controlled by the control electronics to convert the signal from the switch to an available wavelength. The detection of the available wavelengths can be done through detectors in front of the FDL leading to the input of the OXC as in [3].

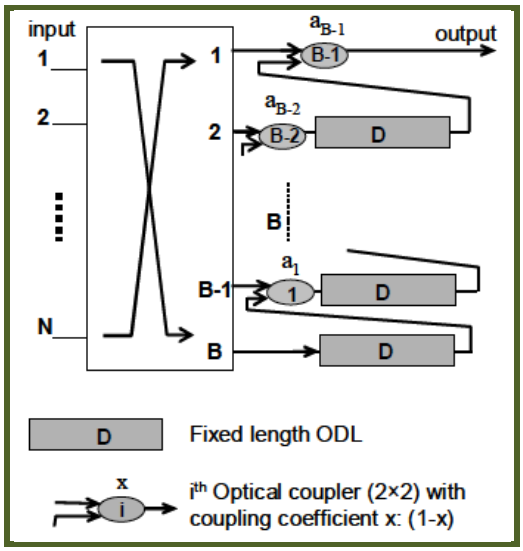

Figure 6.3 A MISO-FIFO buffer structure with fixed length ODLs, taken from [50].

This output block is built by connecting a Tuneable Wavelength Converter (TWC) with an 1xN Arrayed Waveguide Grating (AWG) switch with *n* outputs as the number of wavelengths we are using in the system (see figure 6.4). The TWC can be built with semiconductor optical amplifier March-Zehnder interferometer with a tuneable laser controlled by the control block. The output block can be also built by using fast switching modules. However our choice is attractive because it uses only all-optical passive optical components. The main benefits are simplicity in packaging of such elements and lower power consumption.

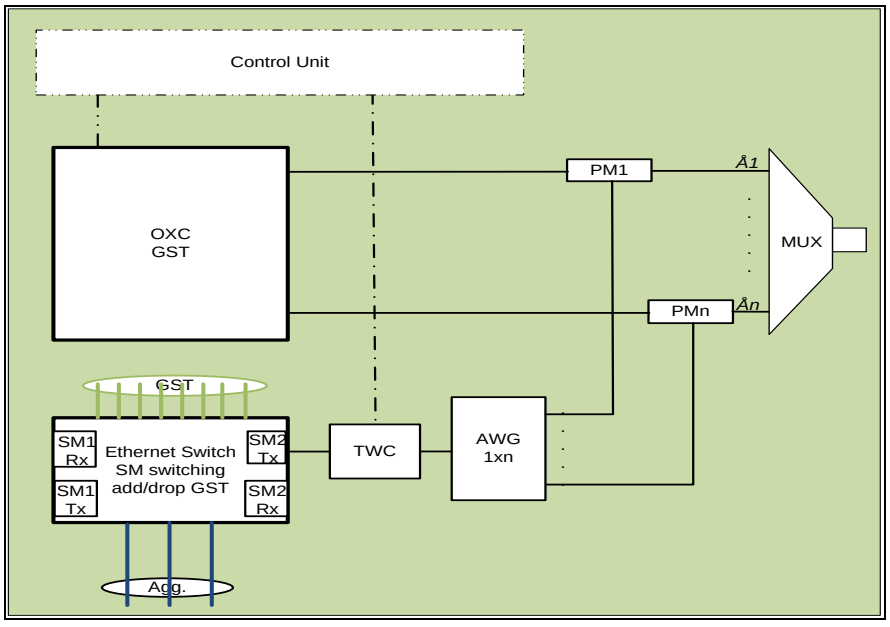

Figure 6.4 The controlled wavelength converting module using a TWC and a 1xN AWG.

5. The control block is responsible for :
    a. Detecting the available wavelengths for the SM traffic.
    b. Controlling the tuneable wavelength converters to insert the SM traffic at the appropriate wavelength and polarization.
    c. Signals the switch to send/buffer traffic at the output interface Tx.

6.2 The node using an electronic packet header

To our knowledge, all optical header processing is not available commercially but is being looked into in the research area. That is why it is easier to deploy a system where the header of the packets is processed electronically, while the GST payload can be switched all optically. The building blocks of an integrated node would be as in figure 6.5. The polarization beam splitters and maintainer couplers are not needed anymore. At the output of the wavelength conversion block we use normal optical couplers. For a simpler and less costly solution, we use only one coupler after the MUX directly to the physical fiber. This is possible because the control unit recognizes the available wavelength after the signal is demultiplexed and controls the TWC. The wavelength conversion block is simpler compared with the first design because the AWG is not needed.

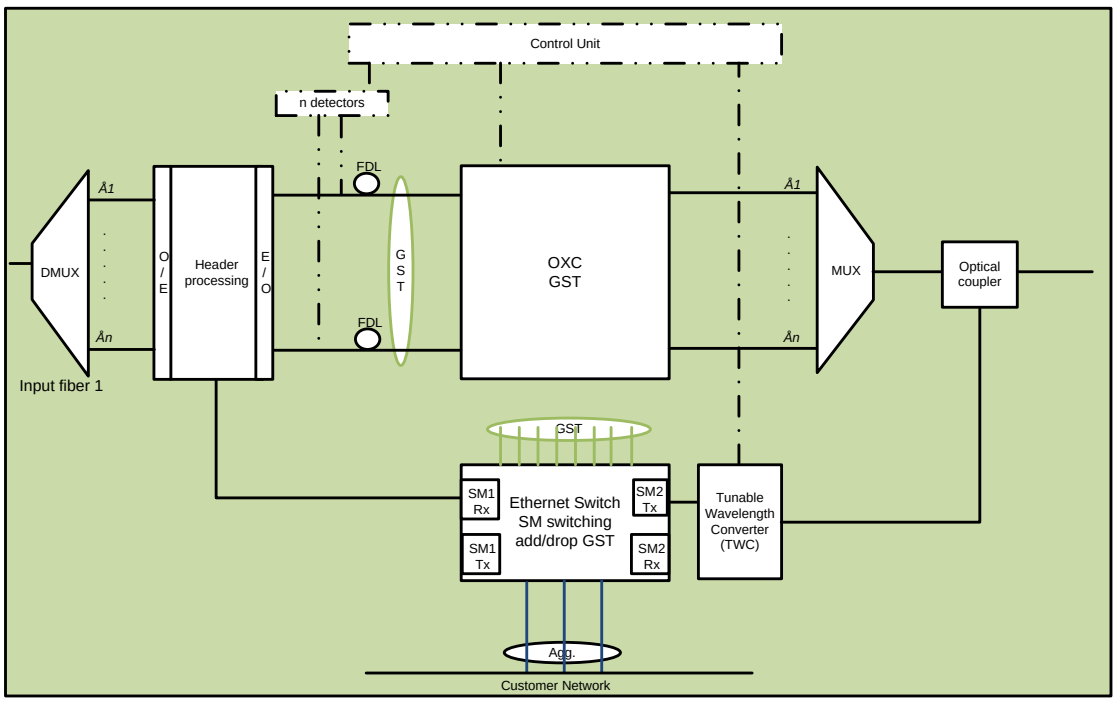

Figure 6.5 Node design with electronic header processing.

At the input is the header processing unit. The header has to be translated to an electronic signal and processed. If the signal is carrying GST traffic, no further processing is required and the header is translated back into optical signal. Furthermore it will be demultiplexed in order for the control unit to monitor the available wavelengths. The SM payload will continue to be translated into an electronic signal and passed to the switch.

The header processing unit is out of the scope of this thesis as there can be multiple ways in addition to the VID or CoS tags to differentiate the traffic. Furthermore GMPLS or OTN could be employed for this purpose. Nevertheless, an optical header could be used to tag the traffic and there has been research work toward all optical packet differentiation.

6.3 Node analysis and limitations

The static wavelength routed optical network is a good solution for traffic with a uniform pattern, while the hybrid architecture is attractive for unbalanced traffic [18]. In [10] are discussed several reservation techniques to solve the contention of the GST and SM traffic with proactive and reactive approach. However, the xSTP solution forces us to use a chain connectivity of the SM ports as discussed previously. This means that if node 1 has to aggregate and packet switch SM traffic to node 2 and node 3 down the line it still is connected to them through the same interface. The total packet/s delivered is dependable on the number and time range that the wavelengths are available. Considering a SWRON the path setup delay is avoided enabling a fixed delay of the GST traffic induced by the FDLs and the signal transmission through the fiber.

Regarding the packet reordering problem, we consider that:

- With the buffering scheme used in the input block the packet sequence is not modified within one hop since it uses a FIFO scheduling mechanism. Furthermore the packet priorities can be modified and changed in the Ethernet switch. In our node design there are both optical buffering for the Ethernet input port and electronic buffering inside the switch. The latter requires payload clock recovery which we discuss in the following section, while the former does not require it.

- The wavelength conversion in the output block maintains the packet sequence.

In a SWRON considering N the number of nodes and n-1 the number of ports for full mesh connectivity, the *network efficiency* is the ratio between the maximal number of lightpaths that can be established and the total number of lightpaths provided [53]. In our case, the number of wavelengths in the network is *n(n-1)* this ratio is 1. The maximum network transport capacity is *Tc=n(n-1)R* where *R* is the bit-rate of all the channels.

The lightpath allocation and minimal wavelength requirements in a WRON are extensively studied and heuristic algorithms have been used for this purpose [53]. GMPLS [55] offers a set of protocols for dynamic lightpath setup and can be used by the control/management block to allocate the lightpaths in a WRON. However, in this node design we are not focused in analysing the optimized number of wavelengths. It is important for the xSTP purpose that the nodes should sense the GST paths as point to point connections and the proposed scheme in figure 6.2 allows the switches to be independent of the wavelength assignments.

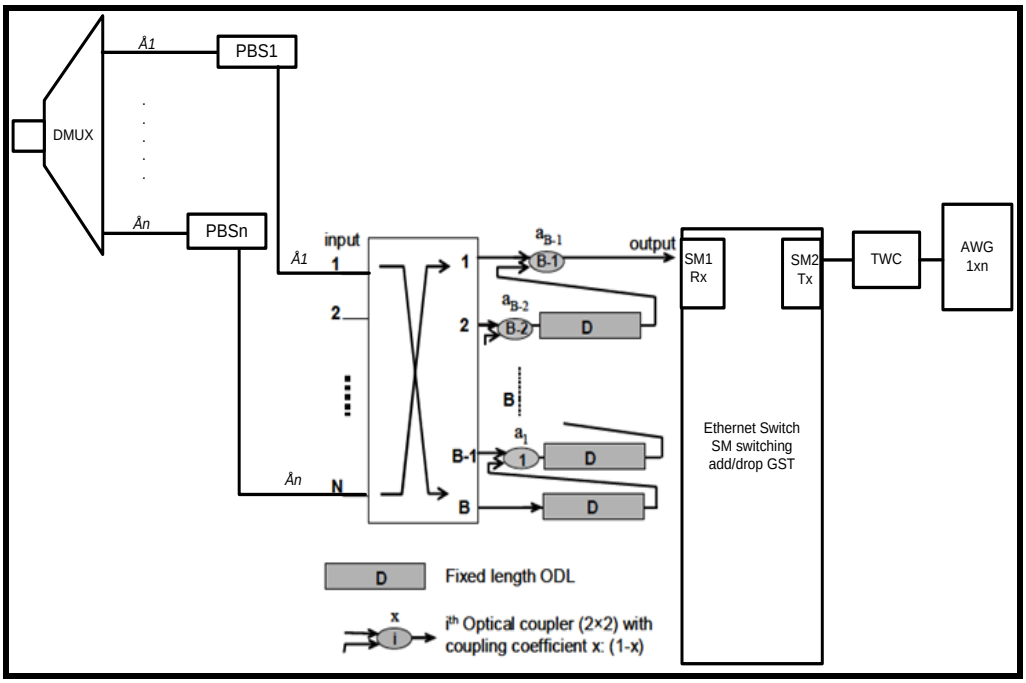

Figure 6.6 The input block design.

A major issue in optical packet-switched networks is contention resolution [52]. In electronic routers this problem is tackled using random access memory (RAM), which is not practically feasible today in the optical domain. Since all-optical buffers are technologically hard to realize, there seems to be a consensus that they should be avoided as much as possible

or at least limited to a minimum. This approach leads to designing optical networks where contention is either avoided or mainly resolved in the wavelength and/or space/fiber domain. The use of the Ethernet switch allows for an electronic buffering and QoS dependent scheduling of the packets. However, in the input block as shown in figure 6.6, the optical MISO-FIFO memory is needed to delay the packets coming simultaneously toward the SM1 port.

Case 1 (pktsize=256B). The input fiber is using a coarse WDM with W=16 wavelengths, each with a capacity of 1GB because of the end node interfaces. Let us suppose that at a given time *t* all the wavelengths are carrying SM traffic which will be detected by the PBSs and directed toward the Ethernet switch. The SM packets are the fixed size of 256B, the mean Internet packet size. The link capacity is determined by the capacity of the SM port in the Ethernet switch, for example 1Gb. The Optical Delay Line (ODL) has a delay D which should be equal to the time needed to process one MTU frame$D = 1.4\mu s$, while the delay of processing one 256B frame is approximately 0.256μs. The parallel structure automatically allows packets to be written into the buffer at a much higher speed than the link capacity.

The delay time is equal to the queue size normalized by the service rate plus service time. The optical buffer is equal to number of wavelengths multiplying the Ethernet MTU size, for example Sb=16 x 1598=25,6MB. The electronic buffer is specific to the switch and the minimal delay for an SM packet is the time to store and forward it when an available wavelength is immediately present.

The Packet Loss Ratio depends on the GST share as discussed in [3], however it is important the fact that a single SM port is being used. This limits the use of available wavelengths to only one at a specific time, thus not making an efficient use of the resources as in [3]. Furthermore, the number of nodes in the OpMiGua network might affect the packet loss rate because the number of necessary wavelengths in the SWRON for the GST connectivity is given from the number of nodes. An increase in the number of nodes increases the wavelength-conversion thus the PLR related to wavelength availability might be reduced. However, the average number of hops is increased for larger rings leading to increased delay of the SM packets which need to be processed at each intermediary node.

6.4 Problems to be addressed

In packet-based mobile backhaul networks, synchronization is needed to minimize interference between base stations and to guarantee optimal intercell handover performance. The nodes in our network share the optical environment. That is why we must take into consideration the problems introduced by the lack of an Ethernet's synchronization mechanism. In the case when PBS is used the packets are following different paths (GST/SM) all optically. The packets that are added will have a completely different clock from the packets that are already in the line. Even with full-duplex, the electronics in a 1 GbE interface still support the ability to perform this clock regeneration. The use of 1GbE interfaces allows the use of the Polarization Beam Splitter because the 1GbE physical interface supports the clock regeneration for every single packet in the line [6]. In [48] has been demonstrated a functional test bed employing 1GbE switches.

However, the 10/40/100 GbE technologies are implemented only for full-duplex connectivity and in a completely different physical coding than the old Ethernet PCS 8B10B used in 1GbE. The new 64B66B coding decreases the 10GbE overhead from 25% with the old one to 3% [6]. Furthermore, the node does not need to be able to regenerate the clock within one packet in the optical line. In our case this translates to the problem that the inserted packets might have a different clock which is not synchronized with the packets within the optical line. This introduces the difficulty of using 10GbE switches because of the need to synchronize at the bit level. In general terms, given our node design in figure 6.1 with PBS it has been demonstrated to be functional with 1 GbE interfaces. Research work has been carried on finding synchronization schemes for carrier and metro Ethernet. IEEE has created the task force group 802.3bf to standardize Ethernet synchronization mechanisms at the physical level. In [48] is employed an interesting flat synchronization scheme using the topology created by STP to logically spread the synchronization information. The synchronization problems and adapting the schemes for the OpMiGua/Ethernet node can be looked further into in the future work.

The physical aspects of the node are not addressed because the main scope of this work has been the focus on the xSTP protocols. That is why the node design itself needs refinement at the physical layer, which can be looked into in further work. Furthermore the control block is only described in an input/output requirement basis and considered as a black box.

Another important issue to be considered is how the switch is notified by the control electronics to send/stop the SM traffic and to keep buffering it. One possible solution is to send a busy signal imitating the CSMA/CD behaviour. For example, if the switch needs to forward at port SM2 Tx, the control block will keep the Rx side fed with a busy signal.

Chapter 7
Discussion

The main goal of this thesis was to propose solutions for employing Ethernet in an OpMiGua network after analysing how Ethernet and OpMigua may interact. There are several limitations because of the completely contradictory concepts of a loop-free Ethernet network with active parallel paths in OpMiGua. These limitations are shown in the analysis in chapter 4. Furthermore, we proposed schemes of combining Ethernet and OpMiGua:

1. in the network architecture considering OpMiGua as a transparent Service Provider network;
2. Inside a logical hybrid node architecture employing Ethernet instead of the Optical Packet Switch of the OpMiGua node in [3].

In order to verify the correctness of the analysis and the proposed network architectures, all the discussed topologies have been simulated through an Ethernet switched network with an emulator. The results show that the proposed schemes are successful, with regards to the functional interoperability. This is the first model of an OpMiGua-Ethernet network with respect to xSTP; other researchers have developed and analyzed OpMiGua node models [1] and evaluated its performance with 1GbE [8].

In order to be able to distinguish between the traffic classes of service (GST/SM), it is needed to have support of the VLAN tagging. This statement was verified when analysing the problems arising with the xSTP protocols and OpMiGua in chapter 4. The SM traffic will always be blocked unless it is virtually divided from its counterpart. Furthermore, it needs to be switched by one physical port on the incoming side and one on the outgoing side when considering a ring topology. This *SM chain-connectivity* limitation will also limit the statistically multiplexed traffic that can be switched in the case that more than one GST path/wavelength is available in the node. Thus the gained throughput is lower than in [3].

Even though RSTP and MSTP are relatively new STA-based protocols, they still share important drawbacks with STP, such as network underutilization, congestion near the root, and no load balancing. By distributing and directing the traffic over different VLANs, it is possible to achieve a more balanced load across the network and achieve higher utilization of the redundant interconnections. The Link Aggregation Control Protocol [44] could be enabled in the network to enhance the reliability of the SM traffic using multiple physical SM ports in a logical one satisfying the mentioned SM chain-connectivity limitation.

The architecture and node design lack the flexibility of being independent of the number of nodes in the OpMiGua network as it affects the number of wavelengths needed for the SWRON and the number of ports in an Ethernet switch for the GST traffic. However, the number of dedicated SM ports is independent of the number of nodes in the network.

The operation of xSTP protocols is relatively slow. The convergence time in STP is approximately 30-60 seconds while RSTP uses a node by node synchronization protocol that allows the convergence of the whole network in 2-3 seconds [24]. They are becoming not preferable while moving Ethernet in the metro and core networks [31], [32], [37]. One of the main reasons is that the high recovery convergence time leads to high data loss during a down-time for high-throughput networks [7]. Furthermore, in the carrier networks the protection times are preferably less than 50 ms [35] and xSTP does not fulfil the requirements for a carrier transport protocol. When employing Ethernet with OpMiGua, if it is the GST port the one that fails, the guaranteed-service property is no longer offered. Multiple proprietary solutions are used instead in order to achieve re-convergence times comparable with SDH/SONET 50ms recovery time.

The new Ethernet OAM standards [46] define the management and signalling protocols during failure but do not specify the recovery protocol for creating redundant paths. There are different recovery mechanisms used in different implementations. The open RPR and EAPS standards fall short on real networks implementations and basically xSTP is still widely being used today [24]. This is an important reason for the motivation of this work and the ability to introduce Ethernet in the hybrid optical network. However, because of the slow recovery time and the limitations mentioned in chapter 4, we can also conclude that the current Ethernet spanning tree based protocols are hardly efficient with carrier optical networks.

Furthermore, PBB-TE [51] might be used to provide a network solution designed specifically for carrier networks. It creates an independent connection-oriented packet-switched transport layer. This allows various services to be transported transparently through

the network. However, PBB-TE turns off some of the native Ethernet features to realise its MAC addresses management through the control plane. In this case Spanning Tree Protocols are inactive and the network design problems, limitations and reliability issues discussed in our work might be avoided.

Our view is that both domains have their merits and both will be deployed depending on each Service Provider's preference and existing networks. Different applications and customer’s have different dependability requirements on the network. The xSTP recovery times might be tolerable for web browsing customers but with high down-time cost in carrier networks. However, xSTP, PBB-TE and other connection-oriented protocols [54], [55] might as well co-exist within the same SP's networks. Research and development on Ethernet is still an active ongoing process so it is of value to keep working on it.

Chapter 8

Conclusions and Summary

Much research work has been directed toward enhancing the STA-based protocols recovery time after a failure. Furthermore, other research work has focused on hybrid optical packet/circuit switched networks and OpMiGua. However, to our knowledge there are no studies analysing the compatibility of Ethernet and its protection mechanisms in a hybrid optical network. This is the first work addressing the use of Ethernet in OpMiGua. Specifically, the work focuses on the compatibility of the traditional Ethernet's loop-free topology protocols with the redundant multiple traffic service paths of OpMiGua. In this concluding section, we highlight our main achievements in this thesis.

Firstly, after studying Ethernet, its xSTP protocols and OpMiGua we analysed their interoperability. The main problem we found derived from the way the physical connection in the optical domain was logically perceived by the Ethernet node. The dedicated point-to-point lightpaths in an OpMiGua network were recognized as a shared medium by the Ethernet switch. The intermediate nodes responsible for switching the SM traffic did not forward the received BPDUs on the same SM port thus the BPDU transmission through the network was incorrect. This led to a failure in the logical spanning tree convergence process. Different network connectivity scenarios were considered for this analysis.

Secondly, we derived the network physical limitations when implementing a provider's network that integrates both domains with respect to xSTP. The Ethernet switch needs two physically distinct incoming/outgoing ports to switch the SM traffic to its neighbours. We proposed the network architecture *SM chain-connectivity* which creates SM ring connectivity

and full-mesh GST connectivity. This architecture allows for a full spanning tree convergence.

Thirdly, we addressed the problem introduced by xSTP which blocks the redundant paths. The GST and SM ports should be in an active state at any given time in OpMiGua. However, xSTP considers these paths as redundant of each other and blocks them when building the spanning tree. Our solution is to logically differentiate the packets by using the VLAN tag as a label. Different ways of differentiating the traffic through the VLAN tags hierarchy and QoS mappings were proposed. The assignment of VLANs is closely related with the Spanning Tree instances that will allow the interoperation of Ethernet switches with the OpMiGua infrastructure without blocking the traffic. The classification is based on:

1. *the traffic classes,*
2. *the available Spanning Trees in the network.*

MSTP could be used when the diameter of the network surpasses 16 nodes, thus justifying the complexity of using multiple spanning tree instances. However, the use of RSTP with VLAN support is recommended for smaller networks. The proposed architecture ensures that RSTP is used throughout the network because the paths are perceived as point-to-point. Thus, this allows for faster recovery times as compared with legacy STP. However, the per-port VLAN separation solution creates completely different networks. Consequently, there are no redundant links for the SM paths. The resiliency issue might be looked into in further work.

In order to verify the correctness of the analysis and proposed solutions, all the discussed topologies have been simulated through an Ethernet switched network with an emulator. The results show that the analysis and proposed scheme are successful regarding the functional interoperability at the network level.

Finally, we have designed a logical all-optical header processing node as in [3] integrating an Ethernet switch instead of the OPS module. We structure the functionalities expected when integrating the Ethernet switch especially with the focus on the mentioned network physical limitations for a functional xSTP topology.

Chapter 9
Further work

From the findings we have gathered in this thesis, we look at how we can expand and improve on this research as well as speculate possible approaches.

The topologies created by the per-port VLAN separation create completely different networks. Thus, there are no redundant links for the SM paths. This is one of the main limitations of this design. We cannot overcome it by a pure physical solution as we already have shown that only two different and specific in/out SM ports should be assigned to one switching node. A possible solution is to use the Ethernet switch management protocols for this purpose.

Currently two options/trends seem to be dominant, Pure Ethernet with xSTP or PBB-TE as an enhancement for traffic engineering or the MPLS [54], GMPLS [55] and standards compatible or designed specifically for compatibility with optical networks. The xSTP protocols prove unsuitable to fully utilize the performance capacity of an OpMiGua node as in previous works. Since Ethernet is an important technology it is of interest to further look into the available loop-free protocols and adapting them for the OpMiGua characteristics. MPLS over Ethernet uses the Fast ReRoute mechanism for resiliency which achieves recovery times less than 50ms. In addition, an interesting approach is using the S-Tag as an MPLS label for creating connection-oriented paths through VLAN Cross-connect [56]. The GMPLS control plane is used to setup the end-to-end paths throughout the network and to reserve the resources along the path appropriately. The label is considered the S-Tag and the label swapping mechanisms follow the same rule and are locally significant as in GMPLS. This solution is important because it maintains the same Ethernet frame discussed in chapter

3 which allows for the coexistence of bridging functions in the node. The uses of GMPLS resiliency mechanisms either with VLAN cross-connect or not, are less than the desired 50 ms. The use of PBB-TE and MPLS over Ethernet in an OpMiGua underlying network can be analysed and compared with the xSTP recovery mechanism. The comparison could focus on the recovery times and the network architecture limitations discussed in this work.

To conclude, the integrated node design strives to create a functional integrated OpMiGua/Ethernet node. However, as discussed in chapter 5 there are many issues left for discussion and evaluation. Problems introduced by the lack of an Ethernet's synchronization mechanism, the refinement of the control block and especially the electronic header node itself are a wide area for future work. In the scope of the work presented in this thesis the evaluation of the performance of a simple integrated Ethernet/OpMiGua node design with the original OpMiGua node in [8] would be of interest for further work.

References

[1] C.M. Gauger, P.J. Kühn, E. Breusegem, M. Pickavet, and P. Demeester, "Hybrid Optical Network Architectures: Bringing Packets and Circuits Together", in *IEEE Communications Magazine 44*, pp. 36–42, 2006.

[2] The OpMiGua Research Project, available at <http://www.opmigua.com/>, *last checked June 2010.*

[3] S.Bjornstad, D.R. Hjelme, and N.Stol, "A packet switched hybrid optical network with service guarantees", in *IEEE JSAC 24 (8), Supplement on Optical Communications & Networking*, pp. 97-107, August 2006.

[4] H. Frazier, and G. Pesavento, "Ethernet Takes on the First Mile", in *IT Professional*, vol. 3, no. 4, pp. 17-22, July 2001.

[5] A. Kasim, P. Adhikari, N. Chen, N. Finn, and N. Ghani, *Delivering Carrier Ethernet: Extending Ethernet Beyond the LAN***,** McGraw Hill, 2008.

[6] C.F. Lam, and W. Way, "Optical Ethernet: Protocols, Management and 1-100G technologies", *Optical Fiber Telecommunications V, B:Systems and Networks*, chap. 9, pp. 345-401, Elsevier, 2008.

[7] Y. Wong, C. Y. Wong, L. H. Ngoh, and W. C. Wong, "Performance Comparison of Resilient Packet Ring (RPR), Packet Over SDH/SONET (POS) and Gigabit Ethernet (GE) for Network Design," in *International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS)*, pp. 680–687, Montreal, Canada, July 2003.

[8] M. Nord, S. Bjørnstad, O. Austad, V.L. Tuft, D. R. Hjelme, A. S. Sudbø, and L. Eriksen, "OpMiGua Hybrid Circuit- and Packet-Switched Test-Bed Demonstration and

Performance", in *IEEE Photonics Technology Letters*, vol. 18, no. 24, December 15, 2006.

[9] The OpMiGua Project – Final report, Telenor R&I R 32/2006, available at <www.telenor.com/rd/pub/rep06/r_32_06.pdf> , *last checked June 2010.*

[10] A. Kimsas, S. Bjørnstad, H. Overby, and N. Stol, "Reservation Techniques in an OpMiGua Node." in *Proc. Conference on Optical Network Design and Modelling (ONDM),*Greece, 2007.

[11] J. Scharf, A. Kimsas, M. Köhn, and G. Hu, "OBS vs. OpMiGua – A Comparative Performance Evaluation", in *Proceedings of the 9th International Conference on Transparent Optical Networks (ICTON 2007)*, pp. 294–298, Rome 2007.

[12] Careglio et al., "Performance Issues in Optical Burst/Packet Switching", *Towards Digital Optical Networks, Springer Lecture Notes on Computer Science*, vol. 5412, 2009.

[13] S. Bjørnstad, H. Øverby, N. Stol, and D.R. Hjelme, "Protecting guaranteed service traffic in an OpMiGua hybrid network", in *Proceedings of the 31th European Conference on Optical Communications (ECOC)*, vol. 3, September 25-29, 2005, Glasgow, Scotland.

[14] E. Van Breusegem, J. Cheyns, D. De Winter, D. Colle, M. Pickavet, P. Demeester, and J. Moreau, "A broad view on overspill routing in optical networks: A real synthesis of packet and circuit switching?," *Journal on Optical Switched Networks*, vol. 1, no. 1, pp. 51–64, 2004.

[15] H. Øverby, "Network layer packet redundancy in optical packet switched networks", *Optics Express 12*, pages 4881-4895, 2004.

[16] S. Bjørnstad, *Packet switching in optical networks, PhD thesis*, Norwegian University of Science and Technology (NTNU), July 2004.

[17] A. Kimsas, S. Bjornstad, H. Overby, and N. Stol, "Protection Using Redundancy in a Hybrid Circuit/Packet Node Design" , *Proceedings of the European Conference on Optical Communications (ECOC)*, 2006.

[18] M. Nord, “Hybrid Wavelength Routed and Optical Packet Switched Ring Networks for the Metropolitan Area Networks”, in *Proc. of the International Conference on Transparent Optical Networks* (ICTON), 2005.

[19] E. Van Breusegem, J. Cheyns, D. De Winter, D. Colle, M. Pickavet, P. Demeester, and J. Moreau, “Overspill routing in optical networks: a true hybrid optical network design”, *IEEE Journal of Selected Areas of Communication* ***24*** *(Suppl 4),*pp. 13–26, 2006.

[20] S. Bjørnstad, M. Nord, T. Olsen, D. Hjelme, and N. Stol, ”Burst, packet and hybrid switching in the optical core network”, *TELEKTRONIKK*, 2005.

[21] R.Perlman, *Interconnections (2nd ed.): bridges, routers, switches, and internetworking protocols*, Addison-Wesley, Boston, 1999.

[22] R. Perlman, “An algorithm for distributed computation of a spanning tree in an extended LAN”, *SIGCOMM Computer Communications Rev.* 15, pp. 44-53, September 1985.

[23] G. Ibáñez, A. Garcıa-Martınez, A. Azcorra, and I. Soto, “ABridges: Scalable, self-configuring Ethernet campus networks”, *Computer Networks*, vol. 52 (3), 22 February 2008.

[24] M. Huynh, S. Goose, and P. Mohapatra, “Resilience technologies in Ethernet”, *Computer Networks*, vol. 54(1), pp. 57-78, January 2010.

[25] K. Webb, and K. Bagwell, *Building Cisco Multilayered Switched Networks*, Cisco Press, 2006.

[26] B.Y.Wu, and K. Chao, *Spanning Trees and Optimization Problems*, Chapman & Hall, Boca Raton, 2004.

[27] IEEE 802.1D, *IEEE Standard for Local and Metropolitan Area Networks--Media access control (MAC) Bridges* (Incorporates IEEE 802.1t-2001 and IEEE 802.1w), IEEE standardization, 2004.

[28] IEEE 802.1Q, *Virtual LANs*, IEEE standardization, 2006

[29] *Understanding Multiple Spanning Tree Protocol (802.1s)*, Cisco, available at: <http://chand.lums.edu.pk/cs573/resources/MST_802.1s.pdf>, *last checked June 2010*.

[30] *Understanding Rapid Spanning Tree Protocol (802.1w)*, Cisco, available at: <http://chand.lums.edu.pk/cs573/resources/MST_802.1s.pdf>, *last checked June 2010*.

[31] M. Huynh, P Mohapatra, and S. Goose, "Spanning Tree Elevation Protocol: enhancing metro Ethernet performance and QoS", *Computer Communications*, Elsevier, 2009.

[32] M. Huynh, P. Mohapatra, and S. Goose, "Cross-Over Spanning Trees: Enhancing Metro Ethernet Resilience and Load Balancing", in *Proceedings of IEEE BroadNets Conference*, 2007.

[33] F. C. Gartner, "A Survey of Self-Stabilizing Spanning-Tree Construction Algorithms", Technical Report IC/2003/38, EPFL, Technical Reports in Computer and Communication Sciences, 2003.

[34] F. De Pellegrini, D. Starobinski, M. G. Karpovsky, and L. B. Levitin, "Scalable, distributed cycle-breaking algorithms for Gigabit Ethernet backbones", *Journal of Optical Networks*, 5(1):1, 2006.

[35] K Fouli, and M Maier, "The road to carrier-grade Ethernet", *IEEE Communications Magazine*, 2009.

[36] A. Reid, P. Willis, I. Hawkins, and C. Bitrons, "Carrier Ethernet", *IEEE Communications Magazine*, 2008.

[37] L. Fang, N. Bitar, R. Zhang, and M. Taylor, "The evolution of carrier Ethernet services— requirements and deployment case studies", *IEEE Communications Magazine* 46(3):69–76, 2008.

[38] IEEE 802.3 Ethernet (CSMA/CD), IEEE Standards, 2008.

[39] D. Allan, N. Bragg, A. McGuire, and A. Reid, "Ethernet as Carrier Transport infrastructure", *IEEE Communications Magazine*, 2006.

[40] IEEE STD 802.1ah, "IEEE Standard for Local and metropolitan area networks Virtual Bridged Local Area Networks Amendment 7: Provider Backbone Bridges", IEEE Standards, 2008.

[41] M. Huynh, and P. Mohapatra. “Metropolitan Ethernet Network: A move from LAN to MAN2”, *Computer Networks*, Volume 51, Issue 17, p. 4867-4894, 5 December 2007.

[42] “PBB-TE, PBT: Carrier Grade Ethernet Transport”, TPACK white paper, available at <http://www.tpack.com/resources/tpack-white-papers/pbb-te-pbt.html>, *last checked June 2010*.

[43] R. Sanchez, L. Raptis, and K. Vaxevanakis, “Ethernet as a Carrier Grade Technology: Developments and Innovations”, *IEEE Communication Magazine*, vol. 46(9), p. 88-94, 2008.

[44] IEEE 802.1AX-2008, “IEEE Standard for Local and Metropolitan Area Networks--Link Aggregation”, IEEE Standards, 2008.

[45] IEEE 802.17, “Resilient packet ring (RPR) access method & physical layer specifications”, IEEE Standards, 2004.

[46] IEEE 802.1ag, “*802.1ag* - Connectivity Fault Management”, IEEE Standards, 2007.

[47] Graphical Network Simulator based on Dynamips and Dynagen, available at http://www.gns3.net/, *last checked June 2010*.

[48] C. Nicolau, and D. Sala, “Flat soft-synchronization for Gigabit Ethernet”, *Proc. of the 28th IEEE international conference on Computer Communications Workshops*, p. 357-358, Brazil 2009.

[49] M.S. Borella, J.P. Jue, B. Ramamurthy, and B. Mukherjee, “Optical components for WDM Lightwave Networks”, in *Proc. IEEE*, pp. 1274-1307, 1997.

[50] G. Das, R. S. Tucker, C. Leckie, and K. Hinton, “Multiple-input single-output fifo optical buffers with controllable fractional delay lines,” *Optics Express*, Vol. 16, Issue 26, pp. 21849-21864, 2008.

[51] IEEE 802.1Qay, “IEEE Standard for Local and Metropolitan Area Networks---Virtual Bridged Local Area Networks - Amendment: Provider Backbone Bridge Traffic Engineering", 2007.

[52] J.M. Finochietto, F. Neri, K. Wajda, R. Watza, J. Domzal, M. Nord, and E. Zouganeli, "Towards optical packet switched MANs: Design issues and tradeoffs", *Optical Switching and Networking*, Volume 5, Issue 4, p. 253-267, October 2008.

[53] S. Baroni, and P. Bayvel. "Wavelength requirements in an arbitrarily connected Wavelength-Routed Optical Networks", *Journal of Lightwave Technology*, vol. 15, no. 2, 1997.

[54] RFC 3031, "Multiprotocol Label Switching Architecture", IETF standards, 2001.

[55] RFC 3945, "Generalized Multi-Protocol Label Switching (GMPLS) Architecture", IETF standard, 2004.

[56] IETF RFC draft, "GMPLS Control of Ethernet VLAN Cross Connect Switches", March 2006, available at < http://ietfreport.isoc.org/idref/draft-sprecher-gels-ethernet-vlan-xc>, *last checked June 2010*.

Appendix A

Case 1 STP topology without VLANs

```
R1#show spanning-tree
 VLAN1 is executing the ieee compatible Spanning Tree protocol
  Bridge Identifier has priority 32768, address c209.6e07.0000
  Configured hello time 2, max age 20, forward delay 15
  We are the root of the spanning tree
  Topology change flag not set, detected flag not set
  Number of topology changes 3 last change occurred 00:18:02 ago
        from FastEthernet1/1
  Times:  hold 1, topology change 35, notification 2
        hello 2, max age 20, forward delay 15
  Timers: hello 1, topology change 0, notification 0, aging 300

 Port 41 (FastEthernet1/0) of VLAN1 is forwarding
   Port path cost 19, Port priority 128, Port Identifier 128.41.
   Designated root has priority 32768, address c209.6e07.0000
   Designated bridge has priority 32768, address c209.6e07.0000
   Designated port id is 128.41, designated path cost 0
   Timers: message age 0, forward delay 0, hold 0
   Number of transitions to forwarding state: 1
   BPDU: sent 712, received 2

 Port 42 (FastEthernet1/1) of VLAN1 is forwarding
   Port path cost 19, Port priority 128, Port Identifier 128.42.
   Designated root has priority 32768, address c209.6e07.0000
   Designated bridge has priority 32768, address c209.6e07.0000
   Designated port id is 128.42, designated path cost 0
   Timers: message age 0, forward delay 0, hold 0
   Number of transitions to forwarding state: 1
   BPDU: sent 713, received 2

 Port 43 (FastEthernet1/2) of VLAN1 is forwarding
   Port path cost 19, Port priority 128, Port Identifier 128.43.
   Designated root has priority 32768, address c209.6e07.0000
   Designated bridge has priority 32768, address c209.6e07.0000
```

Designated port id is 128.43, designated path cost 0
Timers: message age 0, forward delay 0, hold 0
Number of transitions to forwarding state: 1
BPDU: sent 713, received 1

Port 56 (FastEthernet1/15) of VLAN1 is forwarding
Port path cost 19, Port priority 128, Port Identifier 128.56.
Designated root has priority 32768, address c209.6e07.0000
Designated bridge has priority 32768, address c209.6e07.0000
Designated port id is 128.56, designated path cost 0
Timers: message age 0, forward delay 0, hold 0
Number of transitions to forwarding state: 1
BPDU: sent 713, received 1

R1#show spanning-tree summary
Root bridge for: VLAN1.
PortFast BPDU Guard is disabled
UplinkFast is disabled
BackboneFast is disabled

Name	Blocking	Listening	Learning	Forwarding	STP Active
VLAN1	0	0	0	4	4
1 VLAN	0	0	0	4	4

R2#show spanning-tree

VLAN1 is executing the ieee compatible Spanning Tree protocol
Bridge Identifier has priority 32768, address c20a.6e07.0000
Configured hello time 2, max age 20, forward delay 15
Current root has priority 32768, address c209.6e07.0000
Root port is 41 (FastEthernet1/0), cost of root path is 19
Topology change flag not set, detected flag not set
Number of topology changes 2 last change occurred 00:26:12 ago
from FastEthernet1/15
Times: hold 1, topology change 35, notification 2
hello 2, max age 20, forward delay 15

Timers: hello 0, topology change 0, notification 0, aging 300

Port 41 (FastEthernet1/0) of VLAN1 is forwarding
Port path cost 19, Port priority 128, Port Identifier 128.41.
Designated root has priority 32768, address c209.6e07.0000
Designated bridge has priority 32768, address c209.6e07.0000
Designated port id is 128.41, designated path cost 0
Timers: message age 2, forward delay 0, hold 0
Number of transitions to forwarding state: 1
BPDU: sent 2, received 802

Port 42 (FastEthernet1/1) of VLAN1 is forwarding
Port path cost 19, Port priority 128, Port Identifier 128.42.
Designated root has priority 32768, address c209.6e07.0000
Designated bridge has priority 32768, address c20a.6e07.0000
Designated port id is 128.42, designated path cost 19
Timers: message age 0, forward delay 0, hold 0
Number of transitions to forwarding state: 1
BPDU: sent 802, received 0

Port 43 (FastEthernet1/2) of VLAN1 is forwarding
Port path cost 19, Port priority 128, Port Identifier 128.43.
Designated root has priority 32768, address c209.6e07.0000
Designated bridge has priority 32768, address c20a.6e07.0000
Designated port id is 128.43, designated path cost 19
Timers: message age 0, forward delay 0, hold 0
Number of transitions to forwarding state: 1
BPDU: sent 803, received 0

Port 55 (FastEthernet1/14) of VLAN1 is blocking
Port path cost 19, Port priority 128, Port Identifier 128.55.
Designated root has priority 32768, address c209.6e07.0000
Designated bridge has priority 32768, address c209.6e07.0000
Designated port id is 128.56, designated path cost 0
Timers: message age 1, forward delay 0, hold 0
Number of transitions to forwarding state: 0
BPDU: sent 1, received 803

Port 56 (FastEthernet1/15) of VLAN1 is forwarding
Port path cost 19, Port priority 128, Port Identifier 128.56.

Designated root has priority 32768, address c209.6e07.0000
Designated bridge has priority 32768, address c20a.6e07.0000
Designated port id is 128.56, designated path cost 19
Timers: message age 0, forward delay 0, hold 0
Number of transitions to forwarding state: 1
BPDU: sent 804, received 1

R2#show spanning-tree summary
Root bridge for: none.
PortFast BPDU Guard is disabled
UplinkFast is disabled
BackboneFast is disabled

Name	Blocking	Listening	Learning	Forwarding	STP Active
VLAN1	1	0	0	4	5
1 VLAN 1	0	0	4	5	

R3#show spanning-tree

VLAN1 is executing the ieee compatible Spanning Tree protocol
Bridge Identifier has priority 32768, address c20b.6e07.0000
Configured hello time 2, max age 20, forward delay 15
Current root has priority 32768, address c209.6e07.0000
Root port is 41 (FastEthernet1/0), cost of root path is 19
Topology change flag not set, detected flag not set
Number of topology changes 1 last change occurred 00:23:50 ago
from FastEthernet1/0
Times: hold 1, topology change 35, notification 2
hello 2, max age 20, forward delay 15
Timers: hello 0, topology change 0, notification 0, aging 300

Port 41 (FastEthernet1/0) of VLAN1 is forwarding
Port path cost 19, Port priority 128, Port Identifier 128.41.
Designated root has priority 32768, address c209.6e07.0000
Designated bridge has priority 32768, address c209.6e07.0000
Designated port id is 128.42, designated path cost 0
Timers: message age 1, forward delay 0, hold 0

Number of transitions to forwarding state: 1
BPDU: sent 2, received 737

Port 42 (FastEthernet1/1) of VLAN1 is blocking
Port path cost 19, Port priority 128, Port Identifier 128.42.
Designated root has priority 32768, address c209.6e07.0000
Designated bridge has priority 32768, address c20a.6e07.0000
Designated port id is 128.42, designated path cost 19
Timers: message age 2, forward delay 0, hold 0
Number of transitions to forwarding state: 0
BPDU: sent 0, received 732

Port 43 (FastEthernet1/2) of VLAN1 is forwarding
Port path cost 19, Port priority 128, Port Identifier 128.43.
Designated root has priority 32768, address c209.6e07.0000
Designated bridge has priority 32768, address c20b.6e07.0000
Designated port id is 128.43, designated path cost 19
Timers: message age 0, forward delay 0, hold 0
Number of transitions to forwarding state: 1
BPDU: sent 728, received 1

Port 55 (FastEthernet1/14) of VLAN1 is blocking
Port path cost 19, Port priority 128, Port Identifier 128.55.
Designated root has priority 32768, address c209.6e07.0000
Designated bridge has priority 32768, address c20a.6e07.0000
Designated port id is 128.56, designated path cost 19
Timers: message age 2, forward delay 0, hold 0
Number of transitions to forwarding state: 0
BPDU: sent 1, received 734

Port 56 (FastEthernet1/15) of VLAN1 is forwarding
Port path cost 19, Port priority 128, Port Identifier 128.56.
Designated root has priority 32768, address c209.6e07.0000
Designated bridge has priority 32768, address c20b.6e07.0000
Designated port id is 128.56, designated path cost 19
Timers: message age 0, forward delay 0, hold 0
Number of transitions to forwarding state: 1
BPDU: sent 728, received 1

R3#show spanning-tree summary

Root bridge for: none.

PortFast BPDU Guard is disabled

UplinkFast is disabled

BackboneFast is disabled

Name	Blocking	Listening	Learning	Forwarding	STP Active
VLAN1	2	0	0	3	5
1 VLAN	2	0	0	3	5

R4#show spanning-tree

VLAN1 is executing the ieee compatible Spanning Tree protocol

Bridge Identifier has priority 32768, address c20c.6e07.0000

Configured hello time 2, max age 20, forward delay 15

Current root has priority 32768, address c209.6e07.0000

Root port is 41 (FastEthernet1/0), cost of root path is 19

Topology change flag not set, detected flag not set

Number of topology changes 0 last change occurred 00:22:14 ago

Times: hold 1, topology change 35, notification 2

hello 2, max age 20, forward delay 15

Timers: hello 0, topology change 0, notification 0, aging 300

Port 41 (FastEthernet1/0) of VLAN1 is forwarding

Port path cost 19, Port priority 128, Port Identifier 128.41.

Designated root has priority 32768, address c209.6e07.0000

Designated bridge has priority 32768, address c209.6e07.0000

Designated port id is 128.43, designated path cost 0

Timers: message age 2, forward delay 0, hold 0

Number of transitions to forwarding state: 1

BPDU: sent 1, received 664

Port 42 (FastEthernet1/1) of VLAN1 is blocking

Port path cost 19, Port priority 128, Port Identifier 128.42.

Designated root has priority 32768, address c209.6e07.0000

Designated bridge has priority 32768, address c20a.6e07.0000

Designated port id is 128.43, designated path cost 19

Timers: message age 3, forward delay 0, hold 0

Number of transitions to forwarding state: 0
BPDU: sent 0, received 664

Port 43 (FastEthernet1/2) of VLAN1 is blocking
Port path cost 19, Port priority 128, Port Identifier 128.43.
Designated root has priority 32768, address c209.6e07.0000
Designated bridge has priority 32768, address c20b.6e07.0000
Designated port id is 128.43, designated path cost 19
Timers: message age 2, forward delay 0, hold 0
Number of transitions to forwarding state: 0
BPDU: sent 1, received 668

Port 55 (FastEthernet1/14) of VLAN1 is blocking
Port path cost 19, Port priority 128, Port Identifier 128.55.
Designated root has priority 32768, address c209.6e07.0000
Designated bridge has priority 32768, address c20b.6e07.0000
Designated port id is 128.56, designated path cost 19
Timers: message age 2, forward delay 0, hold 0
Number of transitions to forwarding state: 0
BPDU: sent 1, received 664

R4#show spanning-tree summary
Root bridge for: none.
PortFast BPDU Guard is disabled
UplinkFast is disabled
BackboneFast is disabled

Name	Blocking	Listening	Learning	Forwarding	STP Active
VLAN1	3	0	0	1	4
1 VLAN	3	0	0	1	4

Case 2 xSTP with VLAN separation of GST/SM

VLAN 1 is the default which we do not change.
VLAN 2 is the GST traffic meaning all f1/0 to f1/2 ports of all the switches
VLAN 3 is the SM traffic meaning all the f1/14 and f1/15 ports of all the switches

R1#show spanning-tree brief

VLAN2

Spanning tree enabled protocol ieee

Root ID Priority 32768

Address c209.6e07.0000

Cost 171

Port 41 (FastEthernet1/0)

Hello Time 2 sec Max Age 20 sec Forward Delay 15 sec

Bridge ID Priority 32768

Address c209.6e07.0001

Hello Time 2 sec Max Age 20 sec Forward Delay 15 sec

Aging Time 300

Interface Name	Port ID	Prio	Cost	Sts	Cost	Bridge ID	Port ID
FastEthernet1/0	128.41	128	19	FWD	152	32768 c20a.6e07.0001	128.41
FastEthernet1/1	128.42	128	19	FWD	171	32768 c209.6e07.0001	128.42
FastEthernet1/2	128.43	128	19	FWD	171	32768 c209.6e07.0001	128.43

VLAN3

Spanning tree enabled protocol ieee

Root ID Priority 32768

Address c209.6e07.0000

This bridge is the root

Hello Time 2 sec Max Age 20 sec Forward Delay 15 sec

Bridge ID Priority 32768

Address c209.6e07.0000

Hello Time 2 sec Max Age 20 sec Forward Delay 15 sec

Aging Time 300

Interface Name	Port ID	Prio	Cost	Sts	Cost	Bridge ID	Port ID
FastEthernet1/15	128.56	128	19	FWD	0	32768 c209.6e07.0000	128.56

R1#show spanning-tree

VLAN2 is executing the ieee compatible Spanning Tree protocol
Bridge Identifier has priority 32768, address c209.6e07.0001
Configured hello time 2, max age 20, forward delay 15
We are the root of the spanning tree
Topology change flag set, detected flag set
Number of topology changes 23 last change occurred 00:00:18 ago
from FastEthernet1/1
Times: hold 1, topology change 35, notification 2
hello 2, max age 20, forward delay 15
Timers: hello 0, topology change 32, notification 0, aging 300

Port 41 (FastEthernet1/0) of VLAN2 is forwarding
Port path cost 19, Port priority 128, Port Identifier 128.41.
Designated root has priority 32768, address c209.6e07.0001
Designated bridge has priority 32768, address c209.6e07.0001
Designated port id is 128.41, designated path cost 0
Timers: message age 0, forward delay 0, hold 0
Number of transitions to forwarding state: 1
BPDU: sent 130, received 125

Port 42 (FastEthernet1/1) of VLAN2 is forwarding
Port path cost 19, Port priority 128, Port Identifier 128.42.
Designated root has priority 32768, address c209.6e07.0001
Designated bridge has priority 32768, address c209.6e07.0001
Designated port id is 128.42, designated path cost 0
Timers: message age 0, forward delay 0, hold 0
Number of transitions to forwarding state: 2
BPDU: sent 154, received 84

Port 43 (FastEthernet1/2) of VLAN2 is forwarding
Port path cost 19, Port priority 128, Port Identifier 128.43.
Designated root has priority 32768, address c209.6e07.0001
Designated bridge has priority 32768, address c209.6e07.0001
Designated port id is 128.43, designated path cost 0
Timers: message age 0, forward delay 0, hold 0
Number of transitions to forwarding state: 3
BPDU: sent 177, received 44

VLAN3 is executing the ieee compatible Spanning Tree protocol
Bridge Identifier has priority 32768, address c209.6e07.0000
Configured hello time 2, max age 20, forward delay 15
We are the root of the spanning tree
Topology change flag not set, detected flag not set
Number of topology changes 2 last change occurred 00:01:25 ago
from FastEthernet1/15
Times: hold 1, topology change 35, notification 2
hello 2, max age 20, forward delay 15
Timers: hello 0, topology change 0, notification 0, aging 300

Port 56 (FastEthernet1/15) of VLAN3 is forwarding
Port path cost 19, Port priority 128, Port Identifier 128.56.
Designated root has priority 32768, address c209.6e07.0000
Designated bridge has priority 32768, address c209.6e07.0000
Designated port id is 128.56, designated path cost 0
Timers: message age 0, forward delay 0, hold 0
Number of transitions to forwarding state: 1
BPDU: sent 104, received 12

R2#show spanning-tree brief

VLAN2
Spanning tree enabled protocol ieee
Root ID Priority 32768
Address c209.6e07.0001
Cost 19
Port 41 (FastEthernet1/0)
Hello Time 2 sec Max Age 20 sec Forward Delay 15 sec

Bridge ID Priority 32768
Address c20a.6e07.0001
Hello Time 2 sec Max Age 20 sec Forward Delay 15 sec
Aging Time 300

Interface Name	Port ID	Prio	Cost	Sts	Cost	Bridge ID	Port ID
FastEthernet1/0	128.41	128	19	FWD	0	32768 c209.6e07.0001	128.41

FastEthernet1/1 128.42 128 19 FWD 19 32768 c20a.6e07.0001 128.42
FastEthernet1/2 128.43 128 19 FWD 19 32768 c20a.6e07.0001 128.43

VLAN3
Spanning tree enabled protocol ieee
Root ID Priority 32768
Address c209.6e07.0000
Cost 19
Port 55 (FastEthernet1/14)
Hello Time 2 sec Max Age 20 sec Forward Delay 15 sec

Bridge ID Priority 32768
Address c20a.6e07.0002
Hello Time 2 sec Max Age 20 sec Forward Delay 15 sec
Aging Time 300

Interface Name	Port ID	Prio	Cost	Sts	Cost	Designated Bridge ID	Port ID
FastEthernet1/14	128.55	128	19	FWD	0	32768 c209.6e07.0000	128.56
FastEthernet1/15	128.56	128	19	FWD	19	32768 c20a.6e07.0002	128.56

R2#show spanning-tree

VLAN2 is executing the ieee compatible Spanning Tree protocol
Bridge Identifier has priority 32768, address c20a.6e07.0001
Configured hello time 2, max age 20, forward delay 15
Current root has priority 32768, address c209.6e07.0001
Root port is 41 (FastEthernet1/0), cost of root path is 19
Topology change flag set, detected flag not set
Number of topology changes 13 last change occurred 00:00:34 ago
from FastEthernet1/1
Times: hold 1, topology change 35, notification 2
hello 2, max age 20, forward delay 15
Timers: hello 0, topology change 0, notification 0, aging 300

Port 41 (FastEthernet1/0) of VLAN2 is forwarding
Port path cost 19, Port priority 128, Port Identifier 128.41.

Designated root has priority 32768, address c209.6e07.0001
Designated bridge has priority 32768, address c209.6e07.0001
Designated port id is 128.41, designated path cost 0
Timers: message age 2, forward delay 0, hold 0
Number of transitions to forwarding state: 1
BPDU: sent 53, received 135

Port 42 (FastEthernet1/1) of VLAN2 is forwarding
Port path cost 19, Port priority 128, Port Identifier 128.42.
Designated root has priority 32768, address c209.6e07.0001
Designated bridge has priority 32768, address c20a.6e07.0001
Designated port id is 128.42, designated path cost 19
Timers: message age 0, forward delay 0, hold 0
Number of transitions to forwarding state: 2
BPDU: sent 101, received 95

Port 43 (FastEthernet1/2) of VLAN2 is forwarding
Port path cost 19, Port priority 128, Port Identifier 128.43.
Designated root has priority 32768, address c209.6e07.0001
Designated bridge has priority 32768, address c20a.6e07.0001
Designated port id is 128.43, designated path cost 19
Timers: message age 0, forward delay 0, hold 0
Number of transitions to forwarding state: 1
BPDU: sent 130, received 45

VLAN3 is executing the ieee compatible Spanning Tree protocol
Bridge Identifier has priority 32768, address c20a.6e07.0002
Configured hello time 2, max age 20, forward delay 15
Current root has priority 32768, address c209.6e07.0000
Root port is 55 (FastEthernet1/14), cost of root path is 19
Topology change flag not set, detected flag not set
Number of topology changes 15 last change occurred 00:01:39 ago
from FastEthernet1/15
Times: hold 1, topology change 35, notification 2
hello 2, max age 20, forward delay 15
Timers: hello 0, topology change 0, notification 0, aging 300

Port 55 (FastEthernet1/14) of VLAN3 is forwarding
Port path cost 19, Port priority 128, Port Identifier 128.55.

Designated root has priority 32768, address c209.6e07.0000
Designated bridge has priority 32768, address c209.6e07.0000
Designated port id is 128.56, designated path cost 0
Timers: message age 2, forward delay 0, hold 0
Number of transitions to forwarding state: 1
BPDU: sent 15, received 149

Port 56 (FastEthernet1/15) of VLAN3 is forwarding
Port path cost 19, Port priority 128, Port Identifier 128.56.
Designated root has priority 32768, address c209.6e07.0000
Designated bridge has priority 32768, address c20a.6e07.0002
Designated port id is 128.56, designated path cost 19
Timers: message age 0, forward delay 0, hold 0
Number of transitions to forwarding state: 1
BPDU: sent 137, received 22

R3#show spanning-tree brief

VLAN2
Spanning tree enabled protocol ieee
Root ID Priority 32768
Address c209.6e07.0001
Cost 19
Port 41 (FastEthernet1/0)
Hello Time 2 sec Max Age 20 sec Forward Delay 15 sec

Bridge ID Priority 32768
Address c20b.6e07.0001
Hello Time 2 sec Max Age 20 sec Forward Delay 15 sec
Aging Time 300

Interface Name	Port ID	Prio	Cost	Sts	Cost	Designated Bridge ID	Designated Port ID
FastEthernet1/0	128.41	128	19	FWD	0	32768 c209.6e07.0001	128.42
FastEthernet1/1	128.42	128	19	BLK	19	32768 c20a.6e07.0001	128.42
FastEthernet1/2	128.43	128	19	FWD	19	32768 c20b.6e07.0001	128.43

VLAN3

Spanning tree enabled protocol ieee

Root ID Priority 32768

Address c209.6e07.0000

Cost 38

Port 55 (FastEthernet1/14)

Hello Time 2 sec Max Age 20 sec Forward Delay 15 sec

Bridge ID Priority 32768

Address c20b.6e07.0002

Hello Time 2 sec Max Age 20 sec Forward Delay 15 sec

Aging Time 300

Interface Name	Port ID	Prio	Cost	Sts	Cost	Designated Bridge ID	Port ID
FastEthernet1/14	128.55	128	19	FWD	19	32768 c20a.6e07.0002	128.56
FastEthernet1/15	128.56	128	19	FWD	38	32768 c20b.6e07.0002	128.56

R3#show spanning-tree

VLAN2 is executing the ieee compatible Spanning Tree protocol

Bridge Identifier has priority 32768, address c20b.6e07.0001

Configured hello time 2, max age 20, forward delay 15

Current root has priority 32768, address c209.6e07.0001

Root port is 41 (FastEthernet1/0), cost of root path is 19

Topology change flag not set, detected flag not set

Number of topology changes 15 last change occurred 00:01:02 ago

from FastEthernet1/2

Times: hold 1, topology change 35, notification 2

hello 2, max age 20, forward delay 15

Timers: hello 0, topology change 0, notification 0, aging 300

Port 41 (FastEthernet1/0) of VLAN2 is forwarding

Port path cost 19, Port priority 128, Port Identifier 128.41.

Designated root has priority 32768, address c209.6e07.0001

Designated bridge has priority 32768, address c209.6e07.0001
Designated port id is 128.42, designated path cost 0
Timers: message age 3, forward delay 0, hold 0
Number of transitions to forwarding state: 2
BPDU: sent 33, received 105

Port 42 (FastEthernet1/1) of VLAN2 is blocking
Port path cost 19, Port priority 128, Port Identifier 128.42.
Designated root has priority 32768, address c209.6e07.0001
Designated bridge has priority 32768, address c20a.6e07.0001
Designated port id is 128.42, designated path cost 19
Timers: message age 2, forward delay 0, hold 0
Number of transitions to forwarding state: 0
BPDU: sent 38, received 101

Port 43 (FastEthernet1/2) of VLAN2 is forwarding
Port path cost 19, Port priority 128, Port Identifier 128.43.
Designated root has priority 32768, address c209.6e07.0001
Designated bridge has priority 32768, address c20b.6e07.0001
Designated port id is 128.43, designated path cost 19
Timers: message age 0, forward delay 0, hold 0
Number of transitions to forwarding state: 2
BPDU: sent 83, received 46

VLAN3 is executing the ieee compatible Spanning Tree protocol
Bridge Identifier has priority 32768, address c20b.6e07.0002
Configured hello time 2, max age 20, forward delay 15
Current root has priority 32768, address c209.6e07.0000
Root port is 55 (FastEthernet1/14), cost of root path is 38
Topology change flag not set, detected flag not set
Number of topology changes 8 last change occurred 00:02:21 ago
from FastEthernet1/15
Times: hold 1, topology change 35, notification 2
hello 2, max age 20, forward delay 15
Timers: hello 0, topology change 0, notification 0, aging 300

Port 55 (FastEthernet1/14) of VLAN3 is forwarding
Port path cost 19, Port priority 128, Port Identifier 128.55.
Designated root has priority 32768, address c209.6e07.0000

Designated bridge has priority 32768, address c20a.6e07.0002
Designated port id is 128.56, designated path cost 19
Timers: message age 3, forward delay 0, hold 0
Number of transitions to forwarding state: 1
BPDU: sent 8, received 101

Port 56 (FastEthernet1/15) of VLAN3 is forwarding
Port path cost 19, Port priority 128, Port Identifier 128.56.
Designated root has priority 32768, address c209.6e07.0000
Designated bridge has priority 32768, address c20b.6e07.0002
Designated port id is 128.56, designated path cost 38
Timers: message age 0, forward delay 0, hold 0
Number of transitions to forwarding state: 1
BPDU: sent 89, received 19

R4#show spanning-tree brief

VLAN2
Spanning tree enabled protocol ieee
Root ID Priority 32768
Address c209.6e07.0001
Cost 19
Port 41 (FastEthernet1/0)
Hello Time 2 sec Max Age 20 sec Forward Delay 15 sec

Bridge ID Priority 32768
Address c20c.6e07.0001
Hello Time 2 sec Max Age 20 sec Forward Delay 15 sec
Aging Time 300

Interface Name	Port ID	Prio	Cost	Sts	Cost	Designated Bridge ID	Designated Port ID
FastEthernet1/0	128.41	128	19	FWD	0	32768 c209.6e07.0001	128.43
FastEthernet1/1	128.42	128	19	BLK	19	32768 c20a.6e07.0001	128.43
FastEthernet1/2	128.43	128	19	BLK	19	32768 c20b.6e07.0001	128.43

VLAN3
Spanning tree enabled protocol ieee

Root ID Priority 32768

Address c209.6e07.0000

Cost 57

Port 55 (FastEthernet1/14)

Hello Time 2 sec Max Age 20 sec Forward Delay 15 sec

Bridge ID Priority 32768

Address c20c.6e07.0000

Hello Time 2 sec Max Age 20 sec Forward Delay 15 sec

Aging Time 300

Interface Name	Port ID	Prio	Cost	Sts	Cost	Designated Bridge ID	Designated Port ID
FastEthernet1/14	128.55	128	19	FWD	38	32768 c20b.6e07.0002	128.56

R4#show spanning-tree

VLAN2 is executing the ieee compatible Spanning Tree protocol

Bridge Identifier has priority 32768, address c20c.6e07.0001

Configured hello time 2, max age 20, forward delay 15

Current root has priority 32768, address c209.6e07.0001

Root port is 41 (FastEthernet1/0), cost of root path is 19

Topology change flag not set, detected flag not set

Number of topology changes 3 last change occurred 00:02:38 ago

from FastEthernet1/2

Times: hold 1, topology change 35, notification 2

hello 2, max age 20, forward delay 15

Timers: hello 0, topology change 0, notification 0, aging 300

Port 41 (FastEthernet1/0) of VLAN2 is forwarding

Port path cost 19, Port priority 128, Port Identifier 128.41.

Designated root has priority 32768, address c209.6e07.0001

Designated bridge has priority 32768, address c209.6e07.0001

Designated port id is 128.43, designated path cost 0

Timers: message age 2, forward delay 0, hold 0

Number of transitions to forwarding state: 2

BPDU: sent 18, received 106

Port 42 (FastEthernet1/1) of VLAN2 is blocking
Port path cost 19, Port priority 128, Port Identifier 128.42.
Designated root has priority 32768, address c209.6e07.0001
Designated bridge has priority 32768, address c20a.6e07.0001
Designated port id is 128.43, designated path cost 19
Timers: message age 3, forward delay 0, hold 0
Number of transitions to forwarding state: 0
BPDU: sent 18, received 103

Port 43 (FastEthernet1/2) of VLAN2 is blocking
Port path cost 19, Port priority 128, Port Identifier 128.43.
Designated root has priority 32768, address c209.6e07.0001
Designated bridge has priority 32768, address c20b.6e07.0001
Designated port id is 128.43, designated path cost 19
Timers: message age 2, forward delay 0, hold 0
Number of transitions to forwarding state: 1
BPDU: sent 19, received 99

VLAN3 is executing the ieee compatible Spanning Tree protocol
Bridge Identifier has priority 32768, address c20c.6e07.0000
Configured hello time 2, max age 20, forward delay 15
Current root has priority 32768, address c209.6e07.0000
Root port is 55 (FastEthernet1/14), cost of root path is 57
Topology change flag not set, detected flag not set
Number of topology changes 0 last change occurred 00:03:20 ago
Times: hold 1, topology change 35, notification 2
hello 2, max age 20, forward delay 15
Timers: hello 0, topology change 0, notification 0, aging 300

Port 55 (FastEthernet1/14) of VLAN3 is forwarding
Port path cost 19, Port priority 128, Port Identifier 128.55.
Designated root has priority 32768, address c209.6e07.0000
Designated bridge has priority 32768, address c20b.6e07.0002
Designated port id is 128.56, designated path cost 38
Timers: message age 3, forward delay 0, hold 0
Number of transitions to forwarding state: 1
BPDU: sent 0, received 101