POLITECNICO DI MILANO

Faculty of Information Engineering
Degree Master Course in Telecommunication Engineering
Electronic and Information Department

Energy-aware Traffic Engineering for Carrier Grade Ethernet in Metropolitan Area Network

Supervisors: Prof. Antonio Capone
Prof. Brunilde Sansó

Graduation Thesis of:
Daniele Corti
Matr. 734607

Academic Year 2009 – 2010

Contents

List of Figures

List of Tables

List of Acronyms

Abstract

Italiano

La crescita esponenziale del settore ICT avvenuta negli ultimi anni ha comportato un aumento considerevole dei consumi di energia ad esso correlati. Nei prossimi anni si prevede che l'utilizzo di Internet crescerá ulteriormente, facendone uno dei maggiori consumatori di elettricitá. É per questo motivo che si é sviluppato un grande interesse nel limitare il consumo energetico dei sistemi di Information Technology.

Questa tesi affronta il problema del consumo energetico nelle reti di Area Metropolitana (MAN) che impiegano Ethernet a livello Carrier Grade. L'obiettivo é di minimizzare i consumi di switch e link mantenedo parte della rete inattiva in una modalitá di sleep. Questo é reso possible effettuando traffic engineering e usando il protocollo Multiple Spanning Tree per instradare il traffico.

I risultati hanno mostrato come sia possibile diminuire considerevolmente i consumi, pur mantenendo livelli accettabili di QoS.

English

The exponential growth of the ICT sector in the last few years resulted in a considerable increase in energy consumption. In the coming years it is expected that Internet will grow further, making it one of the largest consumers of electricity. That's why there is a lot of interest in minimizing the energy consumption of Information Technology systems.

This thesis addresses the problem of energy consumption in the Metropolitan Area Networks (MAN) using Ethernet at the Carrier Grade. The objective is to minimize the consumption of switches and links while keeping the network in an inactive sleep mode. That has been made possible through traffic engineering and employing Multiple Spanning Tree Protocol to route traffic.

The results showed that it is possible to substantially reduce power consumption while maintaining acceptable levels of QoS.

Acknowledgements

This thesis is probably the arrival point of my academic studies (who knows?).

I would like to use this opportunity to thank all the people who have helped me, in different ways, during the last years.

First of all, I would like to sincerely thank Professors Antonio Capone and Brunilde Sansó who gave me the opportunity to carry out this thesis at the École Polytechnique de Montréal. I am very grateful to them for their constant support, guidance, and advice.

I would like to thank also all the people at GERAD for helping me in the last few months.

I am grateful to my friends Fla, Anna, Do, Carlo, Gio, Marco, Teo, Guglielmo, Mario, Prose, Ele, Moja, Bea, Ross, Covi, Chiara. Each of them knows why, and I am too shy to write it here. It is likely that I have to honor them with a speech lasting more than two sketched words this time.

I would like to thank Marco, Paola, Gabri, Luca, Andrea, Don, Ricky, Richard, Ste, Dario, Matteo, Annoni, my colleagues and friends at the Politecnico di Milano, for their friendship during these years. Without them it would be very hard to get at the end of those five years. I would like also to thank Silvia and Filippo, with whom I shared my stay in Canada.

Last but not least, I thank my mother Donatella and my father Valerio, for always supporting me when I need them and encouraging me in difficult times. I thank my sister Ilaria, my aunt Sabrina, for the valuable suggestions she gave me, my uncles Beppe and Roberto, my aunt Grazia and my grandparents Sandrina, Luigi, and Nando. Thanks to all!

Introduction

Minimizing energy consumption in the area of Information and Communication Technologies (ICT) is becoming an increasingly important issue. The bandwidth requirements of networking applications are doubling every 18 months, as reported in [13] and in the next several years global Internet traffic will likely maintain a similar if not higher growth rate. With the increase of IP traffic, the amount of equipment required to route this traffic must grow. As a consequence, network equipment total power consumption will grow as well. That is why the Internet is rapidly becoming a major consumer of electricity with measurable economic and environmental impact and the efficient use of energy in communications is becoming a key issue for industry, society and government.

According to [9], in 2007 the electrical energy spent by the Telecom Italia's Network was more than 2 billion kWh ($2 \cdot 10^{12}$ Wh), representing nearly 1% of the total National energy demand, second user only to the National Railways. Other studies have estimated that the current Internet consumption represents at least 2% of the total energy consumption of the planet and even more as the considered access rate increases, up to 100 billions kWh (10^{14} Wh) in the US [7, 12, 22].

It is accepted that electrical energy consumption is a significant contributor to greenhouse gasses [8]. This raises the issue of whether the Internet may be constrained not by the speed of routers, but rather by their power consumption. One billion kWh/yr corresponds to 85 million dollars and about 0.75 million tons of CO_2 [20], which makes the Internet responsible for an impact not only economic, but also environmental. That is why the efficient use of energy in communications is becoming a fundamental question: significant energy could be saved by applying energy efficiency concepts to the design of the communication system. Although some remarkable results have already

been obtained in the area of mobile networks and wireless sensor, the problem is still at the beginning in the sector of wired networks. The majority of power is supposed to be consumed by the metro and access networks, because of the large number of elements compared to the core. Therefore, the design and the management of those portion of the networks is crucial.

This essay focuses on the energy consumption optimization problem in telecommunication networks addressed through traffic engineering, based on Multiple Spanning Tree (MST). This problem affects the Carrier Grade Ethernet networks employed in the Metropolitan Area Networks.

The thesis has been developed at the GERAD laboratory at the École Polytechnique de Montréal from March to November 2010.

The chapters of the thesis will examine in greater depth the concepts presented in this introduction. The structure of the chapters is as follows.

Chapter 1 describes the overall structure of the global network and some possible way to limit its energy consumption.

In Chapter 2 Carrier Grade Ethernet technology is described, while after a brief summary on Spanning Tree Protocols, Chapter 3 describes how these can be used to improve the performance of Metro Ethernet.

Chapter 4 describes the optimization model proposed to make traffic engineering, considering the minimization of the energy consumption and the load balancing.

Finally a more experimental phase followed, in which optimizations were carried out on random generated networks to test the validity and accuracy of the model. The results are shown in Chapter 5.

The section named Conclusions summarizes the work and present the final evaluation.

Chapter 1

Limiting the energy consumption of the Internet

In this chapter the structure of the global network hierarchy is described. After providing an estimation on how energy consumption is distributed within the network, some proposals on how to limit consumption are given.

1.1 The network structure and its power consumption

As said in the introduction, global energy saving is a matter of public concern, and now there is a lot of interest in reducing energy consumption of IT systems. However the network design problem involves also decisions on communication protocols, message routing, flow control, capacity assignment and load balancing. Due to the complexity of this problem a hierarchical design approach is usually adopted, so that the network design problem concerns different levels.

The present global network hierarchy can be logically split into three main sections:

- Access Network
- Metro Network
- Network Core

The overall structure of the network is shown in Figure 1.1. The elements at the top of the hierarchy consist of few, large backbone nodes connected in a physical mesh.

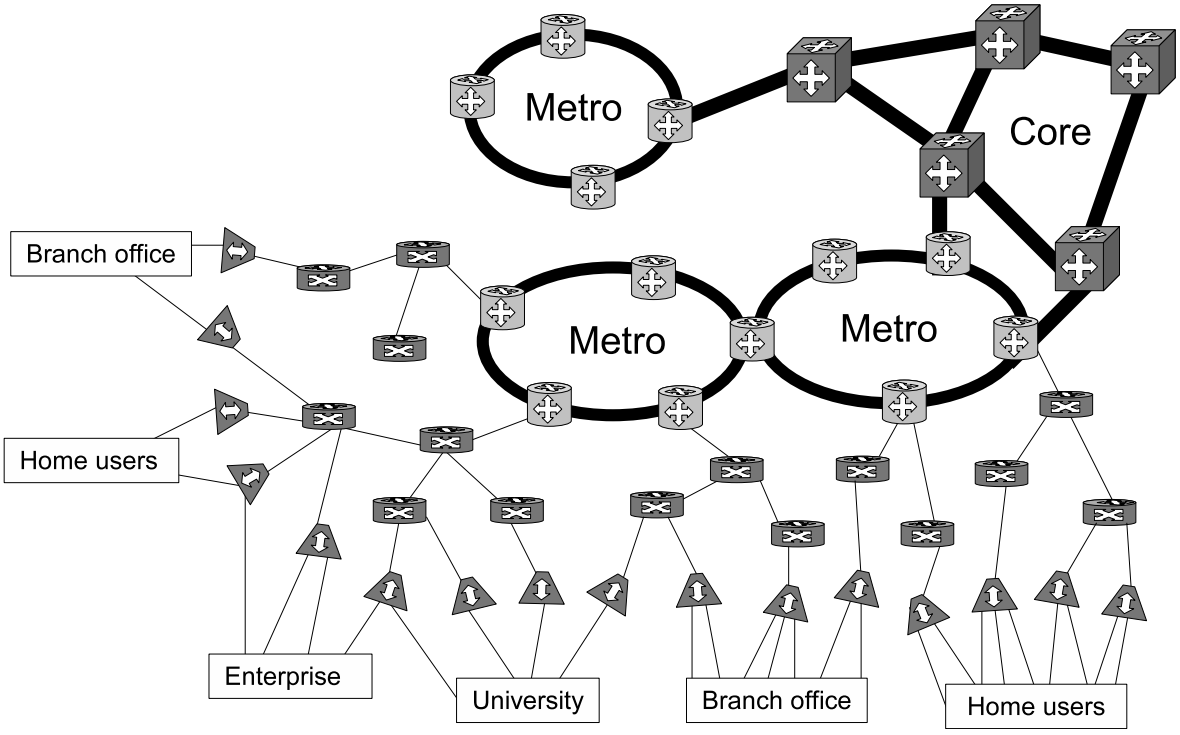


Figure 1.1: Conceptual layout of metro, access and core network.

The backbone routers perform all the necessary routing and also serve as the gateway to neighbouring nodes. This is the core of the network often termed as the long-haul, carrying large volume of communication traffic between different access points. The backbone network has a distributed structure and typically supports multiple paths between nodes, ensuring a high reliability in case of link or node failure.

The next level of hierarchy is represented by the nodes of the metro network, which aggregate the highly fluctuating traffic from the end user and encapsulate the IP packets into a SONET/SDH format for transmission to the network core, serving as interface between the access and the core network.

The exact number of core and edge routers depends heavily on the oversubscription rate used by the Internet Service Provider (ISP) [8]. This rate is equal to $(M - N)/N$, where M is the number of users connected and N is the number that the network can support at a given peak access rate. For example, an oversubscription rate equal to 1 means that only half of the users can simultaneously access the Internet at full speed, or all users at a half of the full speed. In the past, when the Internet was used mainly for website browsing, traffic demand was very bursty. Therefore a user would

load a webpage, read it for a minute or more and then move on to a new webpage. In addition, the number of active users at any time was typically a small fraction of network subscribers. These scenarios allowed ISPs to heavily oversubscribe the network. Nowadays, the development of new multimedia applications such as streaming videos, IPTV, video conferencing and other peer-to-peer applications impose a lower oversubscription rate.

The edge routers together with the links that connect the edge routers to the central node comprise the metro network. This part resembles to a tree: several virtual trees are built together to form a logical mesh, often overlaid on a physical ring. The root of this tree is at one or more of the border core nodes while its leaves terminate at an ISP's Point of Presence (POP).

The access network level represents end users connected to POPs via a multitude of physical media (wireless, optical, DSL). It consists of the curb-side nodes, the network unit at each home that connects the home to the access network, the concentrators and the links that connect these three stages of the network.

To minimize both operational and installation costs required to connect each home to one of the edge nodes, every curb-side node has one or more passive splitters that split a single fibre from an Optical Line Termination (OLT), which resides in an edge node, into several fibres. Each of these fibres connects to an Optical Network Unit (ONU) attached, to each home. Each group of ONUs sharing the same connection to an OLT communicates with it in a time multiplexed order, with the OLT assigning time units to each ONU based on its relative demand. The oversubscription rate can be changed by changing the number of ONUs that share a connection to an OLT. The metro and the access level are the least reliable part of the network.

In [8] a model for the measurements of Internet power consumption is presented. Figure 1.2 shows the total power consumption of the network, versus average access rate when there are 2 million homes and the peak access rate is 100 Mb/s per home.

The average access rate is given by $R(p+1)$ where R is the peak access rate and p is the oversubscription rate. Also included in the plots are the power consumption of the WDM links connecting the edge nodes to the core nodes, the core and the metro plus

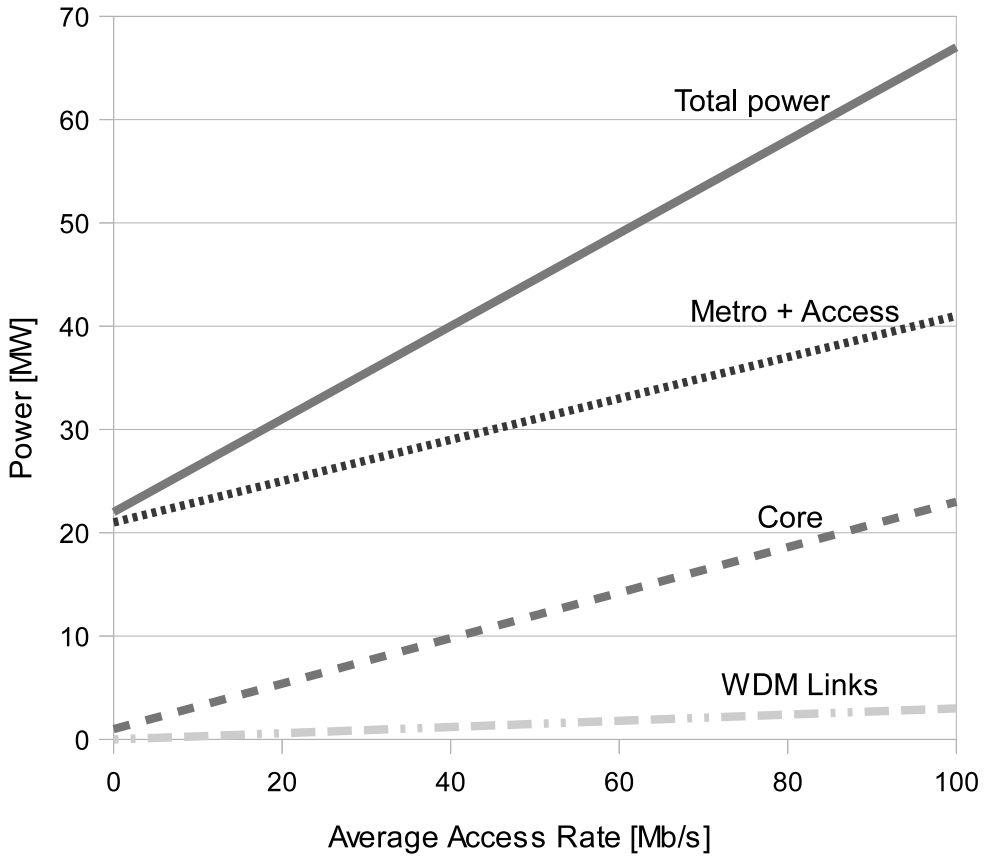


Figure 1.2: Power consumption of WDM links, core nodes, metro plus access network and total power consumption for the full range of average access rates when there are 2 million homes and the peak access rate is 100 Mb/s [8].

access network. Considering this model, the Internet consumes 1% of total electricity consumption. In this scenario the majority of power is consumed by the metro and access networks with the core node only consuming 22% of the power. The WDM links consume only 2% of the power.

Having said this, the network can be designed and managed so as to reduce power consumption. This is the purpose of green networking, which is discussed in the next chapter.

1.2 Green networking

The target of green networking is to efficiently design and manage networks, taking into account their energy consumption.

The Internet is a very energy consuming system in which a large number of devices are working for guaranteeing full-time IP reachability and service continuity. Although there is a lot of research on energy efficiency in mobile networks and wireless sensor networks, only more recently reducing the consumption of wired networks has piqued some interest, because office/home LAN switches, metro and access networks and backbone networks are continuously consuming a lot of electric power.

The study of power-saving network devices has been introduced over these years, starting from the innovative work of [21]. They note that sleeping strategies appear to be the most appropriate way in which is possible to maximize energy conservation in the Internet. The energy saving issue must be handled at many different levels, such as network architecture, network devices, protocols and management algorithms. The indicated problem implies that architectural change may be required for the Internet to enable reduction of power consumption since the major protocols used in the current Internet do not consider for example intermittent operation (i.e. sleep and wake up) of network equipment. Thus, some innovations in protocol and architecture design is required for enabling fully energy-efficient Internet architecture. They proposed uncoordinated and coordinated sleeping mechanisms to reduce the power consumption of routers and discussed impact of the mechanisms on LAN switch and routing protocols.

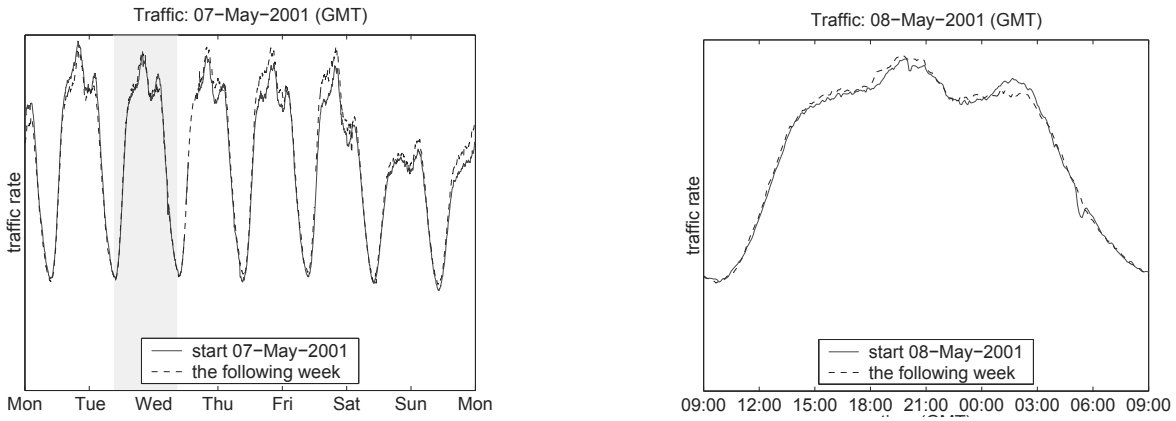


Figure 1.3: Periodicity of network traffic [36].

Following their work, there are several works related to the improvement of the energy efficiency of the Internet. Some of them are listed in the following subsections.

1.2.1 Sleeping techniques

Internet is dimensioned to satisfy peak conditions. On the other hand the traffic level is not constant but varies over time in a periodic and predictable way. In this respect, Figure 1.3 shows the total traffic entering a region of the network at a Point of Presence over two consecutive weeks in May 2001, focusing on daily and weekly variations in the traffic. The unbroken line represent the first weeks data, while the dashed line shows the second weeks data [36]. It's easy to notice how the traffic takes a regular course on a daily basis. This feature can be exploited to reduce the energetic consumptions during low traffic periods, for example applying sleeping strategies to network devices and trying to make link utilization proportional to the energy consumption.

This is what the Energy Efficient Ethernet (EEE) seeks to obtain with its 802.3az task force [6] which is expected to produce a new standard, introducing energy efficiency enhancements to the existing Ethernet. Current Ethernet standards require both transmitters and receivers to operate continuously on a link, thus consuming energy all the time, regardless of the amount of data exchanged. This task force focuses on the possibility of defining two operation modes for transmitters and receivers: active and low power "sleep" mode for idle links intervals in order to obtain energy savings [33]. This saves significant power consumption (close to 90%) compared to leaving

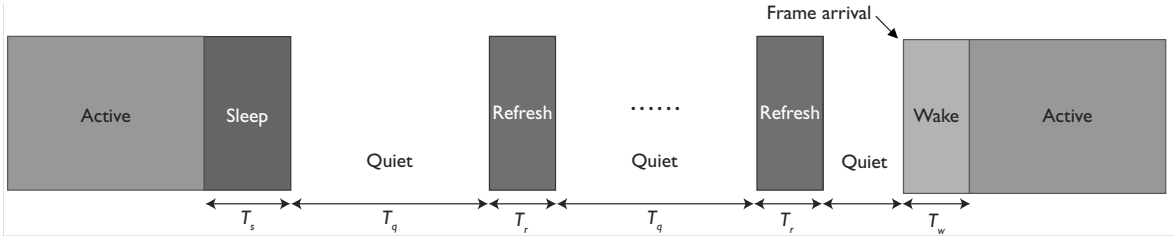


Figure 1.4: State transition example of a link, according to the IEEE 802.3az task force [33].

the devices on full power but with no load. The main idea behind EEE is to put the Physical Layer (PHY) into sleep (low-power) mode when no data is being transmitted and wake it quickly when new data arrives, without changing his speed. The sleep mode freezes the elements in the receivers and wakes them in just few microseconds. Such sleep/active operation requires only minor changes to the hardware.

Figure 1.4 shows a state transition example of a link, according to the Institute of Electrical and Electronic Engineers (IEEE) 802.3az task force. The draft provides for a sleep time T_s (the time needed to enter sleep mode) and a wake-up time T_w (the time required to exit sleep mode). The transceiver spends T_q in the quiet (energy-saving) period. Finally, the draft also considers the scheduling of short periods of activity T_r to refresh the receiver state to ensure that the receiver elements are always aligned with the channel conditions [33].

Concerning energy efficiency of EEE, the sleep-time T_s and the wake-up time T_w are crucial parameters, especially for high-speed links, since if the transmission time is too short, most of the time and energy is spent on waking up and putting to sleep the links rather than on actual data transmission. For the same reason, energy overhead is particularly high for small data frames. This leads the energy consumption of today's EEE standard to be away from proportionality with the system's load.

Higher efficiency levels could be achieved if the device is awoke only when a certain number of data frames are ready for transmission. From an energetic point of view there are benefits collecting data frames and sending them as a single unit of burst over the link (burst transmission [34]) because this way the time the device spends in sleep-mode is maximized and, unlike the previous case, the proportional relationship between energy and load can be reached. On the other hand the additional delay

which is introduced should not be allowed to grow excessively, because it could affect the upper-layer protocols and the applications' performance.

Figure 1.5 shows the energy consumption versus traffic load for three Ethernet speeds: 100 Mbps, 1 Gbps and 10 Gbps. Current standard operate at maximum power all the time, consuming full energy regardless of the traffic load. EEE (called here Frame EEE and plotted with the solid line) allows energy consumption to be more proportional to the traffic load, which should save much energy. Employing the burst transmission (called here Burst EEE) the proportional relationship between energy and load is achieved. In both EEE standards, power consumption in sleep mode is assumed to be 10% of that in active mode.

Implementing energy savings for routers is much more complex, because they often exchange routing information through routing protocols, and thus they cannot simply enter sleep mode even if there is no data traffic. In addition, if a router enters a sleep mode, it makes itself inactive and unresponsive to what happens in the network, threatening to cause the virtual disconnection of the topology. There are many connection paths in the network and a number of those paths could be blocked if some core routers do not process any packets. Coordination is necessary to reorganize some connection paths so that traffic can be aggregated along these paths only, while allowing the network devices on the idle routes to go into sleep mode.

In this regard, [22] proposes a distributed routing protocol that coordinates how routers in wired networks go into power saving mode without degrading quality of service nor network connectivity. During peak hours the protocols will not degrade the network performance, while during non-peak hours some routers can enter sleep mode without degrading the quality of service and adversely affecting network connectivity.

A power saving router has two states: working and sleeping (see Figure 1.6). In the working state, the routing operation is the same as a traditional router, but in the sleeping state the router enters sleep mode so that it will not process any packets until it switches back to the working state.

The state of a router remains unchanged when the network is not busy and it is in the sleeping state, or when the network is busy and it is in the working state. A router

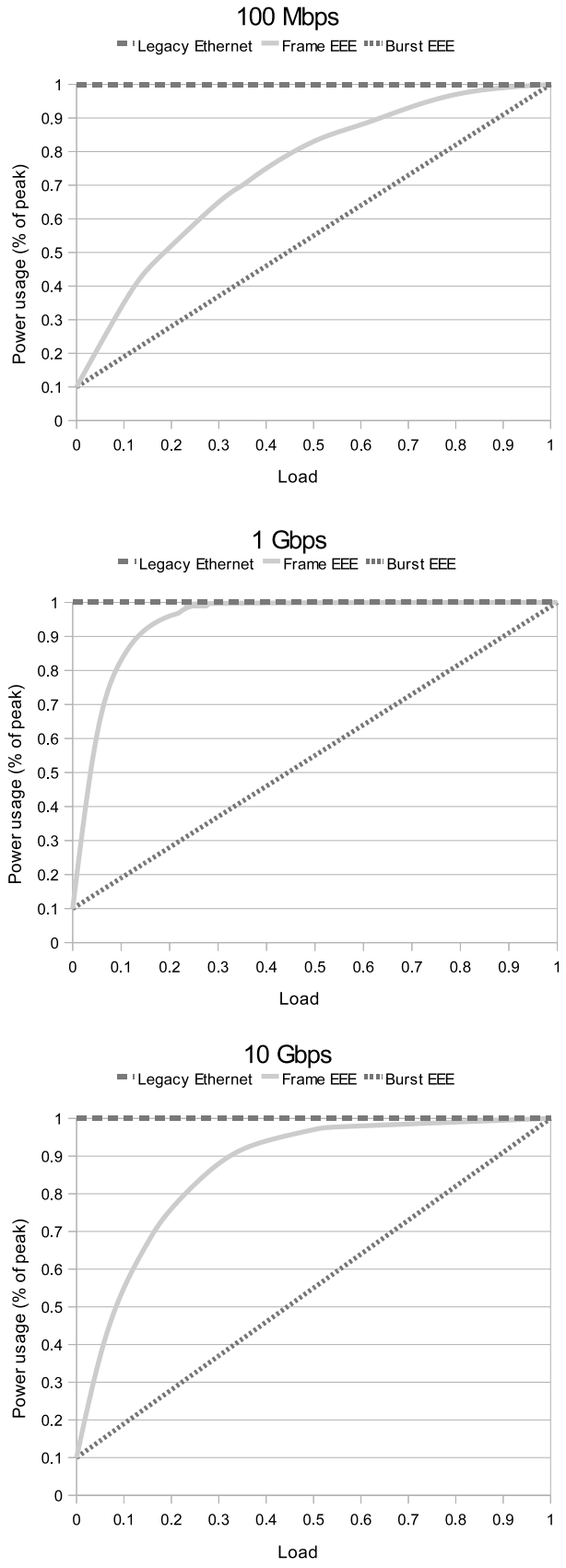


Figure 1.5: Energy consumption versus traffic load for three Ethernet speeds: 100 Mbps, 1 Gbps and 10 Gbps[34].

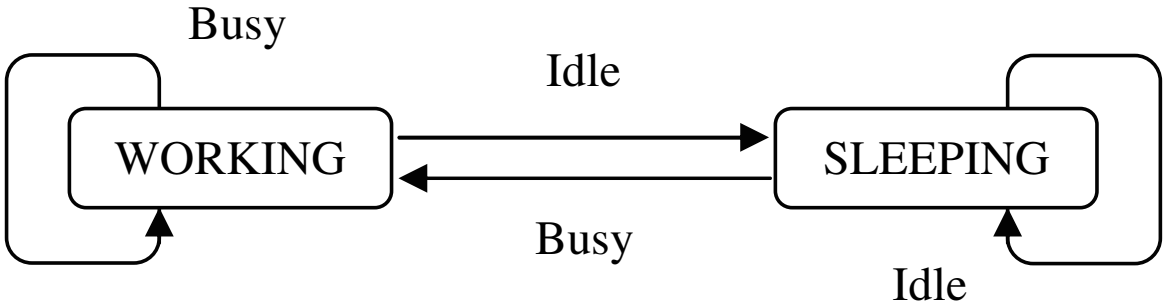


Figure 1.6: Two states and transitions of a power saving router. [22]

detects the network loading by measuring the maximum utilization among all the links attached to itself (U_{max}). This is not the real overall network loading but it's highly related to. When the U_{max} value of a router which is in a working state goes below a threshold, if the network connectivity can be maintained, the router send a message to the coordinator (which was previously elected among all routers) to get a permission to sleep. If the router get a positive reply, then it broadcast a rebuilt routing table and then go to the sleeping state within a certain period. At the end of this sleeping period it will wake up and connect to the network again. Then if the U_{max} value of the coordinator is greater than a threshold, the router will have permission to switch to the working state. Otherwise it will switch back to the sleeping state and sleep in another sleeping period.

1.2.2 Adaptive Link Rate

It is estimated that the Ethernet Network Interface Controller (NIC) consume considerable quantities of energy, up to hundreds of millions of dollars just in the U.S. The energy used by Ethernet links is growing rapidly as the default link data rate increases, even if they are underutilized. In [20] and [19] the authors focus on the opportunity for significant energy savings with user imperceptible impact to packet delay by operating links at a lower data rate during low utilization periods. Energy savings are possible since employing a lower data rate means lower energy consumption: 1 Gbps Ethernet links consume about 4 W more than 100 Mbps links, and a 10 Gbps Ethernet link may consume up to 20 W more. This technique is called Adaptive Link Rate (ALR) and consist in adaptively varying the Ethernet link rate to match the offered load or

utilization, making energy consumption proportional to the link utilization. ALR is intended to use existing Ethernet data rates and is meant to be employed primarily for edge links. However, some energy efficiency enhancements should be made to the existing Ethernet. The key improvements needed to develop an ALR technique are:

1. defining a sufficiently quick mechanism for determining how the link data rate is switched;

2. creating a policy to determine when to switch the link data rate, in order to maximize the energy savings (that is, maximize time spent in a low link data rate) while minimizing increased packet delay. Thus, the performance trade-off is packet delay versus energy savings.

In [20] the authors illustrate one switching mechanism and three different switching policies.

The ALR mechanism must be as quick as possible, in order to avoid great packets delays. The ALR Media Access Control (MAC) Frame handshake mechanism is a fast handshake implemented using Ethernet MAC frames, able to successfully be completed in less than 100 microseconds. Firstly, the link that determines a need to increase or decrease its data rate requests a data rate change using an "ALR REQUEST MAC frame". Secondly, the receiving end link acknowledges the data rate change request with either an "ALR ACK MAC frame" if it agrees to change the data rate or an "ALR NACK MAC frame" if it does not agree. After an "ALR ACK" response the link data rate can be switched and the link resynchronized. Figure 1.7 shows this procedure.

The ALR policy should determine when to switch the link data rate preventing an oscillation between rates. The simplest policy is ALR Dual-Threshold Policy, based on output buffer queue length. Two threshold are introduced: if the output buffer queue length exceed a certain value, then the data rate must be transitioned to high. On the other hand, if the output queue decreases below a threshold, then the data rate can be reduced to low. Better results in therm of limiting the oscillation between rates are obtained through ALR Utilization-Threshold Policy, which uses as threshold values the

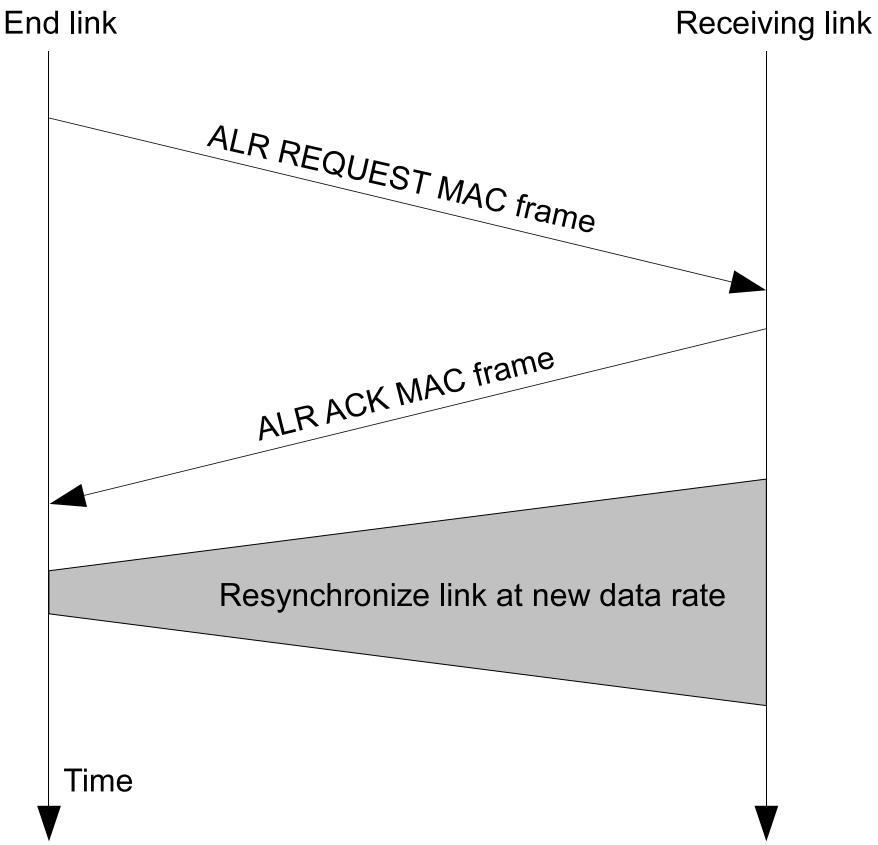


Figure 1.7: Timing diagram for an ALR MAC frame handshake mechanism [20]

utilization of the link instead of the queue length. The last ALR policy is ALR Time-Out-Threshold Policy. It improves the Dual-Threshold Policy introducing timers that force the protocol to maintain the link at a certain speed for a predetermined period of time.

1.2.3 Energy cost-aware routing

Electricity is produced by government utilities and independent power producers from a variety of sources. Different regions may have very different power generation profiles. Energy producers and consumers are connected to an electric grid through complex network of transmission and distribution lines. Since electricity cannot be easily stored, supply and demand must be continuously balanced. To do this in the U.S. an organization determines the price of electricity taking into account demand and supply of energy. For this reason electricity price is subject to continuous oscillations, both temporal and geographic, that make it vary on an hourly basis and are often not well correlated at different locations.

Figure 1.8 shows how the hour of day affects the price differentials for three location pairs. For instances, if two cities are in different time zones, peak demand does not overlap. For Palo Alto-Richmond, there is a strong dependency on the hour. Before 5 am (eastern), Richmond has a significant edge, while after 6 am the situation has reversed. From 1-4 pm neither is better. For Boston-NYC there is a different kind of dependency: from 1 am to 7 am neither site is better, while at all other times Boston has the edge. The effect of hour-of-day on Chicago-Peoria is less clear.

Figure 1.9 shows how the daily average peak prices vary in different regions of the U.S., according to their power generation profiles. Note how the elevation in 2008 caused by the growth of natural gas prices does not affect the hydroelectric dominated north-west region (first plot), while it determines a record peak price of the energy in Texas (third plot), where 86% of the energy is generated using natural gas or coal.

Moreover, these variation are not negligible, as much as a factor of 10 from one hour to the next. That is why the cost of electricity and its availability becomes important in the management of a network.

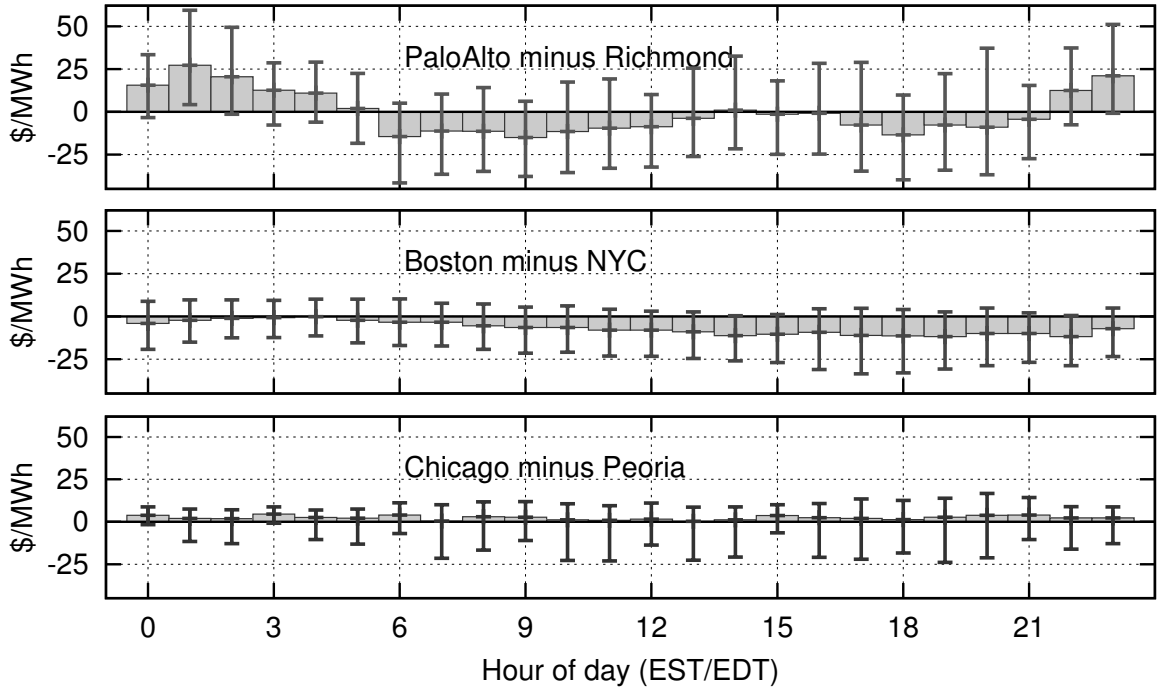


Figure 1.8: Price differential distributions for three location pairs at each hour of the day [32].

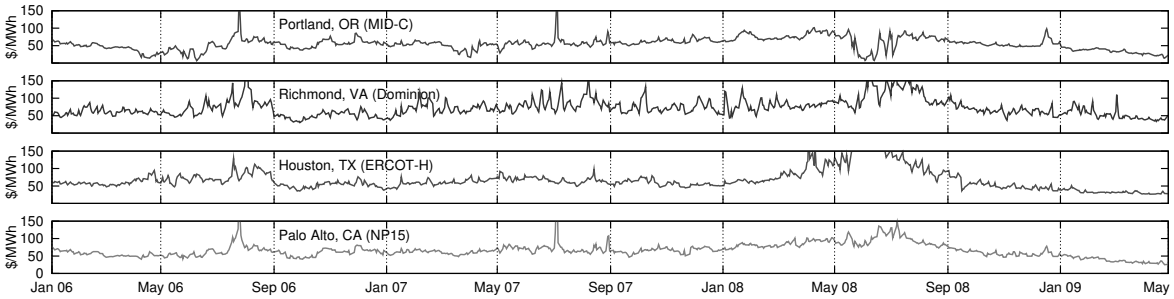


Figure 1.9: Daily average peak prices in different city of the U.S. [32].

In [32] the authors claim that large companies whose system are geographically distributed could save a significant amount of money using a cost-aware request routing policy that preferentially maps client requests to locations where energy is cheaper. However, this technique minimizes the electricity cost of the system but not his energy consumption.

1.2.4 Next Generation Access Network

Even the choice of the network architecture affects the energy demand of the ISP. A new evolution of the network is foreseen in the nearby future, with the deployment of the Next Generation Access Network (NGAN), which is expected to bring about greater energy efficiency than the legacy network, implying new challenges on energy savings and making Internet greener. NGAN will also enable higher Internet speed, achieved through the progressive optical fibre deployment in the following architectures:

- Fiber to the Exchange (FTTE)
- Fiber to the Cabinet (FTTC)
- Fiber to the Building (FTTB)
- Fiber to the Home (FTTH)

Figure 1.10 shows how the copper access network will progressively shorten employing this architectures, boosting the performances of the xDSL systems.

FTTE is the current network architecture employed for the wideband.

In FTTC the fiber reaches the cabinet, which becomes a very important element of the network since it deals with opto-electric conversions. This architecture is quite expensive: due to the size of the cabinet, it could occupy space on sidewalks, which requires obtaining the necessary permits. Moreover, being the roadside, it could be subject to vandalism, which could damage the complex and expensive optical/electric technology that is integrated in the cabinet. FTTC is mainly widespread in the Netherlands and United Kingdom but it is not suitable in other countries such as Italy because of the access network conformation, in which each cabinet should serve too many end users.

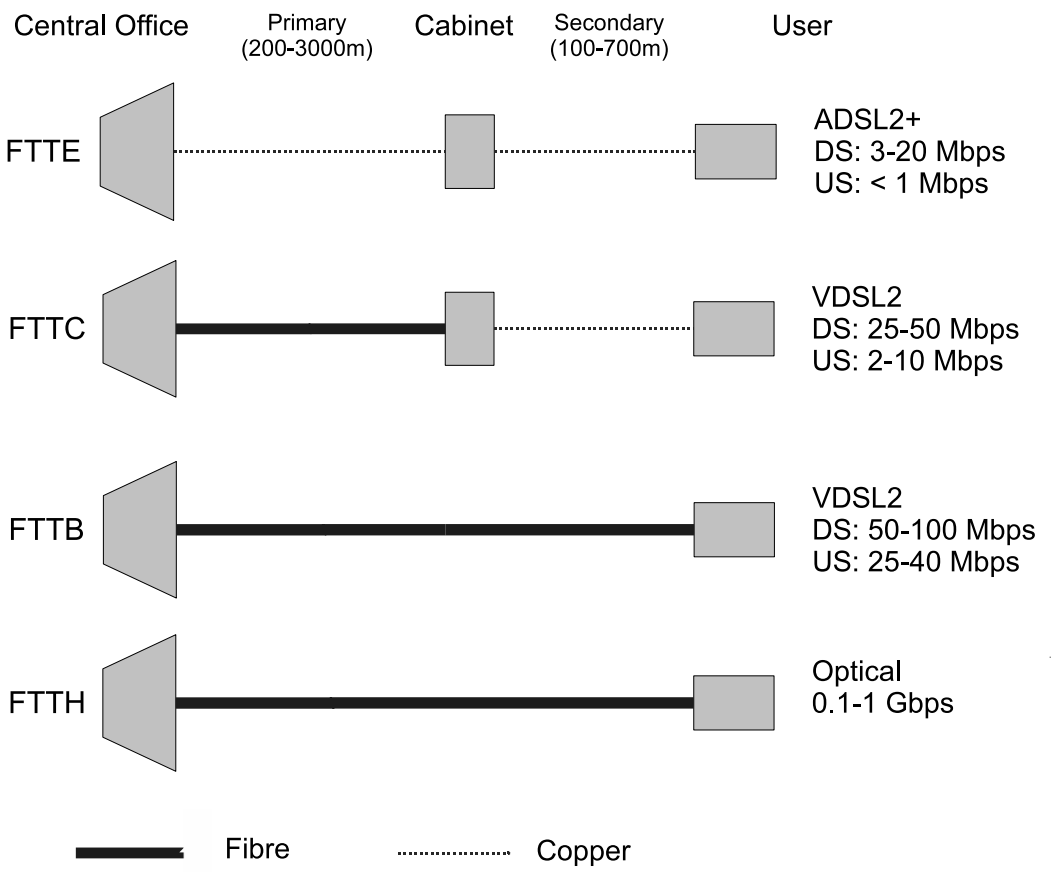


Figure 1.10: Fibre deployment in FTTx architectures.

Switching to the FTTB, there are no more cabinets and the fiber arrives up to the buildings.

The last architecture is the FTTH, in which the fiber comes directly into homes. This is the more expensive solution.

All these architectures employ Ethernet as a Carrier Grade in their Metro network.

Chapter 2

Carrier Grade Ethernet

In this chapter, after a brief description of the evolution of the Ethernet standard, Carrier Grade Ethernet technology is described. The chapter concludes by describing the current deployments of this technology.

2.1 Ethernet standards employment

Simplicity and the high bandwidth available at low cost made Ethernet an attractive choice in various networking deployments. Ethernet is a plug'n'play technology at the link layer developed to provide connectivity in Local Area Networks (LANs). Originally set to 10 Mb/s in the 1980s by the Ethernet physical layer 802.3 standard, its transmission rates have evolved to higher speeds (100 Mb/s and 1 Gb/s), reaching 10 Gb/s with the IEEE 802.3ae standard in 2002 as can be seen in Figure 2.4. In parallel the physical media migrated from coaxial cables to twisted pairs and fiber optic cables.

The majority of these extension were incremental innovations that merely increased the access speed of the LAN. 10 Gigabit Ethernet (10GbE), however, was of a different nature. Its standardization over twisted pair cable (known as 10GBASE-T) required a sophisticated PHY layer and took more than three years to develop. Current 10GbE applications are mostly used in Metropolitan Area Networks (MANs), Wide Area Networks (WANs) and Carrier Networks. 10GbE was the first Ethernet standard to include interoperability with carrier grade transmission system such as Synchronous Optical Network (SONET) and Syncronous Digital Hierarchy (SDH). Within enter-

Speed	Port Name	Reach	Medium	Description
40Gb/s	40GBASE-KR4	1 m	Electrical backplane	4x10 Gb/s
	40GBASE-CR4	7 m	Cu cable	4x10 Gb/s
	40GBASE-SR4	100 m	Multi Mode Fiber	4x10 Gb/s
	40GBASE-LR4	10 km	Single Mode Fiber	4x10 Gb/s
100Gb/s	100GBASE-CR10	7 m	Cu cable	10x10 Gb/s
	100GBASE-SR10	100 m	Multi Mode Fiber	10x10 Gb/s
	100GBASE-LR4	10 km	Single Mode Fiber	4x25 Gb/s
	100GBASE-ER4	40 km	Single Mode Fiber	4x25 Gb/s

Table 2.1: 40GbE and 100GbE physical layer specifications [35].

prises it is used to interconnect servers in data centers. Its use in a LAN environment is limited by the restricted capacity of disk drives to process high data rates. Accordingly, 10GbE can be seen as a platform innovation suitable for two fields of use: high-speed server interconnections and carrier networks.

In regards to this, Ethernet Passive Optical Network (EPON) standard (802.3av), the extension of Ethernet LANs to access environments, also reached 10 Gb/s in 2009, improving another standard published in 2004 of 1 Gb/s speed (802.3ah, known as Ethernet in the First Mile). Thereafter the IEEE 802.3 Higher Speed Study Group (HSSG) has continued his work.

The most recent IEEE draft, 802.3ba was approved in November 2009 and it has been ratified on June 2010. At the beginning the HSSG only targeted a data rate of 100 Gb/s (100GbE) for higher speed Ethernet, but after considering applications for the server and storage, added a data rate of 40 Gb/s (40GbE). Currently, the two data rates are being specified simultaneously. Four types of PHY devices are foresaw for the 100 Gigabit Ethernet (100GbE) and another four for the 40 Gigabit Ethernet (40GbE). Physical layer specifications for 40 and 100 Gigabit Ethernet are defined in Table 2.1.

Focusing on the 100GbE family, the 100GBASE-CR10 PHY supports transmission of 100GbE over 7 m of twin axial copper cable across 10 differential pairs in each direction. The 100GBASE-SR10 PHY is based on Multi Mode Fiber (MMF) optical technology and supports transmission of 100GbE across 10 parallel fibers in each direction. The effective date rate per lane is 10 Gb/s. The 100GBASE-LR4 PHY is based on Dense Wavelength Division Multiplexing (DWDM) technology and supports

transmission of at least 10 km over four wavelengths on a Single Mode Fiber (SMF) in each direction. The effective data rate per lambda is 25 Gb/s. The 100GBASE-ER4 PHY improves the previous PHY specification supporting transmission of at least 40 km.

Gigabit Ethernet real innovation is the development of an architecture that is flexible and scalable, able to support both 40GbE and 100GbE and that could scale to future Ethernet speeds. A unique aspect of the 40Gbe/100Gbe PHY is the adoption of a multi-lane transmission, which means that the 100 Gb/s speed can be achieved for example by using four lanes of 25.78125 Gb/s per lane. In this way the signal rate needed in each lane at the interface can be slower than the total data rate of 100 Gb/s. What is unique about 100GbE is that the number of lanes is almost freely selectable. However, the number of lanes should be as few as possible by using 25 Gb/s Wavelength Division Multiplexing (WDM) transmission in long-distance applications (10 km, 40 km), whereas for short-distances (up to 100 m) transmission, it is possible to have many bundled 10 Gb/s serial interfaces. According to this, the number of physical lanes can be varied to suit the type of Table 2.1.

Starting from 2007, many field trials [40] have been accomplished, demonstrating that 100GbE channels can also be overlaid onto the existing in-service infrastructure. As a result of a tremendous industrial investment 100GbE is able to match, if not exceed, the performance of the traditional 10GbE but with 10 times the capacity for each fiber. The development of 100GbE is justified by the fact that the bandwidth requirements of networking applications are doubling every 18 months, as reported in [13]. In the next several years global Internet traffic will likely maintain a similar if not higher growth rate, which means that networking applications will need Terabit Ethernet in 2015. Figure 2.1 shows global IP traffic predictions up to 2012.

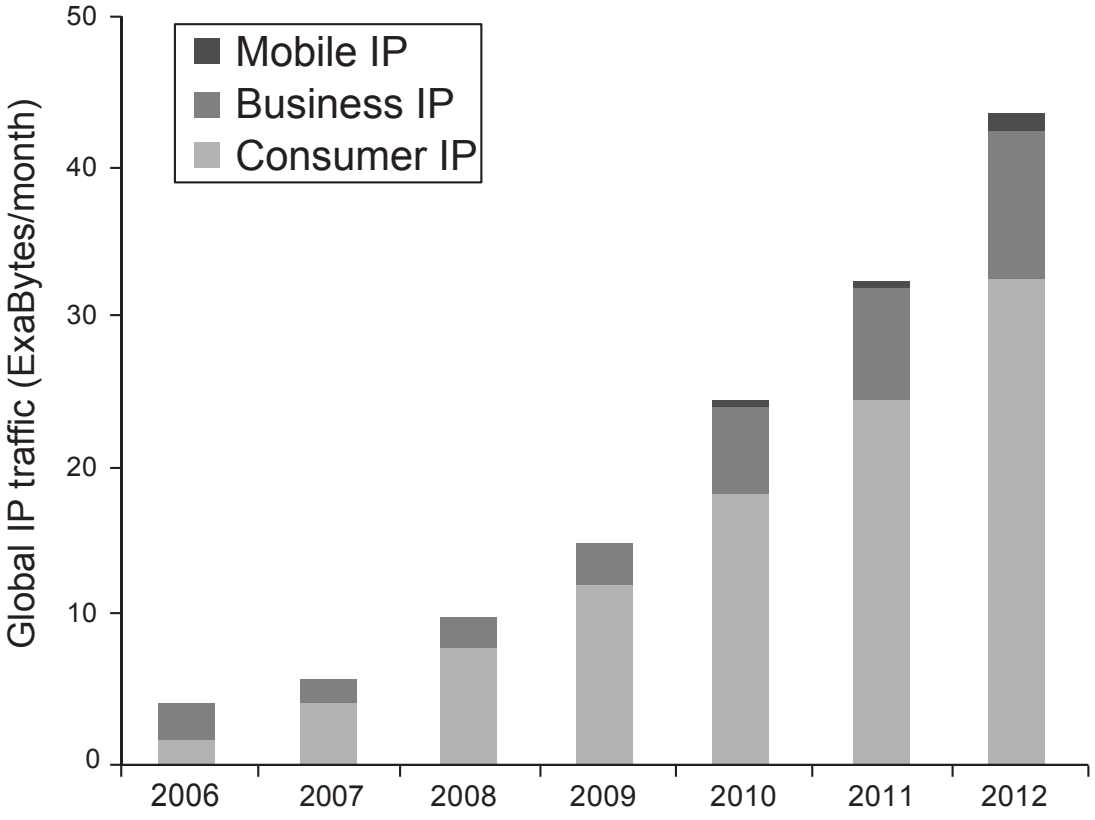


Figure 2.1: Global IP traffic growth according to [40].

Another motivation for transitioning to a higher speed networking interface is to maintain or increase system density. There are many parameters which limit the overall system density including:

- Power dissipation
- Module height, width and length
- Switching capacity
- Electrical interface.

The introduction of 100GbE, thanks to its technical feasibility and its broad market potential, will drive the development of the next generation of electrical and optical technologies, reducing cost and power per 100GbE port and maximizing the usable port densities per system. At this time IEEE 802.3 has also formed a new study group that will examine the need for the development of a 40Gb/s Ethernet optimized for client applications in the carrier environment.

Ethernet technology has so far been widely accepted in enterprise deployments and millions of Ethernet ports have already been deployed. The introduction of higher speed 100 Gb/s transmission will not only increase the bandwidth of the interfaces and the throughput of the links, but it will also significantly impact the entire network. The advancement of Ethernet technology is pushing Ethernet from local area network environment to metropolitan area network environments. 100GbE over Optical Transport Network (OTN) will be used in the carrier network aggregation and transport areas, providing high quality broadband links with high granularity for the collected data flows. In the near future, high-speed Ethernet will become the dominant client of OTN, while SONET/SDH will gradually vanish. In this regard, OTN should be extended to better support Ethernet: this activity is called extended OTN. International Telecommunication Union - Telecommunication Standardization Bureau (ITU-T) is currently standardizing the next generation OTN interface, Optical Transport data Unit 4 (OTU4), that can accommodate and transport 100 Gb/s-class signals. Figure 2.2 show the evolution of high-capacity OTN and the next generation Ethernet.

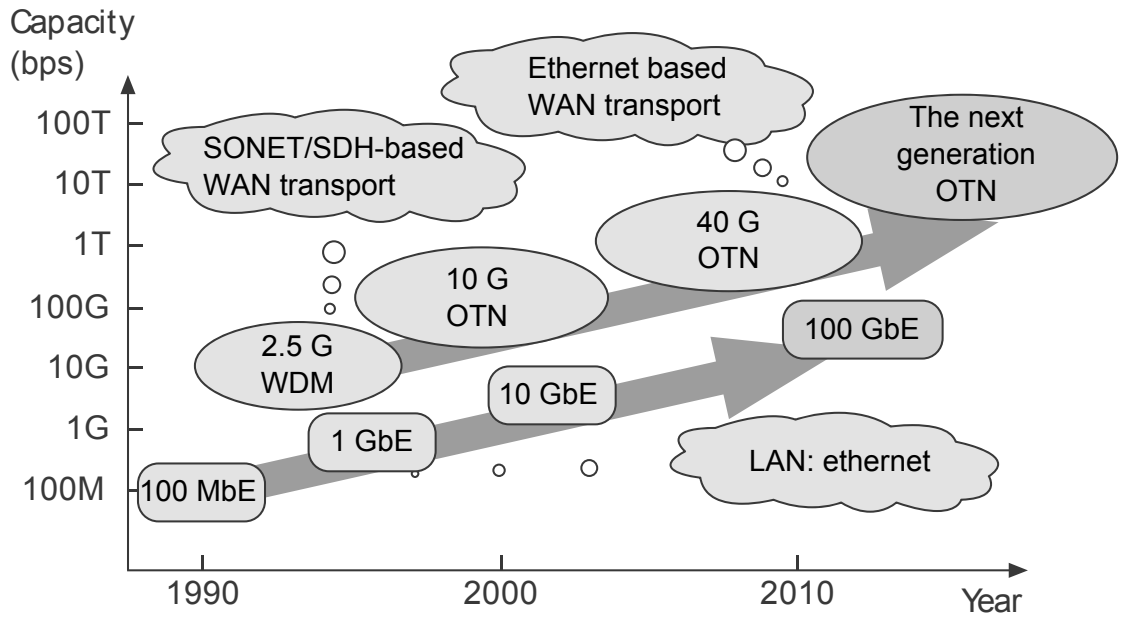


Figure 2.2: Evolution of high-capacity OTN and the next generation Ethernet

However, Metro Networks have different requirements from the early LANs. As such, Ethernet did not have the characteristics required in Carrier Grade (CG) networks and needs to be enhanced if it has to meet these requirements. Among the requirements, traditional Ethernet lacks essential transport features such as wide area scalability, resilience and fast recovery from network failures, Operation, Administration and Maintenance (OAM) capabilities, Admission Control (AC) and advanced traffic engineering functionality, such as load balancing. Consequently traditional Ethernet falls short of providing a good Quality of Service (QoS) as well as security-guarantee levels required by typical transport-network service level agreements Service Level Agreements (SLAs).

As a matter of fact there is a clear convergence to a multi-service network offering more than best effort Internet access, but also voice and video services over the same infrastructure. These services require QoS guarantees and high availability, besides the increased bandwidth.

Nevertheless in the Metropolitan area is expected that Ethernet will play an important role in the near future through the implementation of Carrier Grade Ethernet (CGE). Carrier Grade Ethernet refers to a number of industrial and academic initia-

tives that aim to equip Ethernet with the transport features listed above and that it is missing. In doing so, CGE aspire to extend the all-Ethernet domain beyond the first mile and well into the metropolitan networks. The reasons why Ethernet is considered to be such an attractive technology are driven by the promise of a reduced protocol stack. Ethernet provides very high speeds with copper or low-cost optical interfaces and its management is very simple, so the reductions in cost and complexity are expected to be considerable. Currently installed metro and wide area networks are dominated largely by SONET/SDH and Multi Protocol Label Switching (MPLS). Notwithstanding, most data traffic originates from and terminates at Ethernet LANs. This means that Ethernet is indeed the predominant interface on the endpoints that need to be connected. In addition, most applications and services such as video and business services are migrating toward Ethernet platforms. Some of the problems with the legacy services are the long wait for the service to be installed and activated and the coarse bandwidth granularity. What the enterprises need is rapid provision of on demand services. The enterprises end up paying for the peak bandwidth all the time, even if there is a low bandwidth need in the evening or the weekend, because the service installation is not on demand. In contrast, Ethernet interface can support a variety of bandwidths. Therefore, bandwidth on demand can be easily provisioned. In [24], the authors show that implementing Carrier Grade Ethernet could result in 40% port-count reduction and 20–80% Capital Expenditure (CAPEX) drops compared to various non-Ethernet backbone technology alternatives, providing huge cost advantages to service providers. Therefore, the past few years have seen significant innovations around Ethernet standards in order to meet recent requirements. The IEEE, ITU-T, Internet Engineering Task Force (IETF) and Metro Ethernet Forum (MEF) are currently working on a set of standards, recommendations and technical specifications aimed at revolutionizing the Ethernet from a best effort, plug-and-play LAN technology towards a SLA driven, Carrier Grade WAN technology able to solve the hierarchy and scalability problems of backbone networks. Most of these new standards rely on Spanning Tree Protocols.

2.2 Carrier Grade Ethernet definition

From a carrier point of view Ethernet can be considered to include:

- Ethernet interfaces
- Ethernet services
- Ethernet transport

An Ethernet interface refers to the physical layer media and transceivers used to interface to Ethernet. Ethernet services, on the other hand, are packet based telecommunication services that offer an Ethernet interface (or User-Network Interface) to the customer and ensure reliable delivery of Ethernet packet data. MEF has provided definitions of Carrier Ethernet services, such as E-Line, E-LAN and E-Tree. However, Ethernet services do not necessarily have to be delivered using Ethernet transport. In fact, most implementations of Ethernet services actually use Ethernet over SONET/SDH or Ethernet over MPLS due to their superior Carrier Grade characteristics and reuse of existing infrastructure. Ethernet as a transport technology is something different and is not a prerequisite for delivery of Ethernet services. It has lacked, up to now, the features that carriers require for a wide-scale deployment.

As shown in [18] Ethernet must evolve to achieve the same properties of current Wide Area Network (WAN) technologies. The MEF has defined this evolution as Carrier Ethernet. According to MEF, a Carrier Ethernet Service must posses the following attributes:

- Standardized Services
- Scalability
- Reliability
- Quality of Service
- Service Management
- Backward Compatibility

In its Ethernet service definitions, the MEF has currently defined three standardized service types for the delivery of Carrier Ethernet as shown in Figure 2.3: Point-to-Point (PtP), Multipoint-to-Multipoint (MPtMP) and Point-to-Multipoint (PtMP). These services are defined from the perspective of the User-Network Interface (UNI) i.e., the demarcation point between a provider's Ethernet network and a customer's private network.

An E-Line is a PtP Ethernet Virtual Connection (EVC) connecting two User-Network Interfaces (UNIs) in an Ethernet network. An E-LAN is a MPtMP service offering full transparency to customer control protocols and VLANs (transparent LAN service). Similar to EPON, the E-Tree service offers PtMP connectivity from a root UNI to the leaf UNIs and Multipoint-to-Point (MPtP) connectivity from the leaves to the root. The MEF specifications further associate several service attributes to UNIs and EVCs. These attributes allow a provider to customize the service to end-user expectations and build resulting Service Level Agreements (SLAs). Example of service attributes are physical speed offered at the UNI, bandwidth profile per Class of Service (CoS) identifier, bandwidth profile at each UNI using parameters such as Committed Information Rate (CIR), Committed Burst Size (CBS), Excess Information Rate (EIR), Excess Burst Size (EBS). This approach to offering standardized services requires no changes to customer LAN equipment or networks. By defining a choice of granular bandwidth and QoS options, providers have the freedom to define service that can be customized to the end-user's needs.

Ethernet's scalability is seen as vital to meeting the annual capacity growth demand. Ethernet interfaces are already available in a wide range of speeds and intermediate speeds between standard Ethernet physical rates can be achieved using Link Aggregation (LA). An appealing attribute of Carrier Ethernet is the ability for a user to easily upgrade the bandwidth of an existing Ethernet service via software control. Scalability also allows users to use a network service that is ideal for the widest variety of applications. This means that Carrier Ethernet provides the necessary Quality of Service (QoS) guarantees required to be delivered over a common infrastructure. A

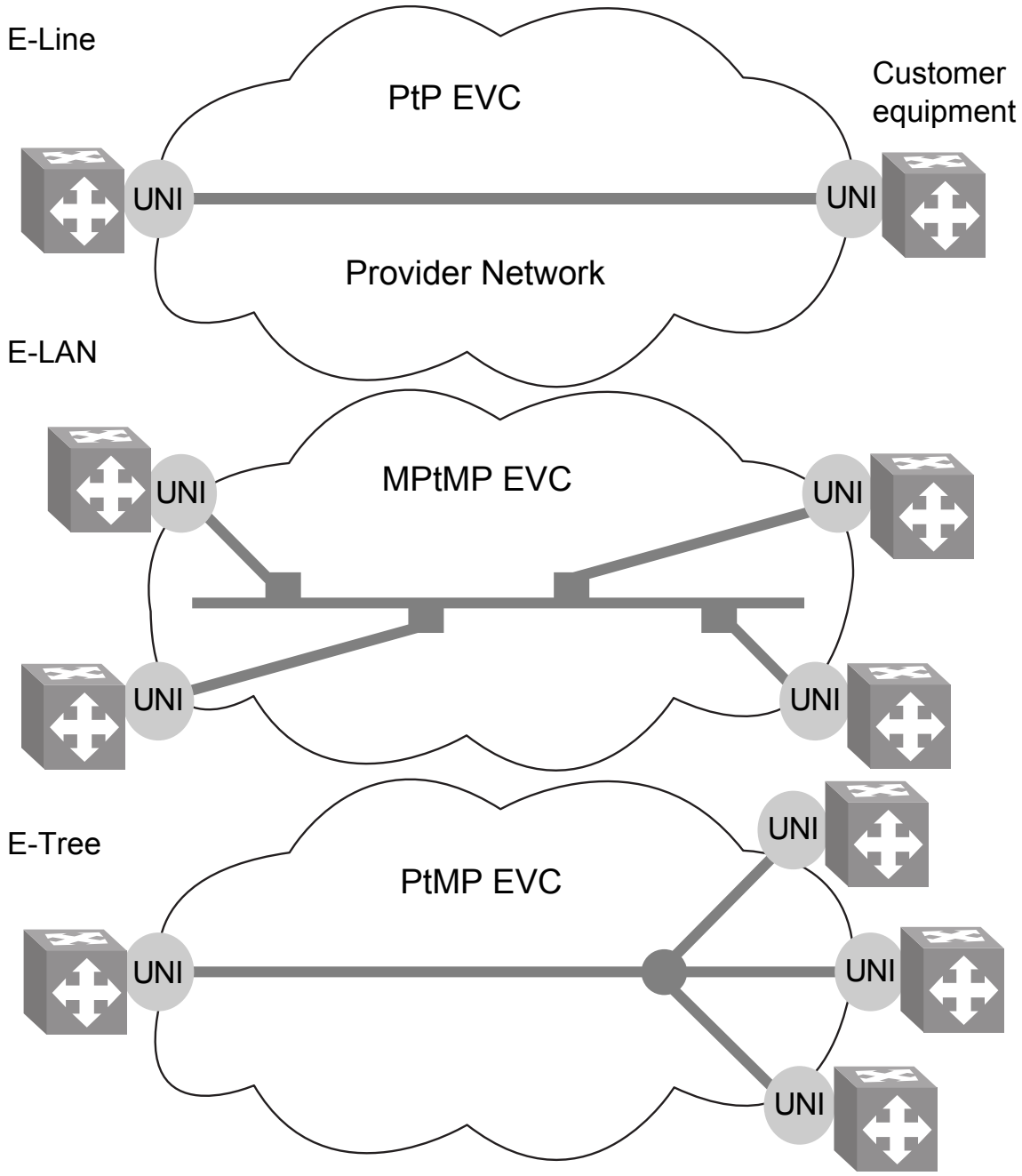


Figure 2.3: Metro Ethernet Forum service definitions: E-Line, E-LAN and E-Tree

third dimension of scalability is geographical reach. It is important to reach a consistent end to end service.

Traditional LAN-based Ethernet was often perceived as a best-effort connectivity mechanism. Carrier Ethernet in contrast must offer the capability to rapidly detect and recover from node, link or service failures. Service providers typically boast ”five 9s”, or 99.999% network availability. One of the benchmark tools for achieving this has been SONET/SDH's ability to provide 50 ms link recovery, as well as protection mechanism for nodal and end-to-end path failures. In order to be adopted, Carrier Ethernet must match these performance levels seen by traditional WAN technologies.

As mentioned before, Carrier Ethernet offers a wide range of granular bandwidth and QoS options. By defining attributes that are associated with the service, advanced Service Level Agreements (SLAs) can be offered to deliver the performance required for a target application.

The fifth critical service attribute that distinguishes Carrier Ethernet is the ability to monitor, diagnose and centrally manage the network, using standards-based tools. In order to do such advanced service management, tools are required to rapidly provision services, diagnose connectivity problems, diagnose faults in the network and measure the performance characteristics of a service.

The last Carrier Grade's attribute is Time-Division Multiplexing (TDM) support. While service providers see substantial growth potential in Ethernet services, existing leased lines are still a significant income source for them. So they must be able to interwork with existing leased lines services as they migrate to a Carrier Ethernet network.

Equipment vendors are challenged with how to add these CG functionalities to Ethernet equipment without losing the cost effectiveness and simplicity that has made it attractive.

2.3 Carrier Grade Ethernet roadmap

The lower time line in Figure 2.4 summarizes the evolution of Ethernet data-link layer. The basic technology standard used for delivering Ethernet service is the IEEE

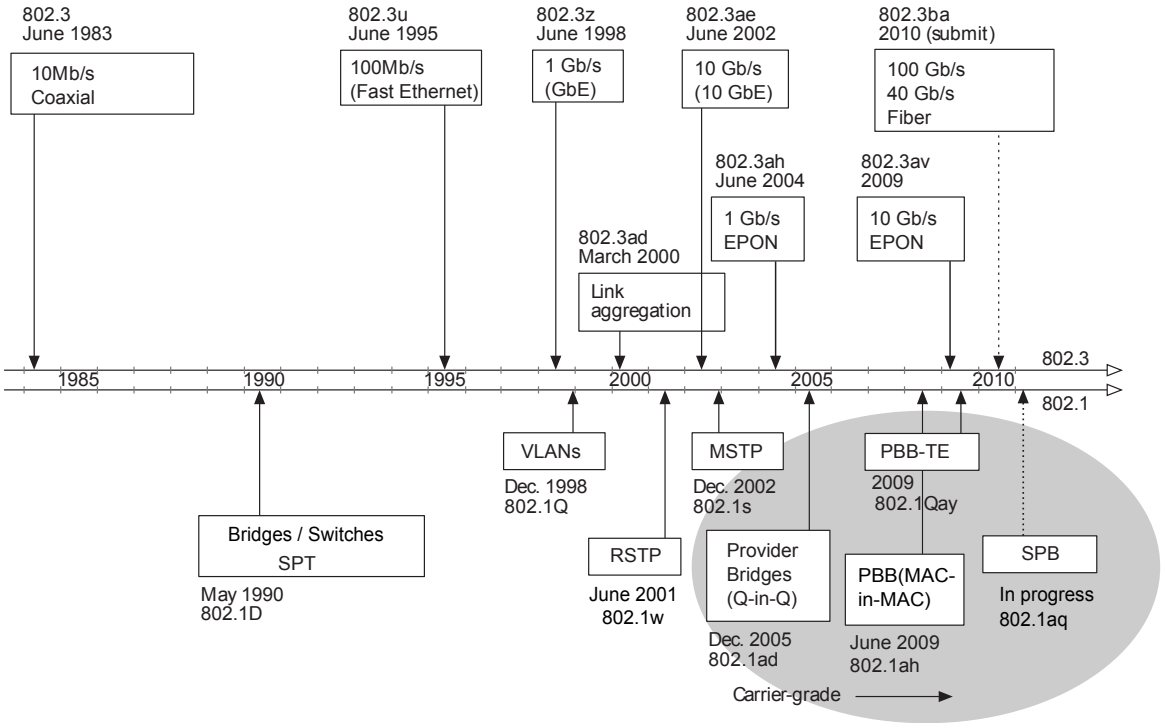


Figure 2.4: Ethernet standards' time line evolution [18].

802.1Q standard. This standard creates Virtual LANs (VLANs) across a common LAN infrastructure and allows enterprises to separate traffic from different departments within a company. Each VLAN identify a logical partitioning of the network and is identified by a 12-bit VLAN ID (VID) (C-Tag). This Customer VID (C-VID) 12 bits limited the number of supported costumers to a maximum of 4094. Although this is sufficient for an enterprise's LANs, it runs into scalability issues when used by a service provider to differentiate customer networks, being not suitable to support Ethernet services in a metropolitan area. In addition, the customers required the same VID field to partition and manage their own networks, leading to a further reduction of the range of VIDs available to the service providers. What is needed is a method for defining secure Ethernet services to individual customers within which each costumer can manage further LANs. There are two standards that support this approach: IEEE 802.1ad Provider Bridges (PB) approved in 2005 (also known as Q-in-Q) and IEEE 802.1ah Provider Backbone Bridges (PBB) which became a standard in 2008 (also known as MAC-in-MAC). The evolution of the Ethernet frame for these standards is presented in Figure 2.5.

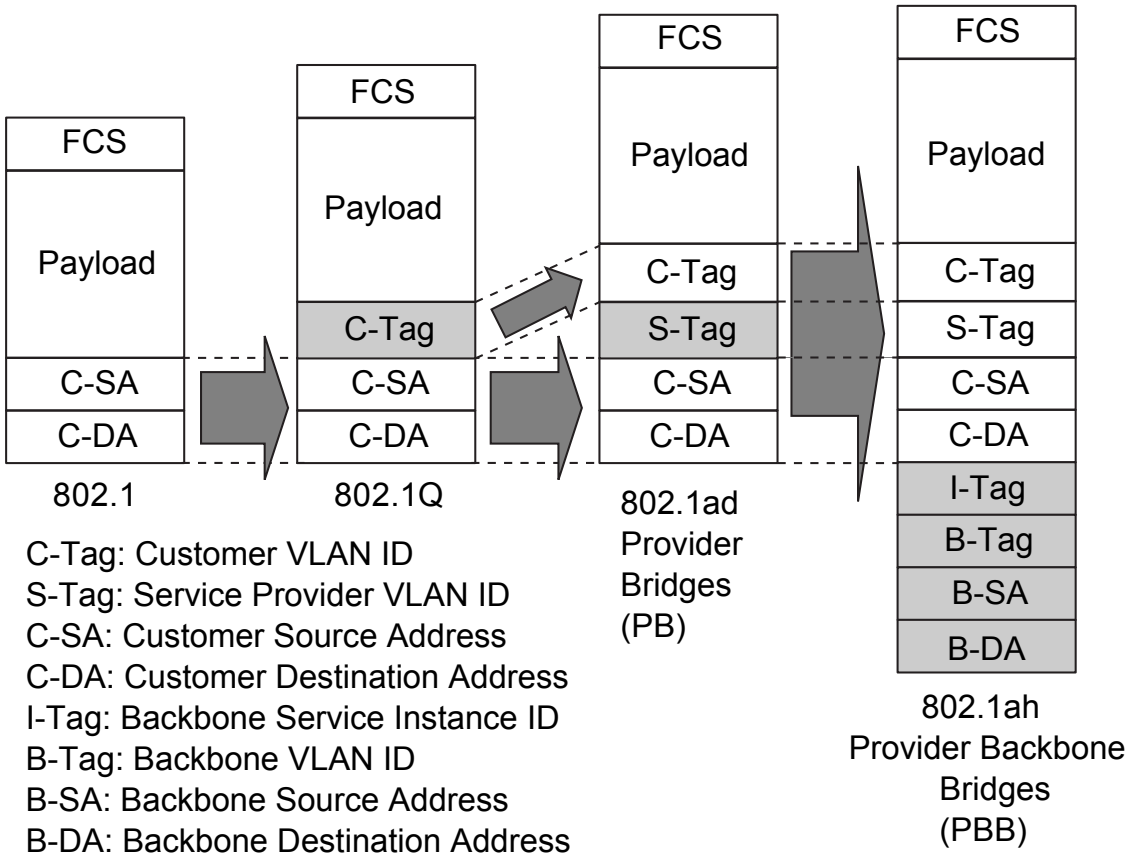


Figure 2.5: Ethernet frame evolution [18].

Trying to mitigate the provider scalability problems, Provider Bridges (PB) simply add an additional VLAN ID (S-Tag). This Service Provider VID (S-VID) is used by the providers to identify the service of a customer network, while the C-VID remains intact and can be used to define different groups of users within a customer's network in a transparent way.

This is often referred to as Q-in-Q. In the Q-in-Q scenario each service requires a different S-VID and because it consists of 12-bit tag, PBs have the same scalability limitation as IEEE 802.1Q (only 4094 services instances can be created). In addition, the provider's and customers' MAC addresses are visible to all network elements making both networks appear as one large network to provider's switches. There's no separation between customer and provider network and any changes to the first one will have an impact on the provider core. For example when a failure occurs in the customer network, the resulting action taken by the Spanning Tree Protocol (STP) can impact the provider network. Moreover, PB cannot provide differentiation between customer and provider Ethernet control protocols which are identified only by their destination MAC address. This causes unpredictable network behaviour and can be solved upgrading the existing Ethernet switches. For this reason, Provider Bridges technology has significant limitations.

Provider Backbone Bridges (PBB) (IEEE 802.1ah) add to the Ethernet frame a MAC header dedicated to the service provider consisting in a backbone source and destination MAC address, a Backbone VID (B-VID) and a Backbone Service Instance ID (I-SID). Essentially, PBB employs an additional Service Provider 16 bit MAC address and basically encapsulates the end user's (this is also referred to as MAC-in-MAC). The outer MAC address is used to forward the Ethernet frames across the Service Provider network. One of the main benefits of PBB is that the 24-bit I-SID allow to enumerate up to 16 million services in the PBB network, completely removing the scalability problems.

Although PBB creates a provider infrastructure that is transparent to the customer networks, they still use best-effort techniques taken from the LAN. Such techniques do not meet the configuration requirements of Carrier Grade operation but,

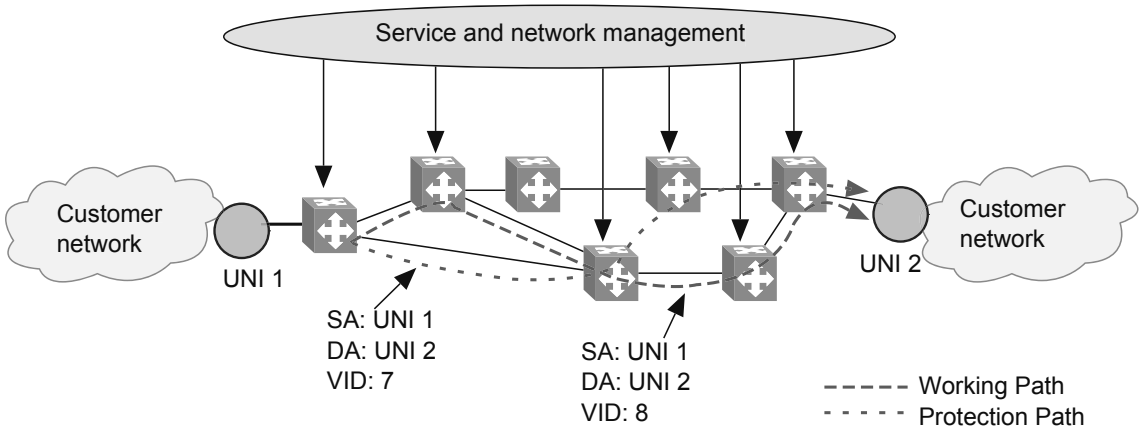


Figure 2.6: Connection oriented path in PBB-TE

due to the modularity of Ethernet specifications, they can be turned off. Following this idea in 2009 the IEEE 802.1Qay standard has been approved. This new technology called Provider Backbone Bridges Traffic Engineering (PBB-TE) provides a connection-oriented forwarding mode in the current Ethernet network without introducing complex and expensive network technologies. It is essentially a variantion of the PBB standard that do not support the Spanning Tree Protocol. In the example in Figure 2.6, two bidirectional Ethernet connections have been configured across the provider network to create a working and a protection path. Each route can be identify by his VID.

Recently, a new class of shortest path routing solutions have been introduced namely, Shortest Path Bridging (IEEE 802.1aq). Shortest Path Bridging (SPB) implements frame forwarding on the shortest path between any two bridges of an Ethernet network. In order to achieve shortest path forwarding, each bridge maintains its own Shortest Path Tree (SPT). SPB form an SPT Region and edge bridges of the Region forward the frames that are incoming to the Region on their own tree. An example is shown in Figure 2.7. In this standard Multiple Spanning Tree Protocol (MSTP) can be applied for the control of SPB, showing that it is still possible to use Spanning Trees in Carrier Ethernets provided that the tree generation and VLAN-Spanning-Tree mapping are performed adequately.

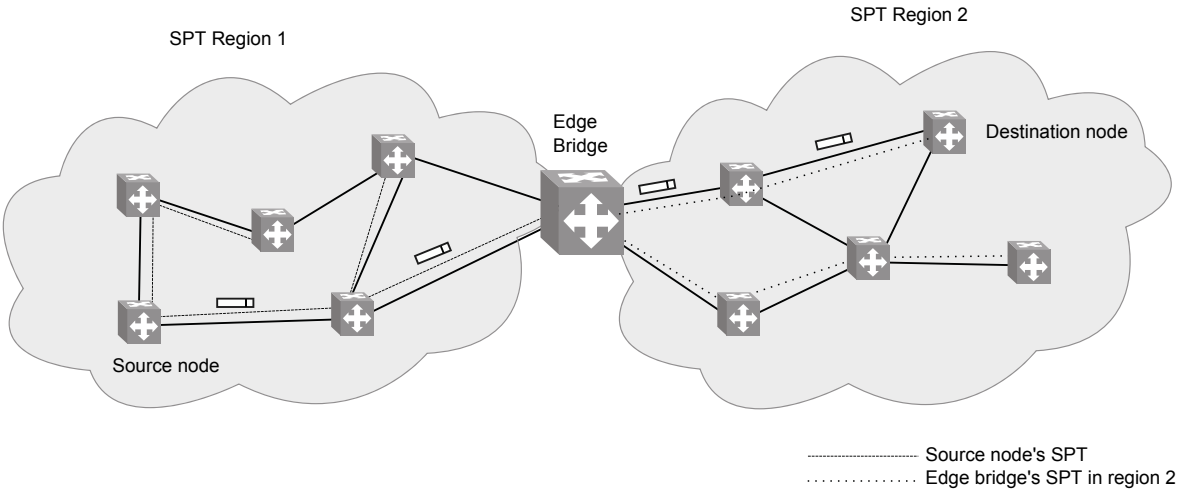


Figure 2.7: Shortest Path Bridging implements frame forwarding on the shortest path.

2.4 Current deployments

At the present time, service providers tend to offer Ethernet over some optical network in Metropolitan Ethernet Network (MEN) so that the services can enjoy the de-facto sub-50ms recovery time. The common trend is to have Ethernet deployed at the access and aggregation part of the metro area network. MPLS is the technology for the metro core. The customer premise equipment can be an Ethernet switch or a router. There are several deployments of Metro Ethernet Network around the globe supporting a wide range of applications. Here are some examples [23].

AT& T installs a multipoint Virtual Private Network (VPN) E-LAN providing high-speed connection among 12 locations of the Clarian Health Center in Indiana, USA [29]. The services include voice and data interconnecting their healthcare systems providing accesses to up-to-date medical research, clinical expertise, and patient values to reduce the variance of care between physicians giving the same patient and physical condition.

The city of Roanoke, Virginia, USA, deploys Ethernet over a SONET network between ten sites. Their application is a web portal that allows citizens to pay parking tickets and tax online, obtain registrations and permits, view parks recreation information, and satellite images of properties for real estates and development purposes.

A utility company in Idaho, USA, the Idaho Falls Powers (IFP), creates a communication network of their own after realizing the flexibility and scalability of metro Ethernet technology avoiding the constrained of the incumbent carriers for new leased

services [28]. It can provide the transportation of multiple services with various CoS. To its customer, IFP offers transport of mission critical data, Ethernet private line, multiple services such as VoIP, and video surveillance.

IFP leases wavelength instead of fiber strands with resiliency and automatic protection for the fiber. A Korean-based wholesale provider of metro Ethernet, PowerComm, offers Ethernet services to residents in Seoul and integrating with their existing cable network at the same time. PowerComm wants to provide Gigabit Ethernet with the reliability as the traditional SONET/SDH but at a lower cost. It requires one metro backbone infrastructure to unify its existing cable network with the new last-mile Ethernet services. PowerComm also implements layer 2 MPLS E-line services and VPLS E-LAN services.

In Spain, a company, called Al-Pi, saw the potential of Metro Ethernet in 2001 and decided to deploy Gigabit Ethernet services in Barcelona. At the start, it offers LAN to LAN services using the enterprises class Ethernet switches. These switches lack features such as sub-50ms resilience, SLA control, high scalability, and end-to-end QoS guarantee. Then it makes a move toward optical Metro Ethernet using dense fibre network. Now it can offer carrier class services to its customers.

Finally, according to [10], operators will introduce soon Ethernet into their mobile transport infrastructure structure. As bandwidth demand increases to support new users and mobile applications, investments in new mobile backhaul infrastructure have to be carried out. Thus, mobile transport networks are being enhanced to provide Ethernet connectivity services according to Metro Ethernet Forum (MEF), leading to an evolution toward Ethernet and IP-based backhaul solutions. In regard to this, carrier Ethernet technologies provide a broad framework of flexible services that can be tailored to meet the requirements of mobile backhaul.

Chapter 3

Traffic Engineering in Metro Ethernet through Spanning Trees

This chapter describes how the protocols based on Spanning Trees have evolved and lists some proposals on how to use them to improve the performance of Metro Ethernet.

3.1 Spanning Tree Protocol Standards

Current Ethernet switching equipment are compliant with protocols such as IEEE 802.1Q and IEEE 802.1D [3]. The IEEE 802.1D Spanning Tree Protocol, which is proposed in its initial version in 1990, is a routing protocol responsible for building a loop-free logical topology over the physical one using a shortest path approach and ensuring connectivity among all nodes, as shown in Figure 3.1. The logical topology is constructed as follows. First, a root bridge is elected (usually the one with the smallest bridge ID is chosen). Then every other switch selects the port which has the smallest distance from the root as root port. Each bridge can determine the shortest path to the Root Bridge thanks to the information propagated in Bridge Protocol Data Units (BPDU). These packets are exchanged only between adjacent bridges and protocol events are invoked by timers. Finally all bridges select their designated ports: these ports provide the fastest access to a new segment of the network. The other ports, which are neither root nor designated ones, will be blocked ports. The result of this procedure is a tree.

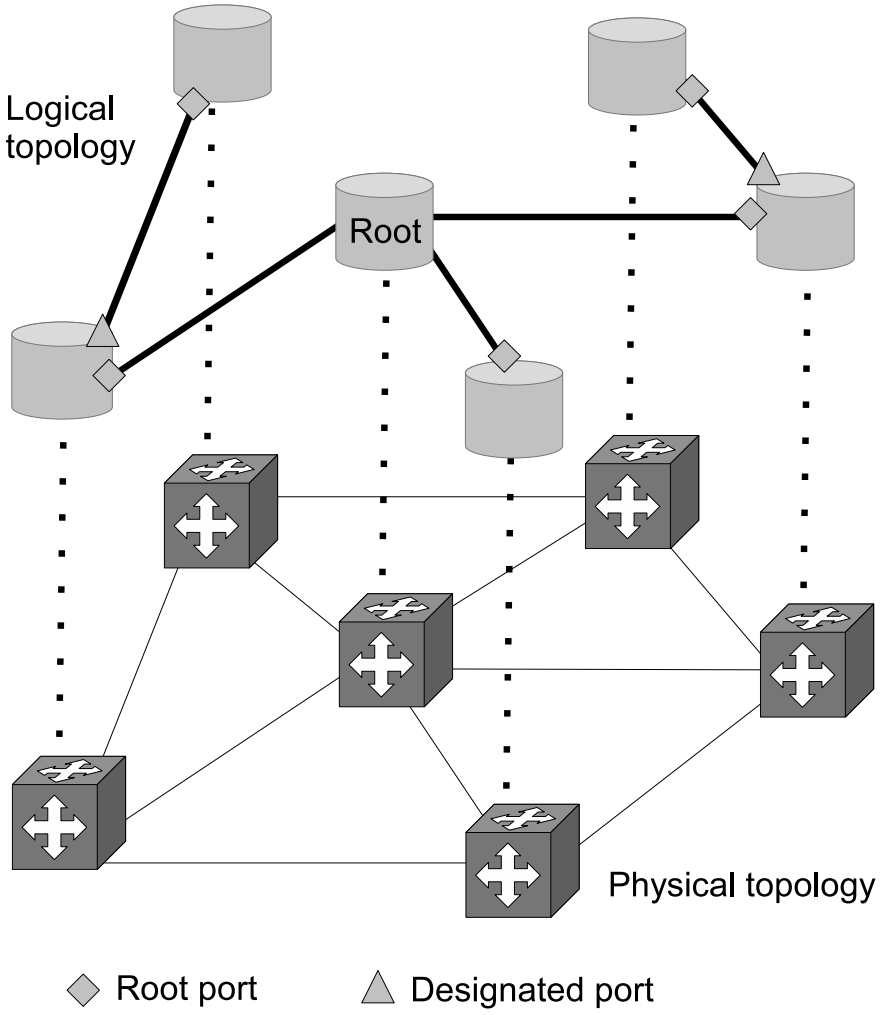


Figure 3.1: Logical topology consisting in a Spanning Tree built over a physical topology.

Unlike IP packets, Ethernet data frames do not have a Time-To-Live (TTL) field. STP prevents loops in the network by blocking redundant links that are not part of the tree. Therefore, the load is concentrated on a single path with no load balancing mechanism and being vulnerable to failures. In a network with n switches interconnected by l point-to-point links ($l > n - 1$), $n - 1$ links are active while the remaining ones do not carry any traffic. Just in case of failures, the blocked links are activated in order to reacquire global connectivity, providing a self-healing restoration mechanism. This is why the STP also suffers from low utilization, not being possible to activate links which would generate a loop.

The IEEE 802.1Q VLAN Protocol enables the operator to define different Virtual LANs and, therefore, create different broadcast domains within the same switching nodes. A VLAN is a set of ports belonging to the same or to different switches that have full connectivity between each other. Traffic demands of different clients are assigned to different VLANs in order to prevent packets from one client to reach ports of other clients, thus optimizing network resources.

All network topology changes, including failures, need a recalculation of a new Spanning Tree to reacquire global connectivity. This reconfiguration requires up to 50 seconds because of the timer based operations and, thus, affects network performance, resulting in poor resiliency.

Another negative aspect concerning the STP is that it lacks of service guaranties, admission control on traffic policing and shaping, which would allow the support of QoS functionality. According to this, although STP has been used for most Ethernet networks, it has several serious shortcoming for MAN deployments.

To achieve faster convergence, the Rapid Spanning Tree Protocol (RSTP) IEEE 802.1w [4] was introduced. It is an evolution from the STP which uses a negotiation mechanism to repair the connectivity in case of failures instead of a timer specified by the root bridge as in the STP case. This make it possible to accelerate the convergence to a new Spanning Tree, reducing the convergence time by 1 to 3 seconds.

In order to achieve a better utilization of the network with a reduced complexity, the IEEE 802.1s working group has introduced a Multiple Spanning Tree Protocol (MSTP)

[5], that allows a switch to participate in Multiple Spanning Trees, one tree for each group of VLANs, as those shown in Figure 3.2.

Existence of multiple overlapped tree instances provides a flexible way of implementing traffic engineering in Metro Ethernet. Firstly, it allows better network resilience, since a link failure does not affect demands assigned to Spanning Trees that do not use it. Secondly, it implies lower network link loads, which makes it possible to maximize the robustness to unpredicted demand growth and to minimize the service disruptions.

With MSTP, a network operator can also define multiple network regions which partition the network. Each region has its own Multiple Spanning Tree Instance, although limited to support only VLANs whose ports are in switches belonging to the same region, which limits the amount of load balancing that can be obtained. Outside the regions the protocol behaves as standard STP, assuring compatibility with legacy equipment: a common Spanning Tree like the one in Figure 3.3 connecting all switches of all regions is set-up by the protocol in such a way that a link failure inside a region does not affect the traffic routing in the other regions.

As can be seen in the following references, traffic engineering of Ethernet using Multiple Spanning Trees is a widely researched topic because of its importance. However the 802.1s draft does not provide any criterion in dividing the network in regions, generating Spanning Trees and mapping them to the VLANs. There is a trade-off between the single region case and the multiple regions case: the single region case provides a better load balancing, because all VLANs can be assigned to any Spanning Tree and not only to the ones belonging to the same region, while the multiple region case may seem superior in terms of failure recovery, as the impact of a network failure is confined to the region where it happened. Another element to be taken into account is that there is no limit to the number of MST regions in a network, but each Multiple Spanning Tree Instance (MSTI) can support up to 16 Spanning Tree Instances (STIs), due to scalability issues.

Despite this, it has been shown in [15] that considering the network as one single region also provides good results in terms of resiliency of the network.

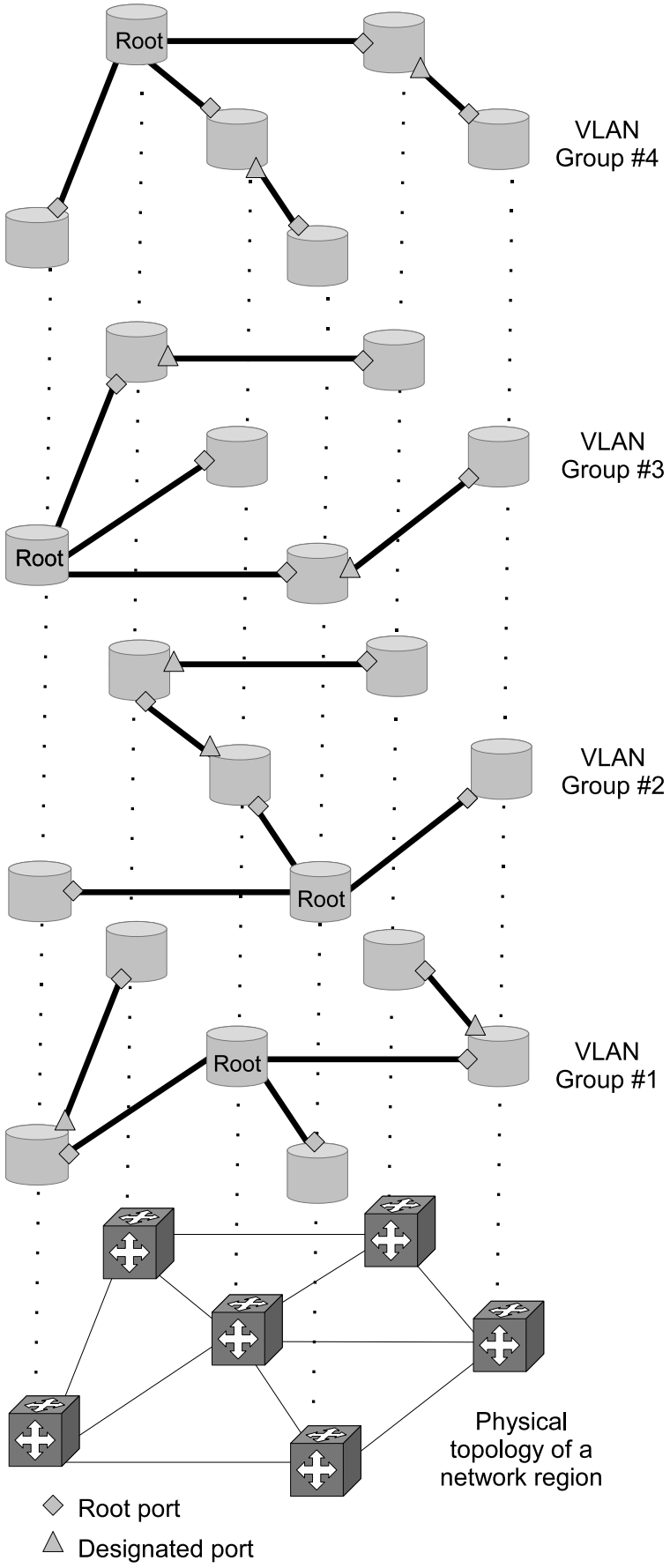


Figure 3.2: Multiple tree instances in a network region.

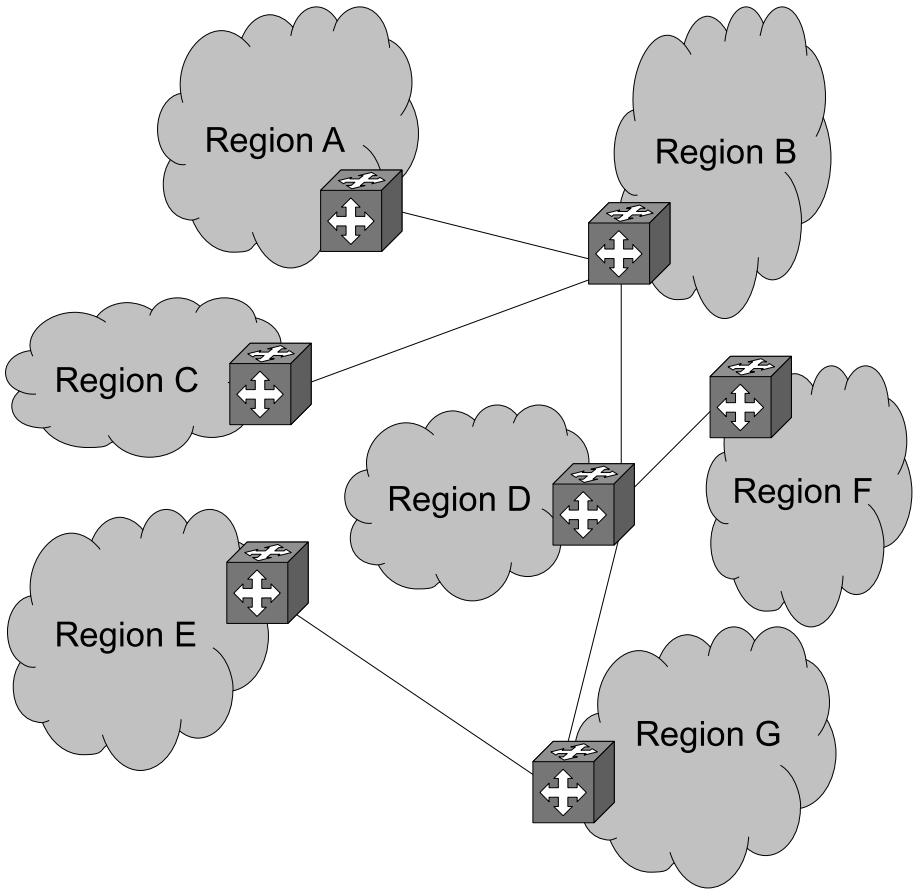


Figure 3.3: A common Spanning Tree connecting all switches of all regions.

Regarding the choice of Spanning Trees (STs), it would be essential to appropriately choose the Spanning Trees to be employed in order to obtain ones own objectives in managing the traffic. This points are crucial for the good performance of the MSTP.

3.2 Employing Multiple Spanning Tree Protocol to enhance Metro Ethernet performance

The problem of how to use the IEEE 802.1s Multiple Spanning Tree Protocol to enhance traffic engineering capabilities of Ethernet networks has been addressed by many authors. In these networks, traffic flows are defined on a per VLAN basis (the simplest VLAN that can be considered is an E-Line VLAN [18], i.e., a point-to-point VLAN which carry a single commodity) and routing of the traffic flows is done through the network based on STs, so that when a VLAN is assigned to a STI, its traffic flow is routed between its end nodes through the unique path defined by the assigned STI.

In [14] the author has considered the single region MSTP case and he has proposed a solving technique to determine the MSTP parameters configuration that minimizes the impact of network failures and optimizes load balancing. This objective is achieved minimizing the Worst case Link Load (WLL), which makes the network vulnerable to unpredictable traffic growth. Among all such solutions, the second Worst case Link Load is minimized, and than the third, an so on.

In [38] other objectives are introduced in order to obtain load balancing, such as to minimize the Average Link Load (ALL) given a Worst Link Load (WLL), or minimize the WLL with a guaranteed ALL. Since the solution of such models can be computationally expensive, also some heuristic solving techniques are proposed.

In [11] the authors claim the importance of an opportune cost metric in the tree construction algorithm. It is shown that the use of dynamic link cost function provides much higher network utilization than in the case when the link cost is constant. In order to achieve a good load balance and decrease the average delay of the network, a configurable link cost metric for the tree construction algorithm, which is a function of both current link load and link delay, is proposed. The algorithm considered, called

$(\alpha, 1-\alpha)$ algorithm, uses a link cost function as follows:

$$W_{ij} = \alpha D_{ij} + (1-\alpha)[C(ij) - B(ij) + R_i - R_{avg}] \tag{3.1}$$

where $0 \leq \alpha \leq 1$ is the link cost parameter, D_{ij} is the link delay between vertices i and j, $C(ij)$ and $B(ij)$ denote the link capacity and the available bandwidth between vertices i and j respectively, R_i is the sum of all traffic requests sourced from node i and R_{avg} is the current average used bandwidth of all links. R_{avg} is calculated as follows:

$$R_{avg} = 1/|E| \sum_{(i,j)\in E} C(ij) - B(ij) \tag{3.2}$$

where E is the set of links (i, j). If $\alpha = 1$, then $W_{ij} = D_{ij}$. The cost is determined only by the link delay, which optimizes the delay performance of the network. If $\alpha = 0$, then $W_{ij} = C(ij) - B(ij) + R_i - R_{avg}$. The link cost is determined only by the link load of the network, which achieves a strong load balance. If $0 < \alpha < 1$, then W_{ij} consists of two parts and the tree construction algorithms considers both the link delay and the available bandwidth.

In [25] the authors use a link cost metric which is function only of the link load:

$$W_{ij} = 1/B_{ij} \tag{3.3}$$

where W_{ij} is the link cost of link (i, j) and B_{ij} is the available bandwidth of link (i, j).

In [31] the link cost function proposed depends on the delay and on the available bandwidth of the links:

$$W_{ij} = D_{ij}/B_{ij} \tag{3.4}$$

where W_{ij} is the link cost of link (i, j), D_{ij} is the delay of link (i, j) and B_{ij} is the available bandwidth of link (i, j).

In [26], two algorithms are introduced: Multiple Spanning Tree Generation Algorithm (MSTGA) and VLAN Spanning Tree Mapping Algorithm (VSTMA). The MSTGA generate Spanning Trees that have the smallest number of links in common, so as to obtain a set of disjoint trees. It has been shown that edge disjoint trees yield near optimal bandwidth allocation and network performance while maintaining resilience to failures. Furthermore, with edge disjoint Spanning Trees, a single link failure affects a

single tree, reducing the number of impacted customers. In order to build the maximum possible number of disjoint trees, a three link metric is used. A metric called Link Usage Count is introduced: $luc(i, j)$ indicates the number of Spanning Tree instances using link (i, j). When building Spanning Trees, the algorithm prefers the links with the smallest $luc(i, j)$. In order to build the maximum number of disjoint trees, using a large number of links attached to the same node in the same tree, should be avoided. A $free - node(n)$ variable, which enumerates the unused links attached to node n is introduced. The last metric takes into account the VLAN location: they define a set of Potential Spanning Tree Links, which contains all the links between two VLAN site locations, and try to use them first while building a new Spanning Tree instance. The VLANs are also grouped in order that similar VLANs belong to the same VLAN Group (VG). This way, each VG will be assigned to one STI, the one that minimizes the number of transit nodes, obtaining an optimal mapping in terms of bandwidth usage. The VSTMA performs a good Spanning Tree mapping but this solution does not take into account the load balancing problem.

The same author, in [27], shows that using pragmatic methods that build a very small number of edge-disjoint Spanning Trees and perform an adequate VLAN Spanning Tree Mapping as shown in [26], it is possible to enhance the 802.1s MSTP, without requiring significant changes in the current Ethernet gear. The solution is called Smart Spanning Tree Bridging (SSTB) and his main objectives are to efficiently leverage the network and minimize bandwidth consumption while performance objectives are met. Figure 3.5 shows the trees build by this algorithm for the network topology in Figure 3.4. SSTB is compared to SPB and seems to be the more compelling between the two technologies, although conclusions are based solely on theoretical models.

The approach presented in [30], named Best Multiple Spanning Tree (BMST), improves the previous algorithm. It finds the best set of edge disjoint Spanning Trees based on the shortest path selection and on links/switches load balancing. Three coefficients correspond to the above criteria, respectively, and allow to weigh the importance of each criterion. Even if the algorithm can find the best answer for small networks, its complexity is too large for large scale networks.

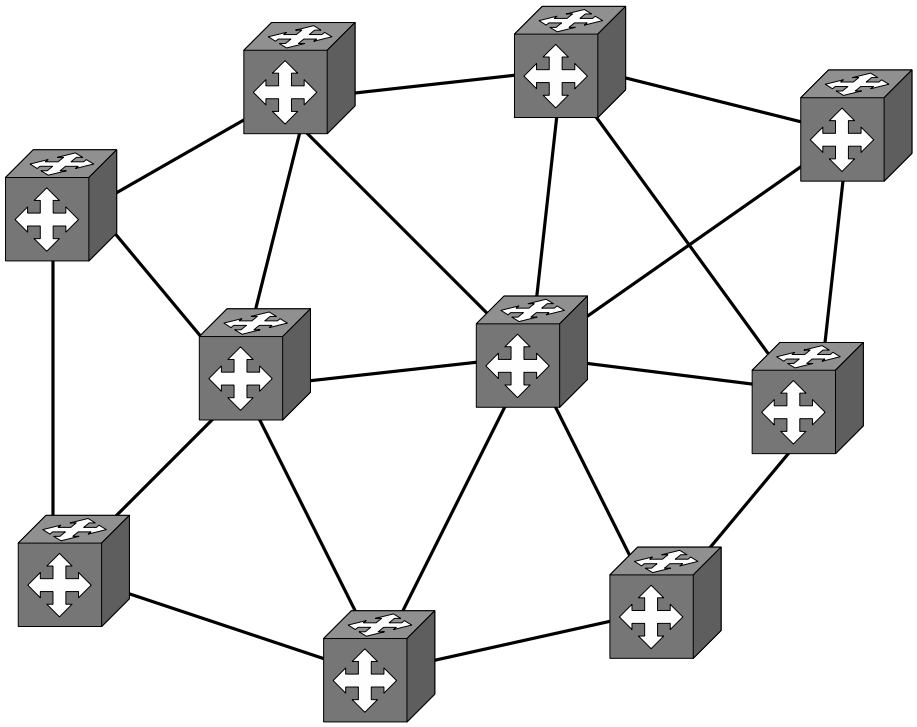


Figure 3.4: Network Topology.

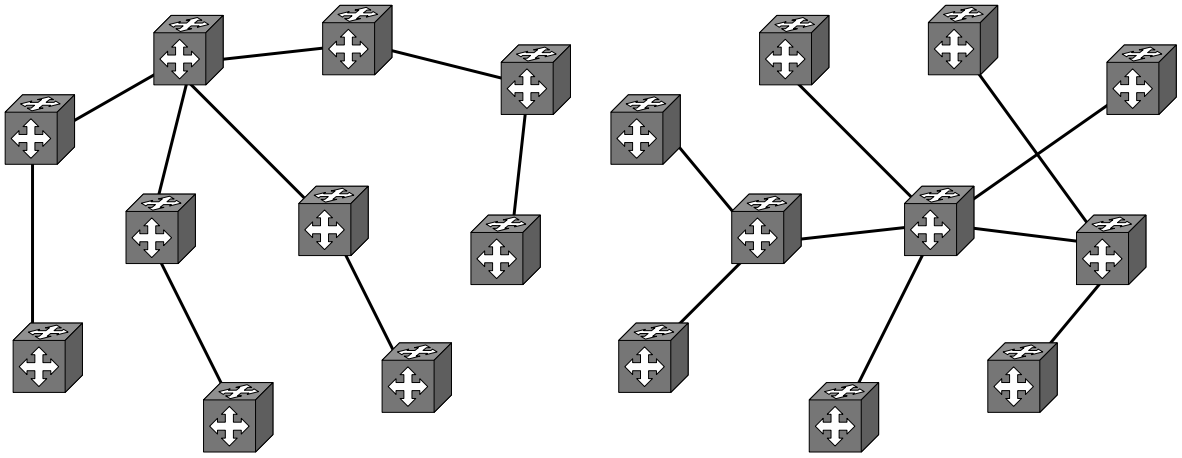


Figure 3.5: Only two disjoint Spanning Trees build by Smart Spanning Tree Bridging algorithm. One tree for each VLAN Group

Finally in [17] a modified version of the MSTP is presented as a control protocol of an SPB network and compared with other possible solutions. A Multiple Spanning Tree Instance is assigned to each bridge, thus each bridge has its own SPT. Figure 3.6 shows the trees built by this algorithm.

In order to meet the congruency requirements, the BPDU have to be extended with a Path Vector, which lists all the bridges comprising a path. This solution does not seem to be able to perform well in a sparse topology, because control information travels over long paths. However, the MSTP is the best choice for better connected topologies because it is more likely that a safe alternate path (to be activated in case of failure) exists.

As can be seen from the preceding paragraphs, the use of MSTP as a means to enhance traffic engineering capabilities in the provision of Ethernet services has been recently addressed by many authors. In all these works, however, they merely address the problems concerning load balancing, the quality of service and how to limit the impact of network failures. This essay proposes to use the IEEE 802.1s MSTP in order to minimize the energy consumption of a given network, finding the best subset of Spanning Trees and the best mapping of the traffic demands to them.

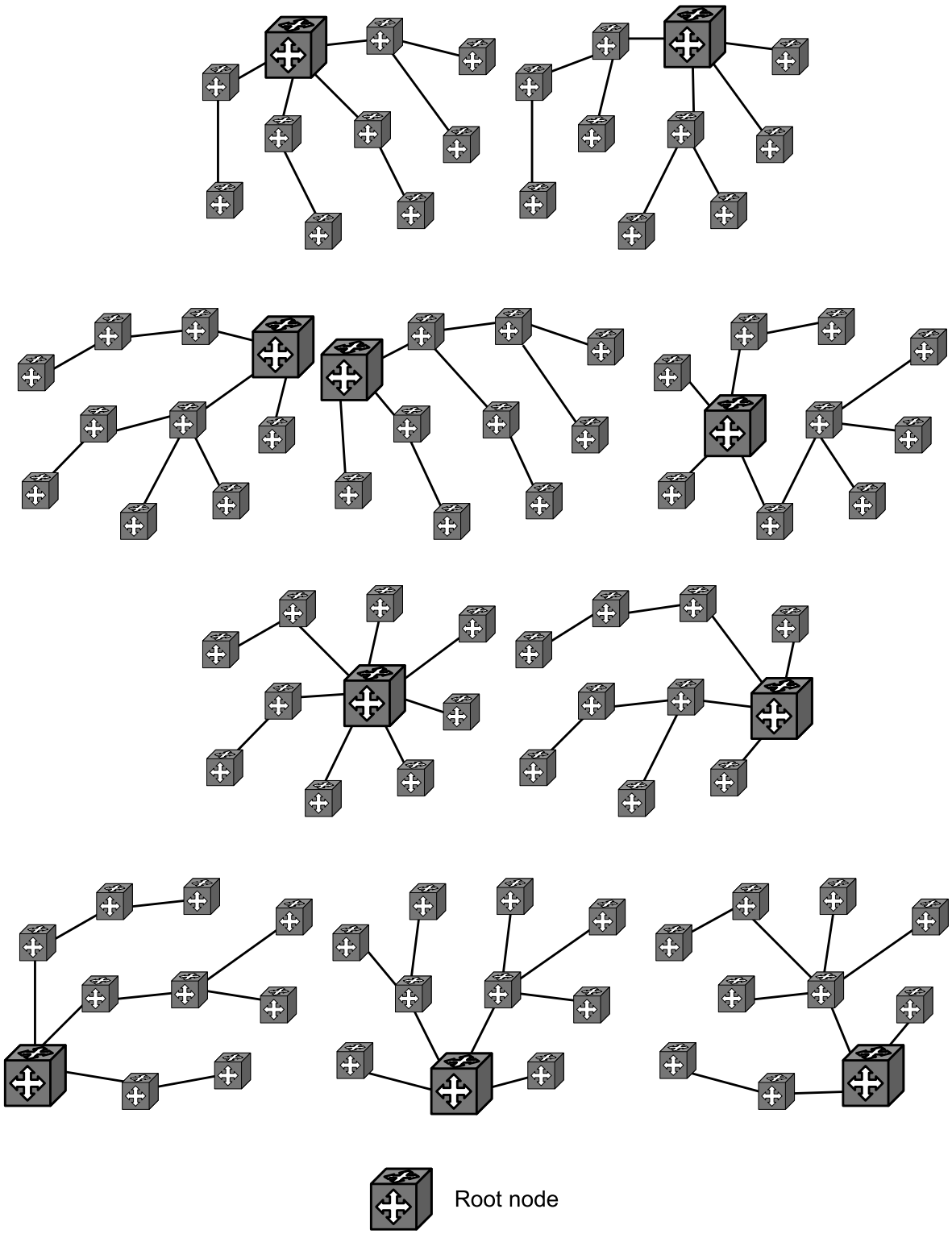


Figure 3.6: With SPB each node has its own Short Path Tree.

Chapter 4

The model

This chapter describes the optimization model proposed to make traffic engineering, considering the minimization of the energy consumption and the load balancing.

4.1 The sketch of the model

The principal objective of this essay is the energy consumption optimization problem in telecommunication networks addressed through traffic engineering, based on Multiple Spanning Tree (MST). This problem affects the Carrier Grade Ethernet networks employed in the Metropolitan Area Networks (MANs).

Consider an Ethernet Carrier network composed by switches connected through point-to-point links. The network could be represented by a graph consisting of nodes and edges. Since Ethernet uses full-duplex links, traffic in both directions must be considered separately. For this reason, edges must be distinguished from arcs, where an arc is equivalent to a directed edge. Each edge has a known full-duplex capacity and an energy consumption, as well as each node has an energy consumption. The network supports a set of VLANs, each of which carries a single commodity characterized by its origin node, its destination node and its traffic demand.

The algorithm presented aims at finding the best Spanning Trees in a given set and the best mapping of the traffic demands to the chosen Spanning Trees, so that part of the network is forced to remain unused, thus allowing to be turned off (or in other words, are put in sleep mode). This leads to use the fewest possible nodes and arcs, while continuing to ensure full connectivity and the management of all requests.

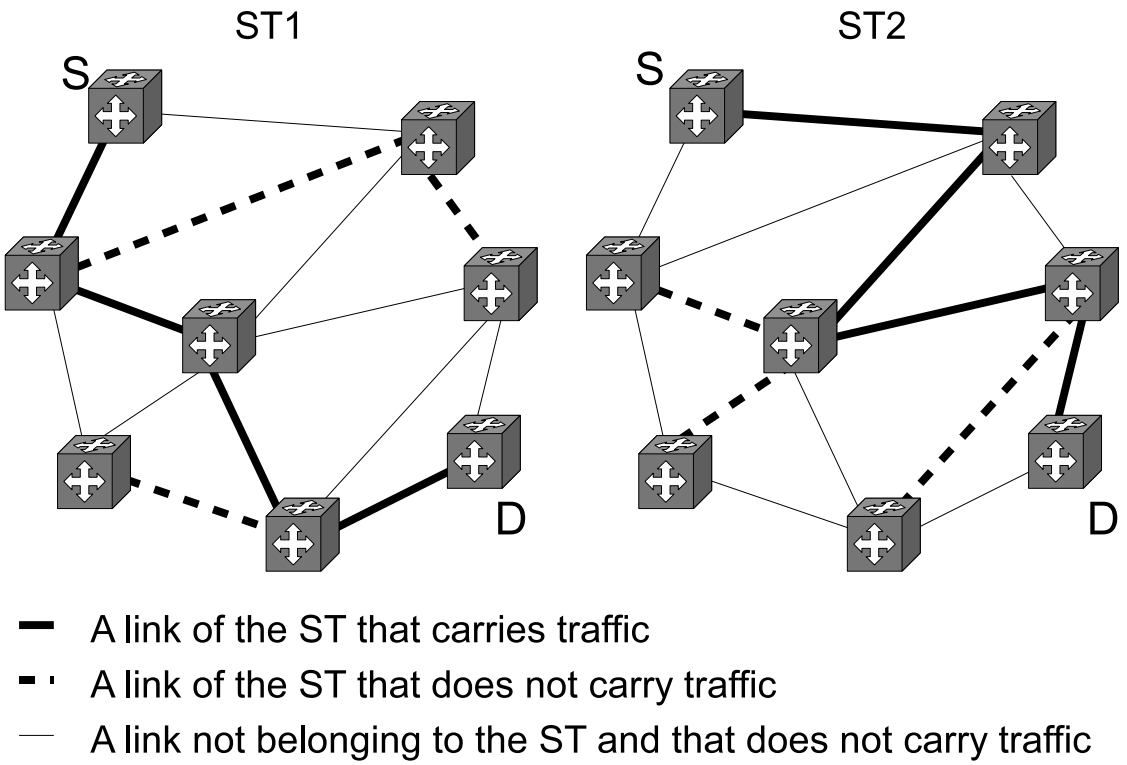


Figure 4.1: A VLAN carrying a commodity with source node S and destination node D matched with two different Spanning Tree results in two different paths between source and destination.

The traffic flows are routed over the network over paths defined by the STs the VLANs have been assigned to. Assigning a VLAN to a ST means determining the route that should be taken by the traffic of that VLAN. Assigning it to a different Spanning Tree means to route it over a different path. An example is given in Figure 4.1. Note that the links which does not carry traffic (even the ones belonging to the Spanning Tree matched with the considered VLAN) can be switched off.

Moreover, suppose that the links in bold in the network of Figure 4.2 are the ones through which the traffic is routed. If a portion of traffic should be sent between node S and node D, this flow, considering the minimization of the energy consumption, will be more likely assigned to a Spanning Tree whose path between the source and the destination node belongs to the portion of the network which has already been employed (and therefore not passing through nodes A, B and C).

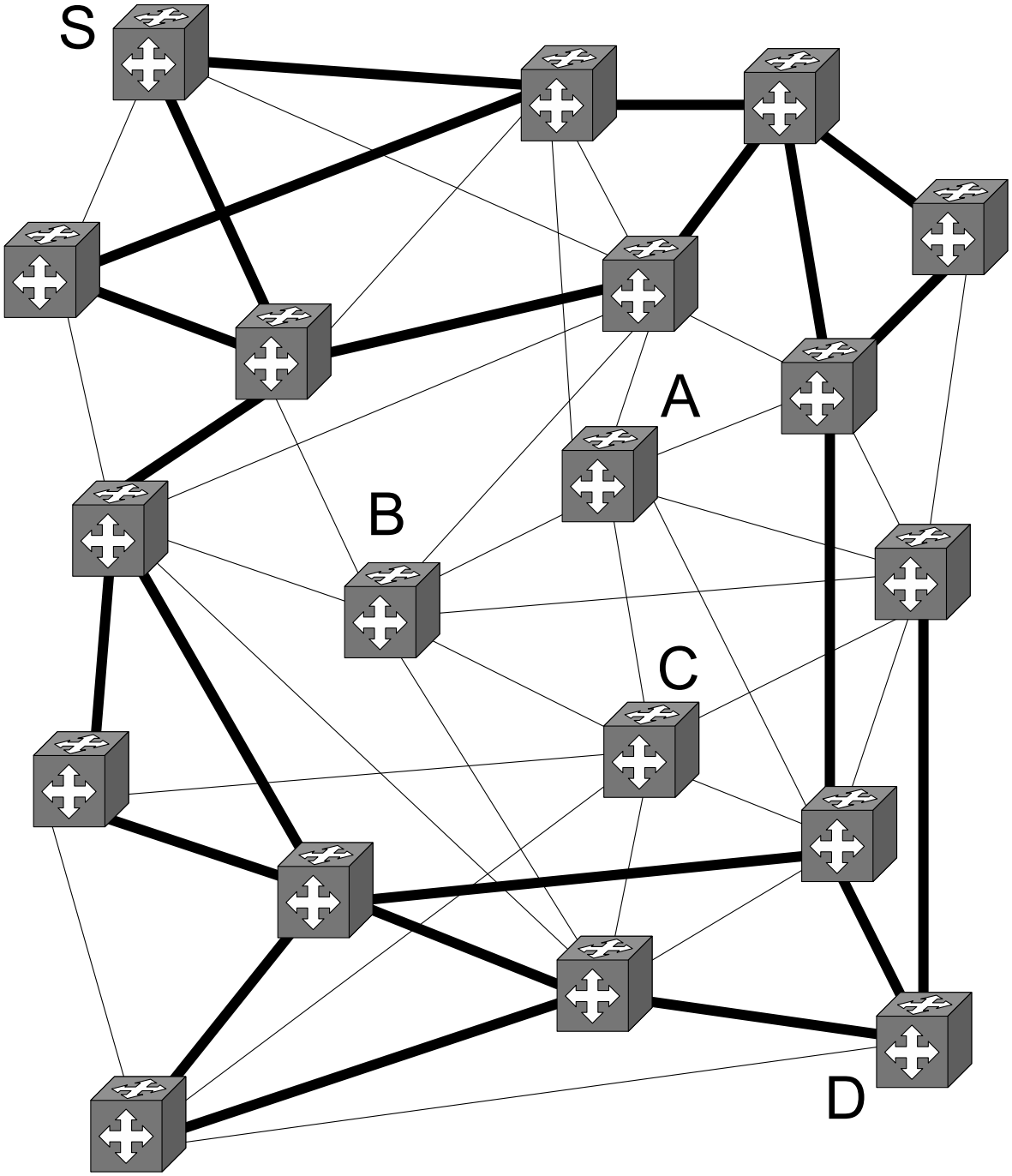


Figure 4.2: An example of how the algorithm should led to use the fewest possible number of nodes and arcs avoiding to route traffic in a portion of the network (in this case the one that includes nodes A, B and C).

4.2 The formulation

After a brief description of the modelling philosophy, additional notation is required before presenting the formulation of the model.

4.2.1 Notational description

The notation is divided in model parameters and decision variables:

Model parameters

N	the set of nodes, each of which represents a switch.
E	the set of edges, each of which represents an undirected link between two nodes.
$G(N, E)$	the graph composed by nodes and edges representing a network.
A	the set of arcs, that are directed links between nodes.
Q	the set of VLANs to be routed.
S	the set of Spanning Trees.
$o_q \in N$	the origin node of the VLAN $q \in Q$.
$b_q \in N$	the destination node of the VLAN $q \in Q$.
$d^q \in \mathbb{R}^+$	the traffic demand of the VLAN $q \in Q$.
$c_{\{ij\}} \in \mathbb{R}^+$	the full-duplex capacity of the edge $\{i, j\} \in E$.
$\varepsilon_{\{ij\}} \in \mathbb{R}^+$	the energy consumption of the edge $\{i, j\} \in E$.
$\varepsilon_i \in \mathbb{R}^+$	the energy consumption of the node $i \in N$.
$K \in \mathbb{N}$	the maximum number of Spanning Trees that can be mapped with the VLAN.
$\alpha \in [0, 1]$	the trade-off parameter that determines how much importance should be given to the energy savings objective ($\alpha = 1$ means that only the energy consumption is considered) and to the load balancing objective ($\alpha = 0$ means that only the load balancing objective is considered, no matter how much energy is spent).

$$a_{ij}^{sq} = \begin{cases} 1 & \text{if arc } (i,j) \in A \text{ is in the path of the Spanning Tree } s \in S \\ & \text{for the VLAN } q \in Q. \\ 0 & \text{otherwise.} \end{cases}$$

The last parameter is as crucial as complex. It is a binary parameter that defines if a VLAN could be routed over a link accordingly to the Spanning Tree it is has been matched to.

Decision variables

The problem of limiting the energy consumption of the network routing the VLAN through the best Spanning Trees (STs) while continuing to ensure full connectivity and the management of all requests can be formulated using four decision variables.

The first variables state which STs have been chosen in the solution among the set S:

$$\lambda_s = \begin{cases} 1 & \text{if Spanning Tree } s \in S \text{ is chosen in the solution.} \\ 0 & \text{otherwise.} \end{cases}$$

The mapping between VLANs and STs is defined by the second binary variables:

$$\phi_s^q = \begin{cases} 1 & \text{if VLAN } q \in Q \text{ is assigned to Spanning Tree } s \in S. \\ 0 & \text{otherwise.} \end{cases}$$

Finally, the last two groups of variables are those that determine whether a node or an edge respectively are switched on or not:

$$x_{\{ij\}} = \begin{cases} 1 & \text{if edge } \{i,j\} \in E \text{ is switched on.} \\ 0 & \text{otherwise.} \end{cases}$$

$$y_i = \begin{cases} 1 & \text{if node } i \in N \text{ is switched on.} \\ 0 & \text{otherwise.} \end{cases}$$

4.2.2 The energy objective function and constraints

Having defined all the parameters and decision variables, the energy consumption model (denoted by $P1$) can be written as follow.

Objective function

$$\min z_1 = \sum_{i \in N} y_i \varepsilon_i + \sum_{\{i,j\} \in A} x_{\{ij\}} \varepsilon_{\{ij\}} + \beta \left\{ \sum_{i \in N} (1 - y_i) \varepsilon_i + \sum_{\{i,j\} \in A} (1 - x_{\{ij\}}) \varepsilon_{\{ij\}} \right\} \quad (4.1)$$

The term z_1 is composed of two parts. The first is the energy consumption of the switched on nodes and edges belonging to the network. The second is the the consumption of the elements in sleep mode. It is weighted by the parameter β, that is set to 0.1 according to the EEE estimates.

Constraints

$$\sum_{s \in S} \sum_{q \in Q} a_{ij}^{sq} \phi_s^q d^q \leq c_{\{ij\}} x_{\{ij\}} \qquad \forall (i,j) \in A \tag{4.2}$$

Constraints (4.2) are the capacity constraints for each arc (i, j). If the edge $\{i, j\}$ is switched off, no traffic can be routed on it.

$$\phi_s^q \leq \lambda_s \qquad \forall s \in S, q \in Q \tag{4.3}$$

Constraints (4.3) guarantee that a single VLAN q can be assigned to a Spanning Tree s only if s has been chosen in the solution.

$$\sum_{s \in S} \phi_s^q = 1 \qquad \forall q \in Q \tag{4.4}$$

Constraints (4.4) guarantee that each VLAN q is assigned to one Spanning Tree s.

$$\sum_{s \in S} \lambda_s \leq K \tag{4.5}$$

Constraint (4.5) guarantees that no more than a fixed number K of trees are in the solution.

$$\sum_{j \in N} x_{\{ij\}} \leq y_i 2K \qquad \forall i \in N \tag{4.6}$$

Constraints (4.6) guarantee that a node i can be switched off only if there are no active links incoming or outgoing from that node.

4.2.3 The load balancing objectives function and constraints

Designing networks based on power consumption only would lead to tree shaped networks with concentrated connections, which are extremely vulnerable to failures. That is why the load balancing objective should also be taken into account.

Considering optimization of load balancing, many different objectives can be set, as it is explained in [38]. One possible objective is to minimize the Worst Link Load

(WLL). Having the maximum link load among all links equal to U for a certain link $(i,j) \in A$, with $0 \leq U \leq 1$, implies that the traffic could increase by a factor of $(1-U)/U$ before that link is saturated. This means that the lower the worst utilization value U is, the more robust the network becomes against unpredictable traffic growth. Another possible objective could be to minimize the Average Link Load (ALL), which maximizes the load over the unused network resources, making it easier to manage new traffic demands.

In this work two of the load balancing objectives proposed in [38] have been considered. To present them, some additional notation and decision variables are required:

Model parameters

$\mu^* \in [0,1]$	the maximum utilization value accepted on each arc $(i,j) \in A$.
$\overline{\mu} \in [0,1]$	the maximum accepted average utilization value over all arcs $(i,j) \in A$.

Decision variables

$u_{ij} \in [0,1]$ represents the utilization of link (i,j).

P2a: first load balancing problem

The first load balancing objective is the minimization of the Average Link Load with a guaranteed optimal Worst case Link Load. The problem (from now on referred as $P2a$) can be modelled as follows. Let z_{2a} be the Average Link Load of the network. Then:

$$\min z_{2a} = \frac{1}{|A|} \sum_{(i,j)\in A} u_{ij} \tag{4.7}$$

s.t.

$$(4.3 - 4.5),$$

$$\sum_{s\in S} \sum_{q\in Q} a_{ij}^{sq} \phi_s^q d^q \leq c_{\{ij\}} u_{ij} \qquad \forall (i,j) \in A, \tag{4.8}$$

$$u_{ij} \leq \mu^* \qquad \forall (i,j) \in A. \tag{4.9}$$

Constraints (4.8), similarly to constraints (4.2), are capacity constraints, but unlike them they make explicit the utilization factor for each arc (i, j) employing the variable u instead of the binary variable x. Contraints (4.9) limit the utilization value of each link below the threshold of μ^*.

P2b: second load balancing problem

The second load balancing objective is the minimization of the Worst case Link Load with a guaranteed optimal Average Link Load. This problem (from now on referred as *P2b*) can be modelled as follows. Let the term z^*_{2b} be equal to the Worst Link Load of the network. Then the objective function can be written as

$$\min\ z^*_{2b} = \max_{ij} u_{ij} \tag{4.10}$$

The minmax function can then be rewritten as follow:

$$\begin{aligned}
&\min\ z_{2b} = \mu && && (4.11)\\
&\text{s.t.}\\
&\quad (4.3, 4.4, 4.5, 4.8),\\
&\quad \frac{1}{|A|} \sum_{(i,j)\in A} u_{ij} \leq \overline{\mu} && \forall (i,j) \in A, && (4.12)\\
&\quad u_{ij} \leq \mu && \forall (i,j) \in A, && (4.13)\\
&\quad \mu \in [0, 1].
\end{aligned}$$

Contraints (4.12) limit the average utilization value over all links below the threshold of $\overline{\mu}$.

4.2.4 The trade-off objective function and constraints

A trade-off between the energy consumption and the QoS should be taken into account. The energy savings objective can be combined with each of the two load balancing objectives thus creating a new model. To do that, some additional notation must be declared.

Model parameters

$\nu^* \in [0,1]$	the maximum utilization value obtained on each arc $(i,j) \in A$ when the energy objective is solved.
$\overline{\nu} \in [0,1]$	the average utilization value obtained over all arcs $(i,j) \in A$ when the energy objective is solved.

P3a: first trade-off model

The first trade-off model (denoted by $P3a$) is the combination between the energy objective 4.1 of the problem $P1$ and the load balancing objective 4.7 of the problem $P2a$. It is defined as follows.

$$\min\ z_{3a} = \alpha z_1 + (1-\alpha) z_{2a} \tag{4.14}$$

Let α determines how much importance should be given to the energy savings objective ($\alpha = 1$ means that only the energy consumption is considered) and to the load balancing objective ($\alpha = 0$ means that only the load balancing objective is considered, no matter how much energy is spent). Then the problem can be written as:

s.t.

$$(4.2 - 4.6),$$

$$u_{ij} = \sum_{s \in S} \sum_{q \in Q} a_{ij}^{sq} \phi_s^q d^q / c_{\{ij\}} \qquad \forall (i,j) \in A, \tag{4.15}$$

$$u_{ij} \leq \mu^* + (\nu^* - \mu^*)\alpha \qquad \forall (i,j) \in A. \tag{4.16}$$

Contraints (4.15) are meant to calculate the utilization of each arc (i,j). Constraints (4.16) limit the utilization of each link below a threshold that varies between μ^* and ν^* depending on the value assumed by the parameter α. μ^* is the maximum utilization value admitted on each arc $(i,j) \in A$, while ν^* is the maximum utilization value obtained on each arc $(i,j) \in A$ when the problem $P1$ is solved. So, if $\alpha = 0$ constraints (4.16) are equivalent to constraints (4.9), while if $\alpha = 1$ constraints (4.16) are irrelevant, because all the utilization parameters take on values below ν^*. For intermediate values of α a linear combination between μ^* and ν^* was assumed.

P3b: second trade-off model

The second trade-off model (denoted by $P3b$) is the combination between the energy objective 4.1 of the problem $P1$ and the load balancing objective 4.11 of the problem $P2b$. It is defined as follows.

$$\min\ z_{3b} = \alpha z_1 + (1-\alpha) z_{2b} \tag{4.17}$$

s.t.

$$(4.2, 4.3, 4.4, 4.5, 4.6, 4.13, 4.15),$$

$$\frac{1}{|A|} \sum_{(i,j)\in A} u_{ij} \leq \overline{\mu} + (\overline{\nu} - \overline{\mu})\alpha \qquad \forall (i,j) \in A. \tag{4.18}$$

Similarly to constraints (4.16), contraints (4.18) limit the average utilization value over all links below a threshold that varies between $\overline{\mu}$ and $\overline{\nu}$ depending on the value assumed by the parameter α. $\overline{\mu}$ is the maximum allowed value of the average utilization value over all arcs $(i,j) \in A$, while $\overline{\nu}$ is the average utilization value over all arcs $(i,j) \in A$ obtained when the problem $P1$ is solved. So, if $\alpha = 0$ constraints (4.18) are equivalent to constraints (4.12), while if $\alpha = 1$ constraints (4.18) are irrelevant, because the average utilization values will be equal to $\overline{\nu}$. Again, for intermediate values of α a linear combination between $\overline{\mu}$ and $\overline{\nu}$ was assumed.

Chapter 5

Experimental approach

This chapter lists the results obtained applying the model described in Chapter 4. Before doing that, it explains how the data and the parameters that serve as input to the model are obtained.

5.1 Preprocessing

5.1.1 Network data

The networks considered in this essay for all the optimization problems have been randomly generated through an instance generator implemented in the C++ programming language. Having fixed the number of nodes and edges, this program is able to generate a network from a full-mesh grid maintaining the edges in the network with a predetermined probability p and discarding the others. Knowing the topology of a real metropolitan network would be useful, but these informations are not easily disclosed by private corporations or ISP, as they are deemed confidential data. In these networks, all the links capacity were set to 100 Gbps. Concerning traffic demands, they too have been generated through the instance generator. Origin and destination have been randomly selected among all nodes and their bandwidth requests have been randomly generated between 0.1 Gbps and 10 Gbps.

To present computational results we have chosen a Cisco CRS-3 16-Slot Single-Shelf System [1]. This device is compatible with the Cisco CRS-3 1-Port 100 Gigabit Ethernet Interface Modules [2], each of which is responsible for the ingress and the

egress packet processing, so that in order to build up a 100 Gbps full-duplex link, we must employ two modules (one at each end).

Regarding the energy consumption parameters, they have been estimated from the datasheets of the devices mentioned above. The maximum power consumption of a router when chassis is fully configured with line cards and with traffic running is estimated to be 12320 Watts, while the energy consumption of a single line card is 150 Watts (therefore the energy consumption of a full-duplex link is considered to be 300 Watts). Finally, the power consumption in the low-power mode is assumed to be 10% of that in the active mode, in line with the estimates provided by different manufacturers during the standardization process of EEE [33].

5.1.2 Build the Spanning Tree

In addition to data on network topologies and traffic matrices, the model needs to have as input a large number of Spanning Trees of the considered networks. These are obtained through the Algorithm 5.1.1, which then adds them to the set S.

Algorithm 5.1.1: ST GENERATION(E, N, s)

$Tree \leftarrow \emptyset$
$S \leftarrow \emptyset$
for each $(i, j) \in E$
  do $cost(i, j) \leftarrow Uniform[0, 1]$
while $cont \leq s$
  do
    while $card(Tree) < card(N) - 1$
      do
        $BestEdge \leftarrow$ MinCost(E)
        $E \leftarrow E \backslash BestEdge$
        NoLoop$(BestEdge)$
    if $(\nexists\, T \in Forest : T = Tree)$ **and** criterion$(Tree)$
      then
        $S \leftarrow Forest \cup Tree$
        Update$(cost)$

Kruskal's algorithm has been used in order to find the Spanning Trees which are the elements of the set S. Each Spanning Tree has been generated by assigning random costs to the links with a uniform distribution, so that Kruskal's algorithm could give at each cycle a different Spanning Tree as output.

The function MinCost(E) provides the minimum cost link in the set E. The procedure NoLoop($BestEdge$) determine if the edge given as input will create a loop and adds it to the set $Tree$ if not. The procedure is defined by the following algorithm:

Algorithm 5.1.2: NoLoop($BestEdge$)

$$
\textbf{if } component[i] = 0
\begin{cases}
\textbf{then} & \begin{cases}
\textbf{if } component[j] = 0 & \\
\quad \textbf{then} & \begin{cases} component[i] \leftarrow NewComponent \\ component[j] \leftarrow component[i] \\ Tree \leftarrow Tree \cup (i,j) \end{cases} \\
\quad \textbf{else} & \begin{cases} component[i] \leftarrow component[j] \\ Tree \leftarrow Tree \cup (i,j) \end{cases}
\end{cases} \\
\textbf{else} & \begin{cases}
\textbf{if } component[j] = 0 & \\
\quad \textbf{then} & \begin{cases} component[j] \leftarrow component[i] \\ Tree \leftarrow Tree \cup (i,j) \end{cases} \\
\quad \textbf{else if } component[j] \neq component[i] & \\
\quad \textbf{then} & \begin{cases} \textbf{for each } n \in N : component[n] = component[j] \\ \quad \textbf{do } component[n] \leftarrow component[i] \\ Tree \leftarrow Tree \cup (i,j) \end{cases}
\end{cases}
\end{cases}
$$

In Algorithm 5.1.2 variable $component[i]$ indicate which component of the temporary Tree the node i belongs to. Its default value is zero. Given an Edge (i,j) four cases must be considered:

- neither node i or j belong to the tree ($component[i] = component[j] = 0$). In this case the edge (i,j) will not create loops and can be added to the set $Tree$;
- node i (or j) does not belong to the tree, while node j (or i) has already been added to the tree. Again the edge is added without creating loops;
- both nodes i and j already belong to the set $Tree$ but in two disconnected components ($component[i] \neq component[j]$). In this case the edge can be added and the two components are merged.
- both nodes i and j already belong to the set $Tree$ in the same connected component ($component[i] = component[j] \neq 0$). In this case the edge cannot be added, otherwise it would create a loop.

The function CRITERION($Tree$) provides an additional criterion to decide whether to add or not the considered tree to the set S. If not mentioned, it's default value is 1.

Finally, function UPDATE($cost$) assigns new costs to the links, according to the uniform distribution. When a tree is complete, it is added to the set S, and also saved into a parameter T such that $T[i, j, s] = 1$ if edge i, j belongs to the Spanning Tree s.

Other ways to generate trees have also been considered. In the first alternative, function CRITERION($Tree$) has been calibrated in such a way that only the trees which involve as few edges as possible could be added to the set S. In this case, the function has been set to 1 only when the number of edges employed to route traffic through the considered tree was below a threshold. Another option concerns the function to update the edge costs. In this case, at the end of each cycle a fixed value is added to the cost of every edge. This value is related to the number of time each edge has been chosen to belong to a tree, in order to discourage the choice of the edges that have been selected more often. This method should generate trees as diverse as possible. However, these alternative tree generator methods do not produce better results than the ones obtained with the basic one.

5.1.3 Parameter a_{ij}^{sq}

The data concerning the Spanning Trees are then used in the model in the form of the binary parameter a_{ij}^{sq} previously defined:

$$a_{ij}^{sq} = \begin{cases} 1 & \text{if arc } (i,j) \in A \text{ is in the path of the Spanning Tree } s \in S \\ & \text{for the VLAN } q \in Q. \\ 0 & \text{otherwise.} \end{cases}$$

The values for parameter a_{ij}^{sq} are calculated through a flow formulation, before running the model for each possible association between Spanning Trees and commodities. Note that given a a Spanning Tree $s \in S$ and a VLAN $q \in Q$, then the path between the origin node o_q and the destination node b_q is unique in s and the values of parameter a_{ij}^{sq} are uniquely determined. a_{ij}^{sq} must satisfy the following constraints:

$$\sum_{(i,j)\in A:T[i,j,s]=1} a_{ij}^{sq} - \sum_{(k,i)\in A:T[k,i,s]=1} a_{ij}^{sq} = \begin{cases} 1 & \text{if } i = o_q \\ -1 & \text{if } i = b_q \\ 0 & \text{otherwise} \end{cases} \quad (5.1)$$
$$\forall i \in N, \forall q \in Q, \forall s \in S$$

5.1.4 Parameter μ^* and $\overline{\mu}$

Two other parameters must be calculated before running the model. The first is μ^*, the maximum utilization value admitted on each arc $(i,j) \in A$. According to [38], its value is obtained by solving the Worst Link Load (WLL) subproblem:

$$\min_{u} \max_{ij} u_{ij}$$

s.t.

$$\begin{aligned} &\sum_{s\in S}\sum_{q\in Q} a_{ij}^{sq}\phi_s^q d^q \le c_{\{ij\}}u_{ij} && \forall (i,j) \in A \\ &\phi_s^q \le \lambda_s && \forall s \in S, q \in Q \\ &\sum_{s\in S} \phi_s^q = 1 && \forall q \in Q \\ &\sum_{s\in S} \lambda_s \le K \end{aligned} \quad (5.2)$$

This problem can be solved by modelling the minmax function using the same approach that has been employed to model the objective function (4.11):

$$\min \mu$$

s.t.

$$\begin{aligned} &\sum_{s\in S}\sum_{q\in Q} a_{ij}^{sq}\phi_s^q d^q \le c_{\{ij\}}u_{ij} && \forall (i,j) \in A \\ &\phi_s^q \le \lambda_s && \forall s \in S, q \in Q \\ &\sum_{s\in S} \phi_s^q = 1 && \forall q \in Q \\ &\sum_{s\in S} \lambda_s \le K \\ &u_{ij} \le \mu && \forall (i,j) \in A \\ &\mu \in [0,1] \end{aligned} \quad (5.3)$$

and then equalling μ^* to μ.

Similarly the parameter $\overline{\mu}$, which represents the maximum admitted value for the average utilization value over all arcs $(i,j) \in A$, can be obtained solving the Average Link Load (ALL) subproblem:

$$\min \frac{1}{|A|} \sum_{(i,j)\in A} u_{ij}$$

s.t.

$$\begin{aligned}
&\sum_{s\in S}\sum_{q\in Q} a_{ij}^{sq}\phi_s^q d^q \leq c_{\{ij\}}u_{ij} && \forall (i,j) \in A \\
&\phi_s^q \leq \lambda_s && \forall s \in S, q \in Q \\
&\sum_{s\in S} \phi_s^q = 1 && \forall q \in Q \\
&\sum_{s\in S} \lambda_s \leq K && (5.4)
\end{aligned}$$

and then equalling $\overline{\mu}$ to the value of the objective function.

5.1.5 Adjusting the objective functions

Finally, the trade-off objective functions (4.14) and (4.17) are actually written as follow:

$$z_3 = \alpha \frac{z_1}{|A|\varepsilon_{\{ij\}} + |N|\varepsilon_i} + (1-\alpha) z_{2a} \tag{5.5}$$

and

$$z_3 = \alpha \frac{z_1}{|A|\varepsilon_{\{ij\}} + |N|\varepsilon_i} + (1-\alpha) \frac{z_{2b}}{10} \tag{5.6}$$

where the term in the denominator $|A|\varepsilon_{\{ij\}} + |N|\varepsilon_i$ and 10 are normalization terms for z_1 and z_{2b} respectively, which otherwise would take much larger values than z_{2a}, making it difficult to manage the trade-off in relation to the parameter α.

5.2 Computational Results

In the following paragraphs the results obtained from the optimizations over different networks are shown. These optimizations can be summarized as follows in Table

Objectives	Data	Function of	Varying
Energy consumption (P1)	Time	|Q|	|N|,|E|
	Energy savings	|Q|	|N|,|E|
	Energy savings	K	|N|,|E|
	Energy savings	|S|	|N|,|E|
Energy (P1) vs Load balancing (P2a)	ALL	|Q|	α
	WLL	|Q|	α
	Energy savings	|Q|	α
Energy (P1) vs Load balancing (P2b)	ALL	|Q|	α
	WLL	|Q|	α
	Energy savings	|Q|	α

Table 5.1: Roadmap of the optimizations performed.

5.1. Each objective function is calculated analyzing the computational time, the energy savings and the QoS indexes as function of some parameters and varying the dimensions of the networks or the trade-off parameter α.

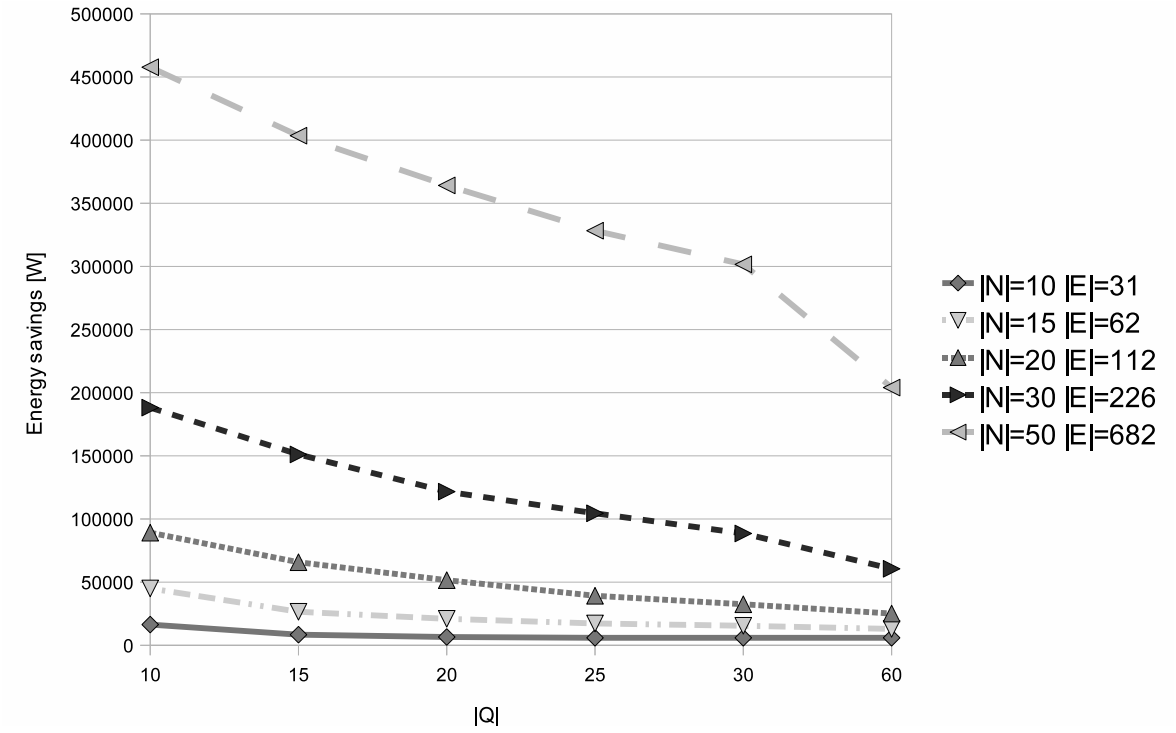
Finally, a comparison between the data obtained by considering the three objectives is shown for networks of different size, different traffic loads and a fixed value of trade-off equal to $\alpha = 0.2$

5.2.1 Minimizing energy consumption

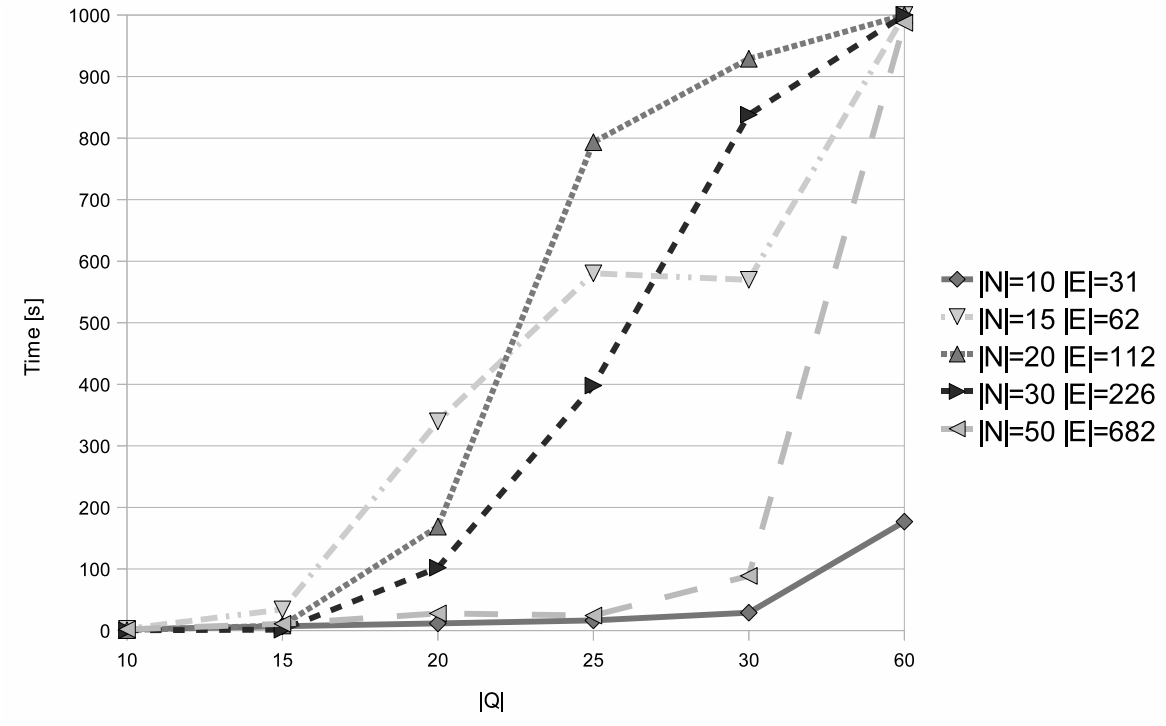
Results were initially obtained by solving energy model $P1$ (4.1) for different traffic loads (i.e. for different number of VLANs $|Q|$ to be routed).

Figure 5.1a shows how much energy is saved by putting to sleep a portion of the network according to the model, compared to the fully switched on consumption.

The data relates to networks of different sizes. It is easy to see that the big savings are obtained with the largest networks, since small networks (from 10 to 20 nodes) are forced to keep active most of their network elements even to route small amounts of traffic. Note that the energy savings achieved decreases proportionally with the increase on the routed VLANs, until the network is saturated and no more savings can be obtained. As expected, the biggest savings are also achieved in case of low traffic levels, when it is easier to find unused network elements. All values are averaged over 15 different instances.



(a) Energy savings



(b) Computational results

Figure 5.1: Data for networks of different sizes and different number of considered VLANs $|Q|$ obtained by applying the model with the energy objective.

$\|N\|$	$\|E\|$	Differential savings (K)				
		2	**3**	**5**	**10**	**15**
10	31	0	0	5940	0	0
15	62	0	0	26152	0	0
20	112	0	0	61902	0	0
30	226	-8658	-540	153828	0	0
50	706	-35982	-18126	406530	+7848	+7848

Table 5.2: Differential energy savings for networks of different sizes and a varying number of MSTI allowed. Differentials have been taken with respect to the case of $K = 5$.

Figure 5.1b shows the computational time required to solve the model. A time limit of 1000 seconds has been imposed. Again, all values are averaged over 15 different instances.

In all cases it was considered that a maximum number of $K = 5$ Spanning Trees were chosen out of $|S| = 30$ possible Spanning Trees. Actually most of the switches support up to 15 Multiple Spanning Tree Instances, but increasing the allowed number of instances to more than 5 does not produce significant benefits, as shown by Table 5.2.

Moreover, as regards to the number of trees used, it tends to be less than 5 (up to even a single tree) when the energy objective function is considered. This is not true when one considers also the load balancing problem $P2a$ or $P2b$: in this cases the number of trees used tends to be equal to 5, according to Table 5.5

The tests also show that increasing the size of set S does not improve the performance of the model in terms of energy savings, but merely expands the computational time.

5.2.2 Trade-off

We recall that two load balancing objective have been considered in this work. The first is the minimization of the Average Link Load with a guaranteed optimal Worst case Link Load. The second load balancing objective is the minimization of the Worst case Link Load with a guaranteed optimal Average Link Load. Each of them results in a trade-off model: problem $P3a$ and problem $P3b$ respectively.

Figure 5.2 shows the results obtained solving problem $P3a$, whose objective is the trade-off between the energy savings and the first load balancing as objectives for a network with $|N| = 10$ Nodes, $|E| = 31$ Edges and $|Q| = 60$ VLANs.

Different values of the parameter α, which defines the weight of the two objective mentioned above, has been considered. The links on which traffic is routed are in bold. The numbers in parentheses in the caption indicate how many edges of each of the five chosen Spanning Trees are used to route traffic. For example, in Fidure 5.2a 4 edges belonging to the first Spanning Tree carry traffic, while in the second Spanning Tree they are 8, and so on. As it can be seen, if one considers only the energy consumption objective, the portion of the network that is employed to route traffic is reduced to a tree. The more the load balancing objective is taken into account, the more the number of edge used increases.

By way of example, Figure 5.3 illustrates in more detail the results concerning the network of Figure 5.2a, the one that considers only the energy consumption objective.

It shows the Spanning Tree Instance (STI) chosen for the network, highlighting the links being used to route the commodities in each Spanning Tree. The links in bold are the ones used to route the traffic. Note that in this case, only four ST have been included in the solution.

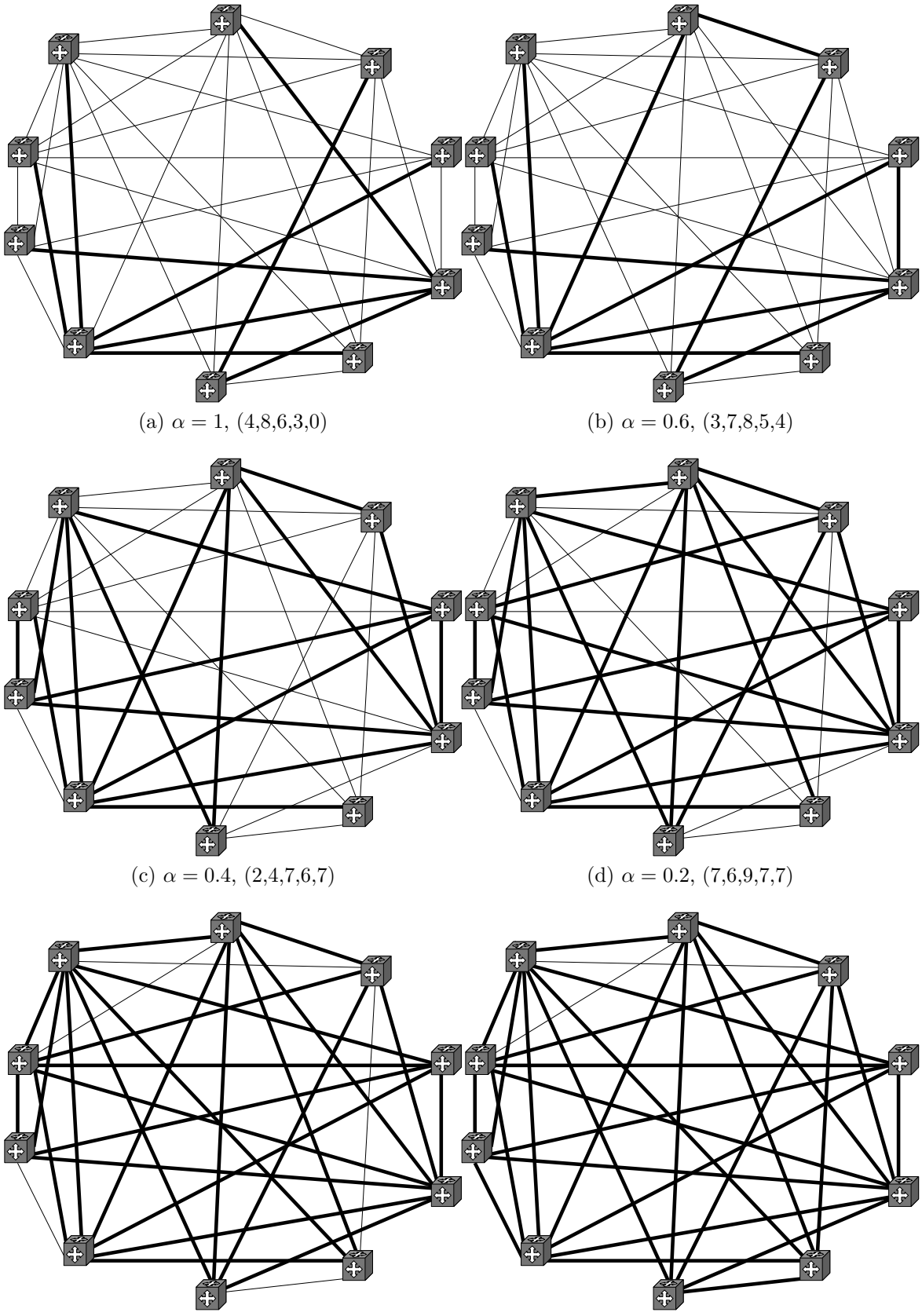
The data in Figures 5.4a, 5.5a and 5.6a summarize the value of the ALL, WLL and energy saving respectively, averaged over 15 different network instances of the type in Figure 5.2 with different values of the parameter α and for different traffic loads when the objective (4.14) is considered.

Similarly, Figure 5.4b, Figure 5.5b and Figure 5.6b summarize the same values when the objective (4.17) is considered.

These graphs show that both trade-off models behave similarly, regardless of the load balancing function that is considered.

5.2.3 Comparison of results

In this section the energy consumption problem $P1$ and the two trade-off problems $P3a$ and $P3b$ are analysed in greater detail, comparing the data obtained for networks



(a) $\alpha = 1$, (4,8,6,3,0)

(b) $\alpha = 0.6$, (3,7,8,5,4)

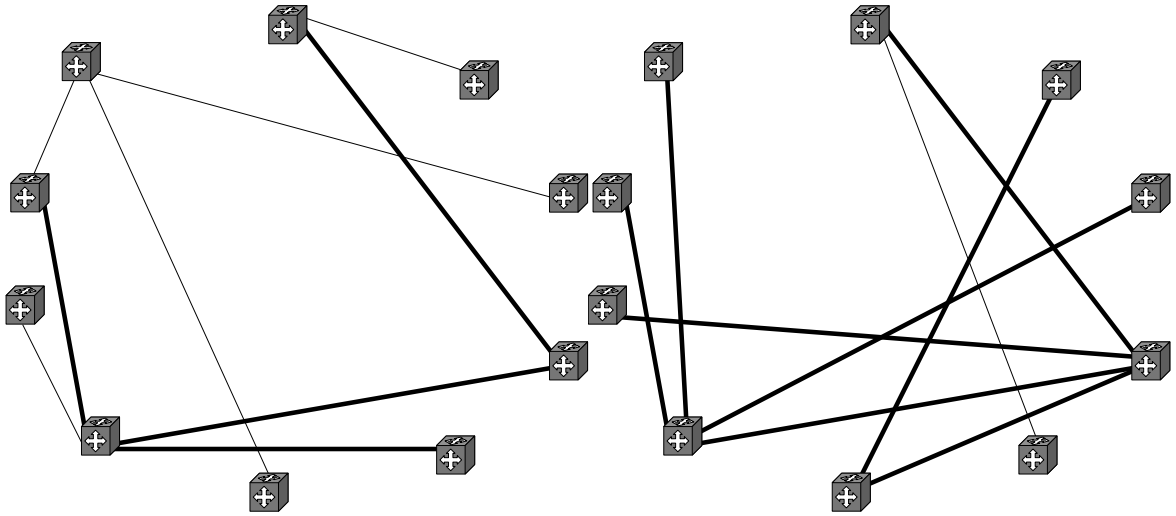
(c) $\alpha = 0.4$, (2,4,7,6,7)

(d) $\alpha = 0.2$, (7,6,9,7,7)

(e) $\alpha = 0.1$, (4,7,8,7,8)

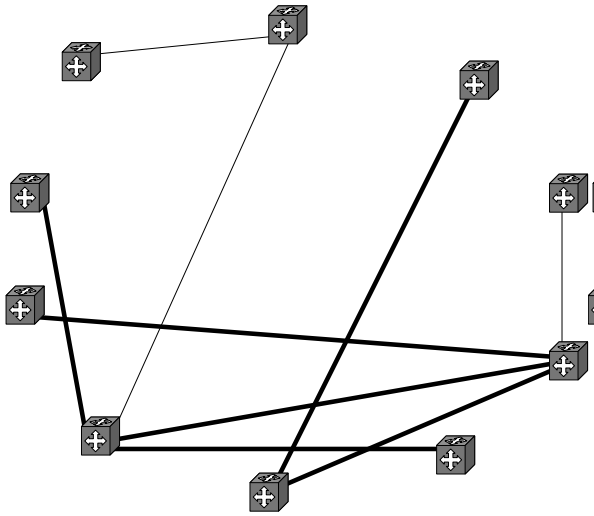
(f) $\alpha = 0$, (8,7,7,9,8)

Figure 5.2: The networks obtained considering the trade-off problem $P3a$ varying parameter α.

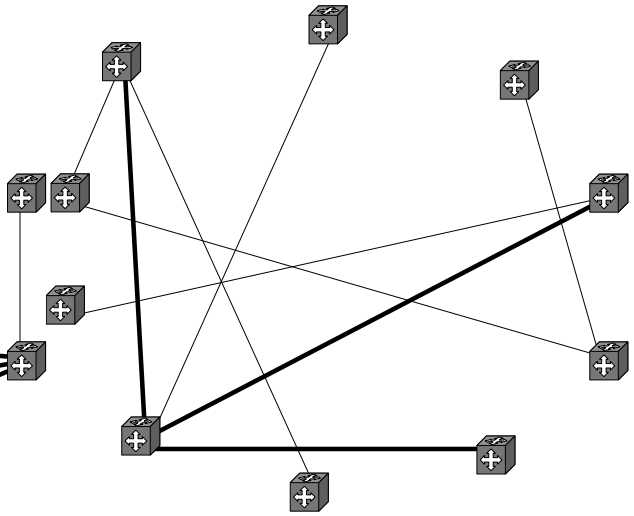


(a) ST1

(b) ST2

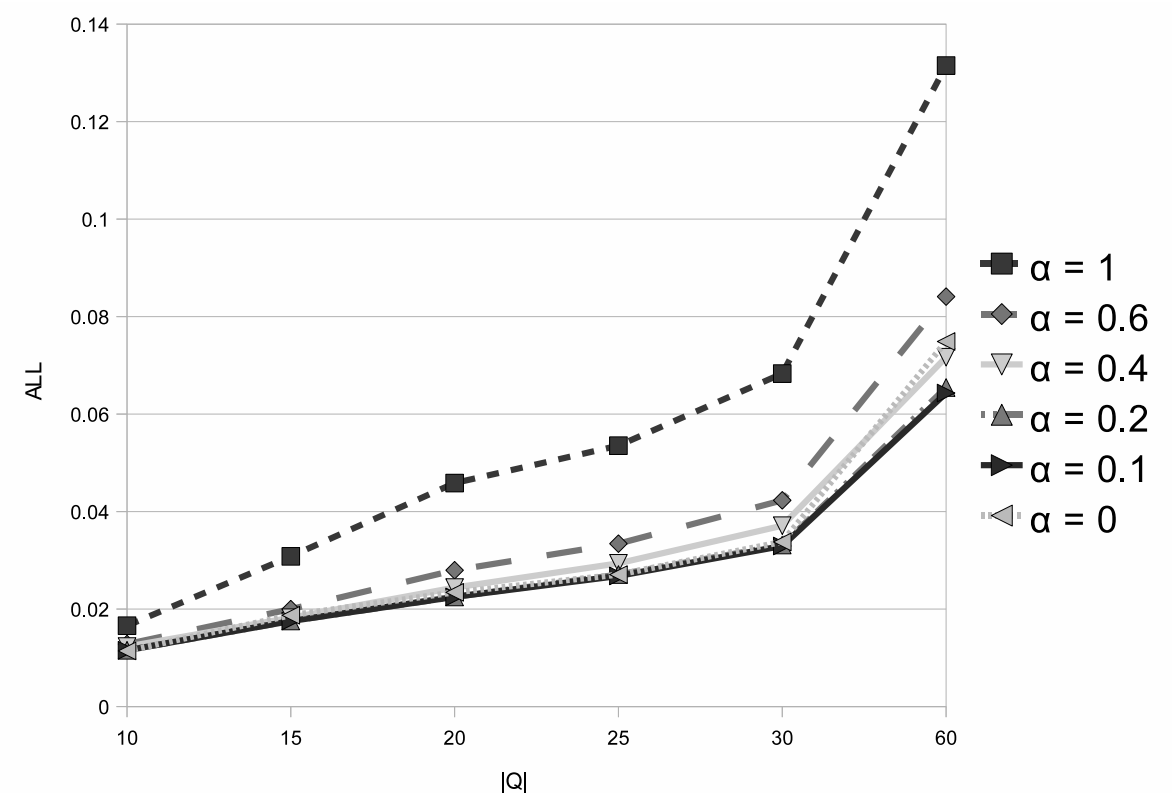


(c) ST3

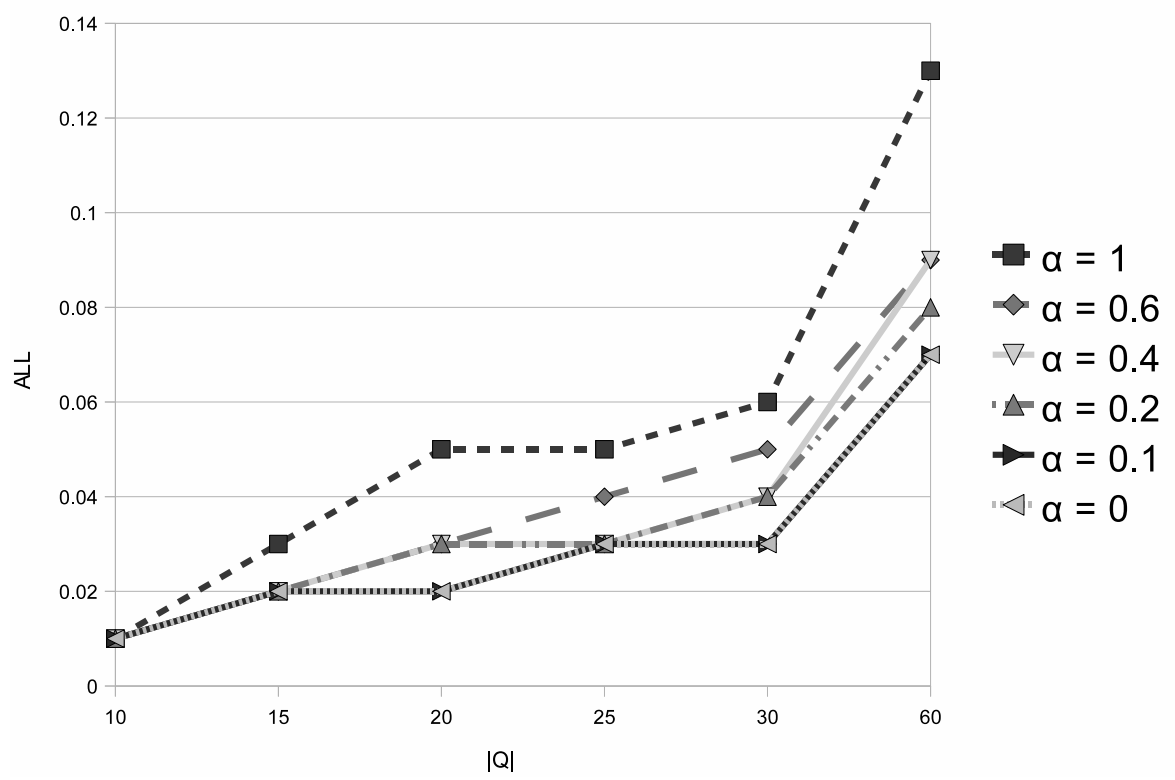


(d) ST4

Figure 5.3: The Spanning Tree Instances chosen for the network of Figure 5.2a.

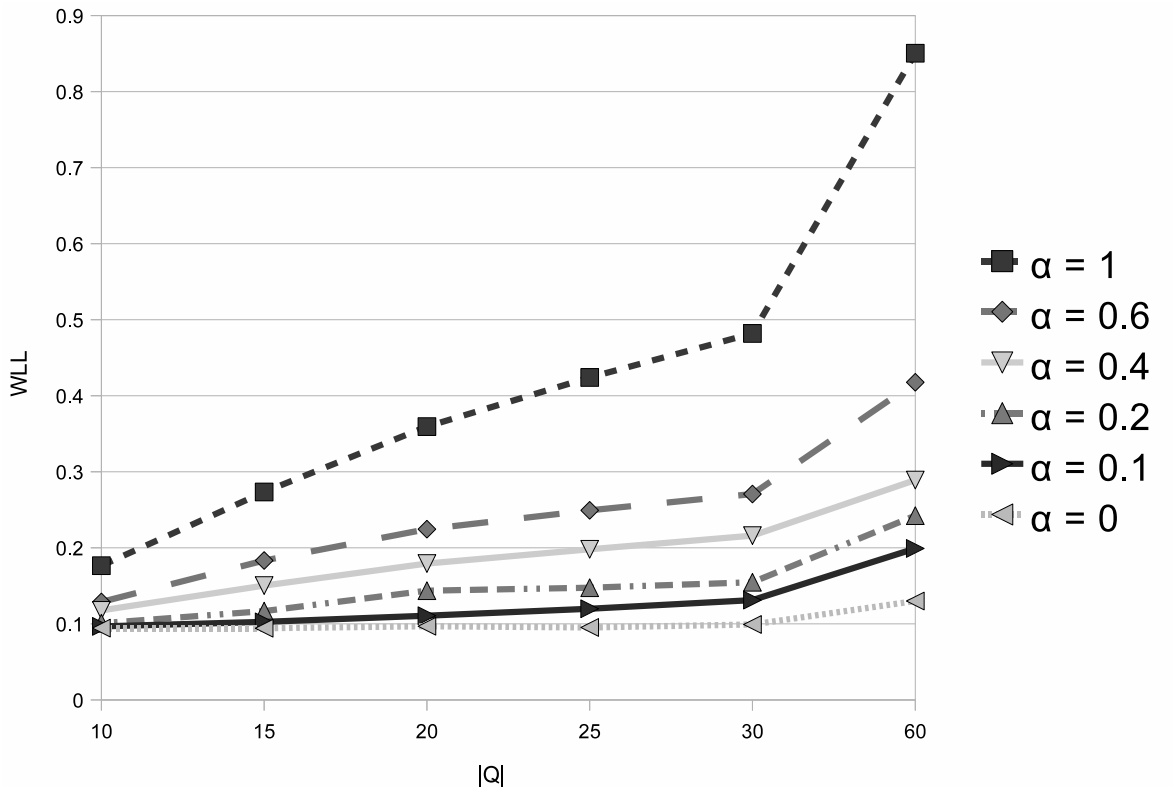


(a) Problems $P3a$

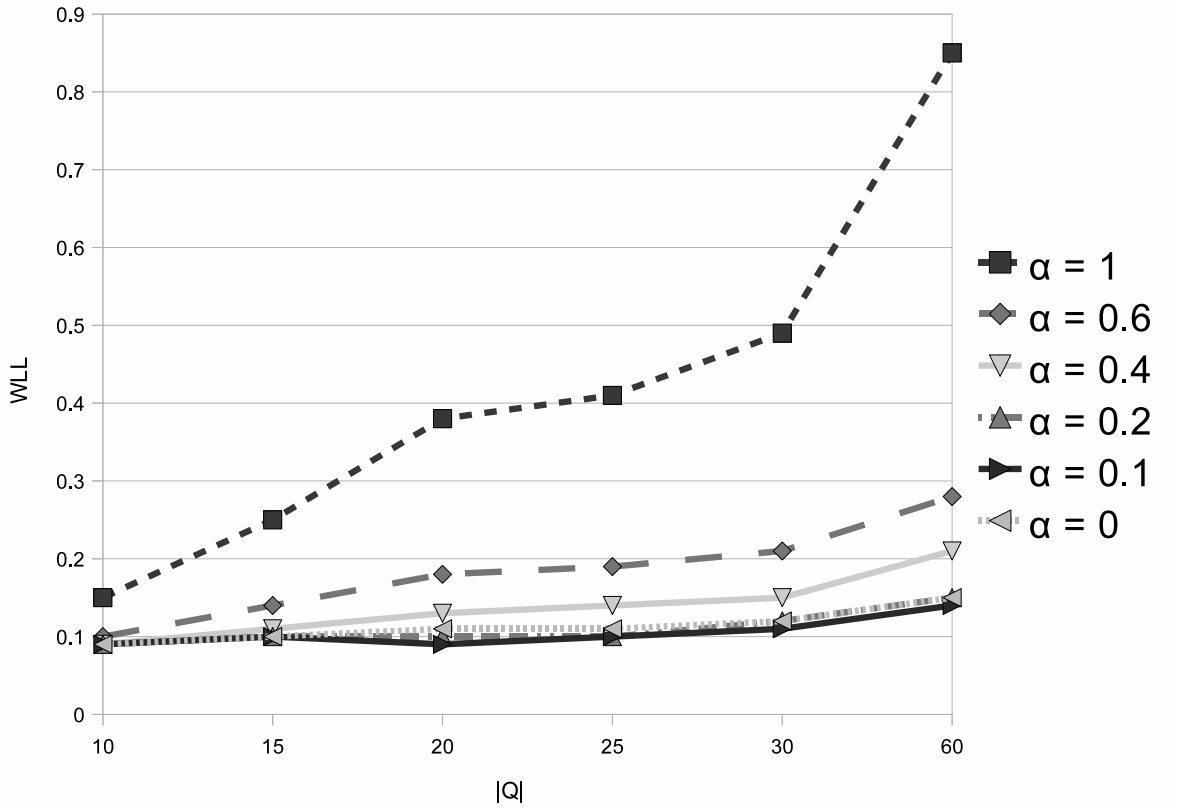


(b) Problem $P3b$

Figure 5.4: Average Link Load obtained for different traffic loads and different values of the parameter α when problems $P3a$ and $P3b$ are considered.

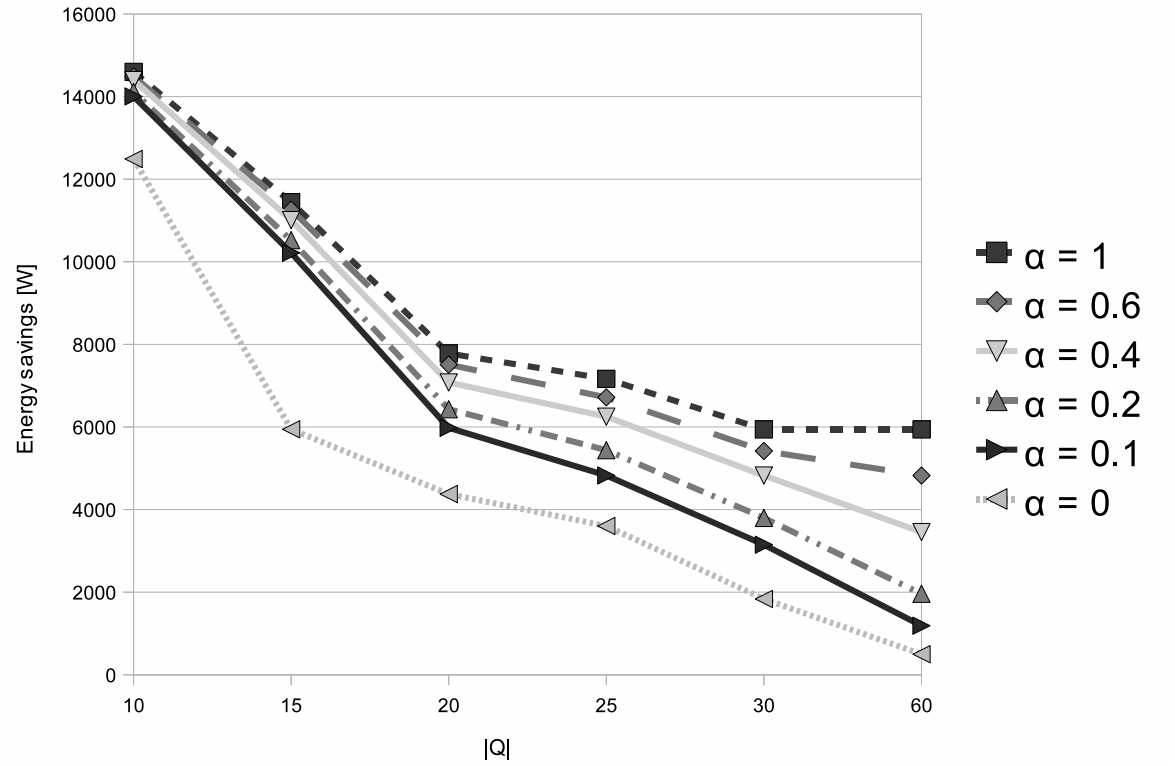


(a) Problems $P3a$

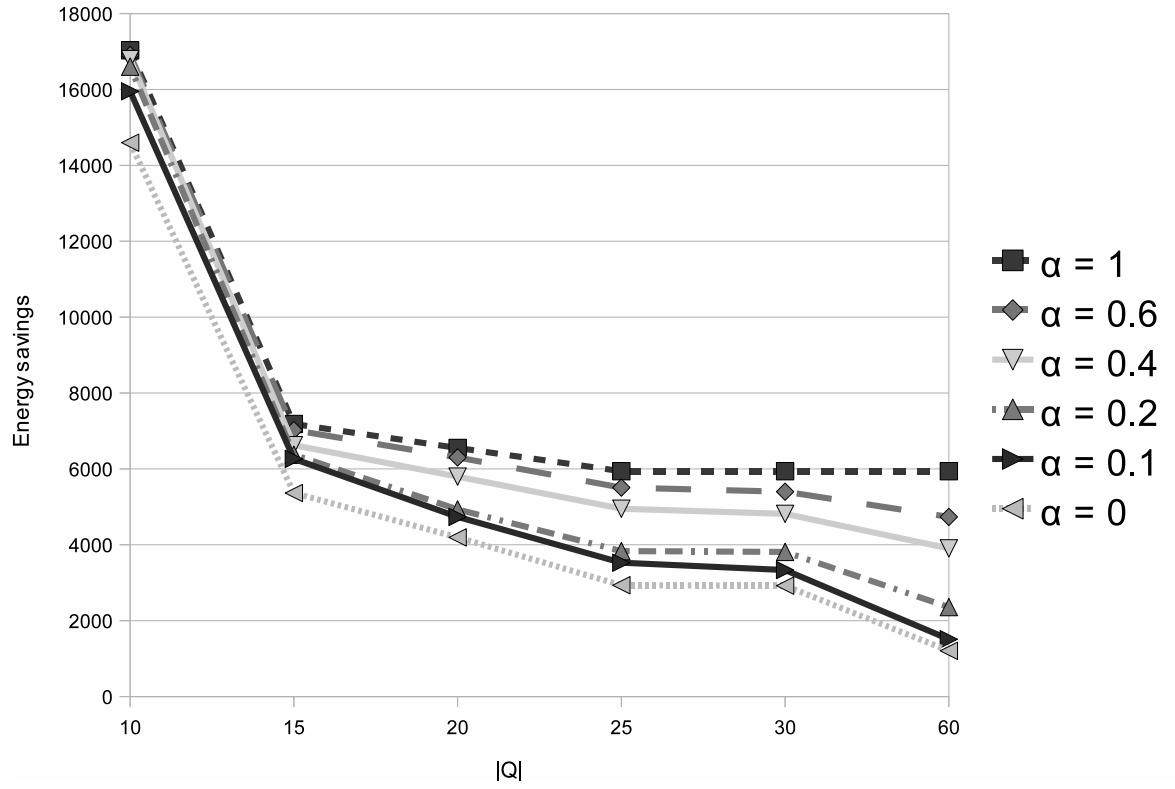


(b) Problem $P3b$

Figure 5.5: Worst Link Load obtained for different traffic loads and different values of the parameter α when problems $P3a$ and $P3b$ are considered.



(a) Problems $P3a$



(b) Problem $P3b$

Figure 5.6: Energy saving obtained for different traffic loads and different values of the parameter α when problems $P3a$ and $P3b$ are considered.

of different size, different traffic loads and a fixed value of the trade-off parameter $\alpha = 0.2$.

Table 5.3 compare the energy savings obtained from the three problems. The data in bold are the highest fraction of savings obtained for each kind of network between the two trade-off problems.

The problem formulation that allows better results in terms of energy savings is the $P3a$, although in most cases problem $P3b$ is just a few percentage points worse.

Table 5.4 shows the load balancing indexes. The data in bold are the lowest level of load for each kind of network.

Concerning the average load on the network the best results are obtained by solving the problem $P3a$, which just minimizes the ALL. This results, however, become less pronounced for larger networks. The problem $P3b$, which minimizes the WLL, instead provides the best performance regarding the maximum load in almost all cases.

The important thing that can be seen looking at Table 5.3 and Table 5.4 is that giving up a few percent of energy savings it is possible to achieve good results not only from the energy point of view but also in terms of load balancing.

In Table 5.6 the computational time necessary to solve the three different problems is compared. The time showed does not include the computational time required to solve the auxiliary problems for the trade-off problem. The data in bold are the highest computational time required for each kind of network. The data underlined are the ones that have reached the time limit of 1000 seconds.

Looking at Table 5.6 it is easy to notice that in the case of small networks (less than 15 nodes) the problems that require more time to be solved are the $P1$ and the $P3b$. For networks of size up to 30 nodes prevails the time required by the problem $P1$, while for even larger networks (50 nodes) is the problem $P3a$ to require the highest computational time. The standard deviation of the computational time varies between less than a second up to more than 400 seconds in the considered networks.

		Fraction of saving		
|N|	**|Q|**	$P1$	$P3a$	$P3b$
10	10	0.15	0.15	0.15
	15	0.08	**0.07**	0.06
	20	0.06	0.05	0.05
	25	0.05	0.04	0.04
	30	0.05	0.04	0.04
	60	0.05	0.02	0.02
15	10	0.27	0.26	**0.28**
	15	0.16	0.15	**0.19**
	20	0.13	**0.12**	0.11
	25	0.1	**0.09**	0.08
	30	0.09	**0.08**	0.07
	60	0.08	0.05	0.05
20	10	0.38	0.38	0.38
	15	0.28	**0.28**	0.27
	20	0.22	**0.22**	0.2
	25	0.17	**0.16**	0.15
	30	0.14	**0.13**	0.12
	60	0.11	**0.09**	0.08
30	10	0.51	**0.5**	0.49
	15	0.41	**0.4**	0.39
	20	0.33	0.33	0.33
	25	0.29	**0.28**	0.27
	30	0.24	**0.23**	0.22
	60	0.17	**0.22**	0.13
60	10	0.65	**0.65**	0.64
	15	0.58	**0.57**	0.55
	20	0.52	**0.51**	0.48
	25	0.47	**0.46**	0.44
	30	0.43	**0.42**	0.4
	60	0.29	0.28	0.28

Table 5.3: The fraction values of saving relative to total consumption of the network for the three different problems $P1$, $P3a$ and $P3b$.

		ALL			**WLL**		
|N|	**|Q|**	$P1$	$P3a$	$P3b$	$P1$	$P3a$	$P3b$
10	10	0.019	0.011	0.011	0.204	0.103	**0.087**
	15	0.029	**0.015**	0.018	0.262	0.109	**0.101**
	20	0.046	**0.023**	0.025	0.371	0.138	**0.096**
	25	0.062	**0.028**	0.030	0.492	0.155	**0.099**
	30	0.071	**0.034**	0.037	0.514	0.166	**0.113**
	60	0.139	**0.067**	0.077	0.867	0.253	**0.152**
15	10	0.009	0.006	**0.005**	0.184	0.094	**0.091**
	15	0.017	**0.009**	0.010	0.269	0.111	**0.094**
	20	0.027	**0.013**	0.014	0.334	0.133	**0.097**
	25	0.036	**0.016**	0.018	0.423	0.148	**0.098**
	30	0.044	**0.019**	0.021	0.526	0.166	**0.100**
	60	0.083	**0.038**	0.044	0.887	0.236	**0.142**
20	10	0.004	**0.003**	0.004	0.154	**0.095**	0.096
	15	0.008	0.005	0.005	0.227	0.103	**0.095**
	20	0.014	0.007	0.007	0.349	0.128	**0.098**
	25	0.022	**0.009**	0.010	0.436	0.146	**0.097**
	30	0.031	**0.011**	0.013	0.524	0.155	**0.098**
	60	0.058	**0.021**	0.027	0.879	0.230	**0.139**
30	10	0.002	0.002	0.002	0.144	0.095	**0.088**
	15	0.004	0.003	0.003	0.203	0.108	**0.106**
	20	0.007	**0.004**	0.005	0.266	0.115	**0.097**
	25	0.011	**0.005**	0.006	0.350	0.135	**0.107**
	30	0.015	0.007	0.007	0.477	0.164	**0.105**
	60	0.034	**0.013**	0.016	0.880	0.220	**0.130**
50	10	0.001	0.001	0.001	0.112	**0.090**	0.095
	15	0.002	0.001	0.001	0.180	0.104	**0.096**
	20	0.003	0.002	0.002	0.250	0.114	**0.099**
	25	0.004	0.003	**0.002**	0.273	0.122	**0.110**
	30	0.004	0.003	0.003	0.357	0.143	**0.117**
	60	0.014	**0.006**	0.007	0.835	0.232	**0.153**

Table 5.4: The Average Link Load (ALL) and the Worst Link Load (WLL) values of the three different problems $P1$, $P3a$ and $P3b$.

		Average number of MSTI		
|N|	**|Q|**	$P1$	$P3a$	$P3b$
10	10	3.6	4.53	4.53
	15	3.4	4.87	4.93
	20	2.67	5	5
	25	3.47	4.93	5
	30	2.67	5	4.93
	60	2.73	5	5
15	10	4	5	4.73
	15	4	4.93	5
	20	3	5	5
	25	2.13	5	5
	30	1.8	5	5
	60	1.93	5	5
20	10	4.6	4.87	4.67
	15	4.47	5	4.93
	20	4.07	5	5
	25	2.87	5	4.93
	30	2.27	5	5
	60	1.5	5	5
30	10	4.93	5	5
	15	4.87	5	5
	20	4.67	5	5
	25	3.53	5	5
	30	3.4	5	5
	60	1.93	5	5
60	10	4.93	4.93	5
	15	5	5	5
	20	5	5	5
	25	5	5	5
	30	4.93	5	5
	60	3.13	5	5

Table 5.5: Average number of MSTI activated for the three different problems $P1$, $P3a$ and $P3b$.

		Computational time [s]		
|N|	**|Q|**	*P*1	*P*3*a*	*P*3*b*
10	10	**1.84**	0.13	0.55
	15	**6.94**	0.31	6.2
	20	11.56	0.4	**34.34**
	25	16.37	0.42	**205.46**
	30	28.97	0.81	**520.36**
	60	176.88	2.4	**1000.83**
15	10	**2.68**	0.19	0.36
	15	**34.23**	0.66	9.91
	20	**339.89**	1.3	134.8
	25	580.24	2.12	**647.13**
	30	569.73	8.73	**828.76**
	60	**1000.04**	37.22	**1001.03**
20	10	0.63	0.2	**1.34**
	15	**8.17**	0.52	6.57
	20	**168.78**	1.34	155.46
	25	**793.33**	4.85	602.62
	30	**928.95**	10.94	891.81
	60	**999.91**	271.22	**1000.96**
30	10	0.42	0.55	**2.57**
	15	1.38	4.36	**6.4**
	20	**102.04**	2.36	41.63
	25	**397.99**	5.78	328.81
	30	**838.06**	94.73	735.86
	60	**999.91**	915.04	**1002.75**
50	10	1.91	**3.84**	1.94
	15	10.83	**21.14**	6.53
	20	27.81	**53.95**	15.14
	25	24.36	**90.43**	85.31
	30	88.79	**245.33**	196.46
	60	987.26	**1000.21**	**1002.6**

Table 5.6: The computational time necessary to solve the three different problems $P1$, $P3a$ and $P3b$.

Conclusions

This thesis addresses the problem of how to use the IEEE 802.1s Multiple Spanning Tree Protocol to improve energy consumption efficiency of Ethernet carrier networks.

After describing the network structure and how the power consumption is distributed within it, Chapter 1 analyzes the literature on sleeping techniques of network elements [22, 33, 34], ALR procedures [19, 20], energy cost-aware routing [32] and NGAN architecture.

Chapter 2 describes the evolution of the Ethernet standard and the Carrier Grade Ethernet technology, while Chapter 3 illustrated some previous proposals [11, 15, 17, 25, 26, 27, 30, 31, 38] on how to use Spanning Trees to perform Traffic Engineering in Metro Ethernet.

A new approach to this problem is presented in Chapter 4. The proposed approach computes the best subset of Spanning Trees and the best mapping of the traffic demands (VLANs) to them. The traffic flows are then routed over the network over paths defined by the STs the VLANs have been assigned to. This is performed in such a way that a portion of the network is forced by the objective function of the model to remain unused, thus making it possible to turn off the elements of that portion of the network, which are put into sleep mode to conserve energy. This leads to use the fewest possible nodes and arcs, while continuing to ensure full connectivity and the management of all requests.

A trade-off between the minimization of the energy consumption and the load balancing objective is also proposed, considering two different ways to manage the QoS indexes.

Finally, Chapter 5 shows the performance of the proposed model for different randomly generated networks, varying the network dimensions and the traffic load.

The experiments show that the energy-aware model presents good results in terms of energy consumption, but at the expense of a bad QoS (the utilization values of the links reaches nearly 0.9 in the worst cases).

However, it is interesting to note that through the models with the trade-off objectives, that take into account not only the minimization of the energy consumption but also the load balancing of the traffic, it is possible to achieve good results also from the load balancing point of view, giving up just a few percentage points of energy savings.

Future work includes the improvement of the proposed model by adopting exact methods to generate the Spanning Tree via column generation; the investigation of alternative solutions for load balancing; and an extensive experimental evaluation in real Metropolitan area network topologies and with real and more complex traffic matrices, to better assess the impact of the proposed solution on the network and on the energy consumption.

Bibliography

[1] https://www.cisco.com/en/US/prod/collateral/routers/ps5763/CRS-3_16-Slot_DS.html.

[2] http://www.cisco.com/en/US/prod/collateral/routers/ps5763/CRS-1x100GE_DS.html.

[3] IEEE Standard for Local and Metropolitan Area Networks. Media Access Control (MAC) Bridges. *IEEE Standard 802.1D-1998*, 1998.

[4] IEEE Standard for Local and Metropolitan Area Networks. Media Access Control (MAC) Bridges – Amendment 2: Rapid Reconfiguration. *IEEE Standard 802.1w-2001*, 2001.

[5] IEEE Standard for Local and Metropolitan Area Networks. Virtual Bridges Local Area Networks – Amendment 3: Multiple Spanning Trees. *IEEE Standard 802.1s-2002*, 2002.

[6] IEEE energy efficient ethernet (EEE). *IEEE Standard 802.3az*, September 2010.

[7] J. Baliga, R. Ayre, W. V. Sorin, K. Hinton, and R. S. Tucker. Energy consumption in access networks. In *OFC/NFOEC*, pages 1–3, February 2008.

[8] J. Baliga, K. Hinton, and R.S. Tucker. Energy consumption of the internet. In *Optical Internet, 2007 and the 2007 32nd Australian Conference on Optical Fibre Technology. COIN-ACOFT 2007. Joint International Conference on*, pages 1–3, June 2007.

[9] C. Bianco, F. Cucchietti, and G. Griffa. Energy consumption trends in the Next Generation Access Network - a telco perspective. In *Telecommunications Energy*

Conference, 2007. INTELEC 2007. 29th International, pages 737–742, September 2007.

[10] P. Briggs, R. Chundury, and J. Olsson. Carrier ethernet for mobile backhaul. *Communications Magazine, IEEE*, 48(10):94–100, October 2010.

[11] W. Chen, D. Jin, and L. Zeng. Design of Multiple Spanning Trees for traffic engineering in Metro Ethernet. In *Communication Technology, 2006. ICCT 2006. International Conference on*, pages 1–4, November 2006.

[12] K. Christensen, P. Gunaratne, B. Nordman, and A. George. The next frontier for communications networks: Power management. *Computer Communications*, 27(18):1758–1770, December 2004.

[13] J. D'Ambrosia. 100 gigabit ethernet and beyond. *Communications Magazine, IEEE*, 48(3):S6–S13, March 2010.

[14] A.F. De Sousa. Improving load balance and resilience of ethernet carrier networks with IEEE 802.1s multiple spanning tree protocol. In *Networking, international Conference on Systems and International Conference on Mobile Communications and Learning Technologies, 2006. ICN/ICONS/MCL 2006. International Conference on*, pages 95–95, April 2006.

[15] A.F. De Sousa and G. Soares. Improving load balancing and minimizing service disruption on ethernet networks using IEEE 802.s MSTP. In *EuroFGI Workshop on IP QoS and Traffic Control, IST Press*, pages 25–55, December 2007.

[16] T.M. Egyedi and M.H. Sherif. Standards dynamics through an innovation lens: Next-generation ethernet networks. *Communications Magazine, IEEE*, 48(10):166–171, October 2010.

[17] J. Farkas and Z. Arato. Performance analysis of shortest path bridging control protocols. In *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, pages 1–6, December 2009.

[18] K. Fouli and M. Maier. The road to carrier-grade ethernet. *Communications Magazine, IEEE*, 47(3):S30–S38, March 2009.

[19] C. Gunaratne, K. Christensen, and B. Nordman. Managing energy consumption costs in desktop pcs and lan switches with proxying, split tcp connections, and scaling of link speed. *International Journal of Network Management*, 15(5):297–310, September 2005.

[20] C. Gunaratne, K. Christensen, B. Nordman, and S. Suen. Reducing the energy consumption of ethernet with adaptive link rate (ALR). *Computers, IEEE Transactions on*, 57(4):448–461, April 2008.

[21] M. Gupta and S. Singh. Greening of the internet. In *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, SIGCOMM '03, pages 19–26, New York, NY, USA, August 2003. ACM.

[22] K. Ho and C. Cheung. Green distributed routing protocol for sleep coordination in wired core networks. In *Networked Computing (INC), 2010 6th International Conference on*, pages 1–6, May 2010.

[23] M. Huynh and P. Mohapatra. Metropolitan ethernet network: A move from lan to man. *Computer Networks*, 51(17):4867–4894, aug 2007.

[24] A. Kirstadter, C. Gruber, J. Riedl, and T. Bauschert. Carrier-grade ethernet for core networks. In *Optical Fiber Communication and the National Fiber Optic Engineers Conference, 2007. OFC/NFOEC 2007. Conference on*, pages 1–3, March 2007.

[25] A. Kolarov, B. Sengupta, and A. Iwata. Design of multiple reverse spanning trees in next generation of ethernet-VPNs. In *Global Telecommunications Conference, 2004. GLOBECOM 2004. IEEE*, volume 3, pages 1390–1395, November 2004.

[26] A. Meddeb. Multiple spanning tree generation and mapping algorithms for carrier class ethernets. In *Global Telecommunications Conference, 2006. GLOBECOM 2006. IEEE*, pages 1–5, December 2006.

[27] A. Meddeb. Smart spanning tree bridging for carrier ethernets. In *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, pages 1–5, December 2008.

[28] MEF. Century old utility provider moves in to the 21st century. `http://metroethernetforum.org/PDFs/CaseStudies/mef-luminous_case-study-idaho-falls-powerf.pdf`, 2004.

[29] MEF. Clarian health deploys a metro area E-LAN solution to support patient care. `http://metroethernetforum.org/PDFs/CaseStudies/mef-att-clarian-health-success-story.pdf`, 2004.

[30] G. Mirjalily, F.A. Sigari, and R. Saadat. Best multiple spanning tree in metro ethernet networks. In *Computer and Electrical Engineering, 2009. ICCEE 2009. Second International Conference on*, volume 2, pages 117–121, December 2009.

[31] M. Padmaraj, S. Nair, M. Marchetti, G. Chiruvolu, M. Ali, and A. Ge. Metro ethernet traffic engineering based on optimal multiple spanning trees. In *Wireless and Optical Communications Networks, 2005. WOCN 2005. Second IFIP International Conference on*, pages 568–572, March 2005.

[32] A. Qureshi, R. Weber, H. Balakrishnan, J. Guttag, and B. Maggs. Cutting the electric bill for internet-scale systems. In *Proc. of SIGCOMM'09*, August 2009.

[33] P. Reviriego, J. A. Hernandez, D. Larrabeiti, and J. A. Maestro. Performance evaluation of energy efficient ethernet. *IEEE Communications Letters*, 13(9):697–699, September 2009.

[34] P. Reviriego, J.A. Maestro, J.A. Hernàndez, and D. Larrabeiti. Burst transmission for energy-efficient ethernet. *Internet Computing, IEEE*, 14(4):50–57, July 2010.

[35] J. Roese, R.P. Braun, M. Tomizawa, and O. Ishida. Optical transport network evolving with 100 gigabit ethernet. *Communications Magazine, IEEE*, 48(3):S28–S34, March 2010.

[36] M. Roughan, A. Greenberg, C. Kalmanek, M. Rumsewicz, J. Yates, and Y. Zhang. Experience in measuring backbone traffic variability: models, metrics, measurements and meaning. In *IMW '02: Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment*, pages 91–92, New York, NY, USA, August 2002. ACM.

[37] R. Sanchez, L. Raptis, and K. Vaxevanakis. Ethernet as a carrier grade technology: developments and innovations. *Communications Magazine, IEEE*, 46(9):88–94, September 2008.

[38] D. Santos, A. de Sousa, F. Alvelos, M. Dzida, M. Pioro, and M. Zagozdzon. Traffic engineering of multiple spanning tree routing networks: the load balancing case. In *Next Generation Internet Networks, 2009. NGI '09*, pages 1–8, July 2009.

[39] H. Toyoda, G. Ono, and S. Nishimura. 100GbE PHY and MAC layer implementations. *Communications Magazine, IEEE*, 48(3):S41–S47, March 2010.

[40] G. Wellbrock and T.J. Xia. The road to 100G deployment. *Communications Magazine, IEEE*, 48(3):S14–S18, March 2010.