NTNU – Trondheim
Norwegian University of
Science and Technology

Economic analysis of QoS differentiation in OPS networks

Fredrik Grøndahl Leistad

Master of Science in Communication Technology
Submission date: August 2012
Supervisor: Harald Øverby, ITEM

Norwegian University of Science and Technology
Department of Telematics

Project Description

This project will analyze the economic aspects of QoS differentiation in Optical Packet Switched networks (OPS). The cost of introducing QoS differentiation in OPS should be studied as well as the required added revenue needed to compensate for the potential increased cost.

Prosjektbeskrivelse (Norsk)

I dette prosjektet analyseres de økonomiske aspekter ved Quality of Service (QoS) differensiering i Optiske Pakke Switchede (OPS) nettverk. Kostnaden ved å introdusere QoS differensiering i OPS blir studert så vel som den påkrevde økte omsetning nødvendig for å kompensere for en potensiell økt kostnad.

Abstract

Quality of Service (QoS) differentiation is the differentiated traffic-handling to achieve multiple service classes of varying quality. In addition to providing service guarantees necessary for delay-sensitive, real-time applications, service differentiation can provide increased income for network providers due to price differentiation opportunities. Moreover, with advances in technology, previous QoS schemes based on buffered networks cannot be used in newer, bufferless, optical networks. Current and future technologies were studied to facilitate economic analysis of QoS differentiation in Optical Packet Switched networks. Empirical data from related studies have been adopted to quantify a relationship between objective measurements of network quality and a user's willingness to pay for that quality. Models that represent network scenarios with and without service differentiation were discussed to address the viability of implementation. The model developed suggests pricing for customer classes based on network parameters before and after implementation, cost of deployment, customer willingness to pay, and business requirements such as Return on Investment. It was determined that under certain circumstances, a network provider may improve revenue via service differentiation.

Preface

This thesis is submitted in partial fulfillment of the requirements for the degree Master of Science (MSc) in the master program for Communication Technology at the Department of Telematics, Norwegian University of Science and Technology (NTNU). The presented work has been carried out in the period February 2012 - June 2012 at the Department of Telematics. I would like to thank my supervisor, Harald Øverby, for invaluable feedback and guidance throughout the project.

Many thanks to fellow students and family who motivated when I was stuck and proofread

Special thanks to my girlfriend Lisa, who proposed new ways of thinking about challenges and who listened passionately during conversations regarding my work.

Acronyms

ADSL Asymmetric Digital Subscriber Line.

APDP Adaptive Preemptive Drop Policy.

ATM Asynchronous Transfer Mode.

BE Best-Effort.

Bpl Packet-loss Robustness Factor.

BurstR Burst Ratio.

Diffserv Differentiated Services.

DWDM Dense Wavelength Division Multiplexing.

FCC Federal Communications Commission.

G-MPLS Generalized Multi-Protocol Label Switching.

GB Gigabyte.

IETF Internet Engineering Task Force.

Intserv Integretated Services.

IP Internet Protocol.

IPTV Internet Protocol television.

ISP Internet Service Provider.

ITU International Telecommunication Union.

LSP Label Switched Path.

Mb Megabit.

MOS Mean Opinion Score.

ms milisecond(s).

NLPRS Network Layer Packet Redundancy Scheme.

NP Network Provider.

O-E-O conversion Optical Electrical Optical conversion.

OBS Optical Burst Switching.

OCS Optical Circuit Switching.

OPS Network Provider.

OSI Open Systems Interconnection.

P-NLPRS Protected Network Layer Redundancy Scheme.

PDP Preemptive Drop Policy.

PESQ Perceptual Evaluation of Speech Quality.

PLR Packet Loss Ratio.

QoE Quality of Experience.

QoS Quality Of Service.

RAM Random Access Memory.

RMSE Root Mean Squared Error.

ROADM Reconfigurable Optical Add-Drop Multiplexer.

ROI Return on Investment.

SDH/SONET Synchronous Optical Networking/Synchronous Digital Hierarchy.

SLA Service Level Agreement.

SSE Sum of Squares Due to Error.

VoIP Voice over IP.

WA Wavelength Allocation.

WDM Wavelength Division Multiplexing.

WTP Willingness To Pay.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Demand for internet capacity has been rapidly increasing. Internet service providers are looking for new ways of increasing capacity and maintaining quality of service while increasing their revenues. One method of much debate is differentiated services, in which the provider guarantees various shares of bandwidth (hence, connection quality) for different prices.

Internet today provides Best Effort Service- the network attempts to process packets as quickly as possible, however there are no Quality Of Service (QoS) guarantees of data delivery. Numerous real-time applications like Voice over IP (VoIP)), Internet TV Internet Protocol television (IPTV) and online gaming require a minimum level of bandwidth as well as a maximum latency to function well. Due to the recent, wide adoption of Wavelength Division Multiplexing (WDM) networks, an important study concerns how service differentiation in WDM networks can be realized for carrying higher priority WDM traffic.

Implementing QoS differentiation is desired not only by users, but by providers, as it permits price differentiation of Internet services [7] which can lead to increased profits. Nevertheless, the deployment of QoS service differentiation comes at a cost. For a Network Provider (NP), the revenue increase from introducing price differentiation must more than cover the deployment cost in order to be attractive and worthwhile. Adaptive Preemptive Drop Policy (APDP) [53] is a method of service differentiation analyzed within this study and both the potential cost of implementing and the revenue required to cover the cost is discussed. The study is supplemented by a mapping between a users' Willingness To Pay (WTP) as a function of the quantitative QoS parameter Packet Loss Ratio (PLR). This relationship is then used to suggest a price for the high priority service class, as a factor of the low priority service class. Lastly, an overview of equipment needed to upgrade an existing network to support QoS differentiation is provided.

1.1 Methods

To achieve the anticipated outcome in this project, a list of methods is set up as follows:

- Obtain knowledge via literature review.
- Identify different types of Quality Of Service.
- Discuss three pricing strategies to most correctly price the service classes.
- Map a quantitative relationship between Quality of Experience (QoE) and Packet Loss Ratio (PLR).
- Analyze two economic scenarios: one with and one without Quality Of Service (QoS) differentiation.
- Compare the two scenarios and calculate a pricing strategy for differentiated services.
- Establish potential costs for deploying QoS differentiation.

1.2 History

With ever increasing throughput demand in the internet, optical networks have largely become the communication system of choice for core networks, mainly due to the advantages of increased speeds and the ability to transmit signals longer distances without the need for signal repetition. Though large parts of the networks have become optical, the switching must still be performed in the electric domain, since optical Random Access Memory (RAM) is still not reality [31]. As the conversion between the different domains is slow, and electronic packet switching has a much lower capacity than that offered by the optical network,

overall throughput in optical networks is limited by the need for electronic buffers. To avoid domain conversion in fully optical packet switched networks, fast optical switches with optical RAM are required. Optical RAM for these purposes is still far from becoming a reality. A hybrid circuit/packet switched approach like Optical Burst Switching (OBS), however, is dubbed as a possible candidate while fully optical packet switching remains in development [8, 5, 57].

Since OBS does not utilize buffers, how to support QoS differentiation at the Wavelength Division Multiplexing (WDM) layer has been the focus of much research [56, 12, 9]. An efficient service differentiation scheme which can provice QoS differentiation at the WDM layer, is Adaptive Preemptive Drop Policy (APDP) [53], which in this thesis is the service differentiation .

Important questions regarding the profitability of supporting QoS differentiation is the cost of deployment, including hardware and software costs, pricing structure and potential profits. Presented within this paper is a pricing strategy as well as a framework for calculating the profitability of QoS deployment.

1.3 Motivation

Although QoS would be beneficial to support real-time and business critical applications, it has never, to the writers' knowledge, been deployed at a large scale in core networks. Quality of Service is about *guaranteeing* service quality. It is about removing the probability that a network transfer could fail because of congestion or a broken network link. A best-effort network, no matter how over-provisioned, can not provide this guarantee [49]. Current needs include important Voice over

IP (VoIP) and video calls. For these services, using today's best-effort network, users experience sudden deteriorated quality or worse, frequent disconnections. Other areas which may benefit include remote live training with user interaction and latency sensitive online games. While even these needs can seem important enough to provide QoS, tomorrow's needs may include e.g a surgeon performing an operation with a robot remotely. Connection disconnects or even minor packet loss in such a case could be crucial to the success of the procedure and the safety of the patient. To manage the multitude of these applications, a network requires QoS in addition to best-effort service. The willingness to pay for these services already exists. For stock traders, miliseconds less delay can transfer to millions of dollars in profits/savings [16]. Professional online gamers may just need that little extra advantage to take home the big pot.

Most developed and adopted QoS mechanisms, (e.g Integretated Services (Intserv) and Differentiated Services (Diffserv)), are centered around the Internet layer (or Network layer in the Open Systems Interconnection (OSI) model) - more specifically, Internet Protocol (IP). These QoS schemes are based on packet switching and require a buffer to differentiate individual classes of traffic. The QoS schemes presented in this paper do not require a buffer and are hence possible to apply in the next-generation OBS networks. Employing service differentiation at a lower OSI-layer will provide better scalability due to less complex scheduling algorithms [56]. Furthermore, QoS at the WDM layer is necessary for carrying some WDM layer traffic like control and management traffic, which would benefit from a higher priority than ordinary traffic.

While some opponents of QoS (through deep packet inspection technologies) argue that over-provisioning of network capacity is more economically viable than QoS implementation [22], they can never guarantee a minimum of service quality the way QoS can. Additionally, while over-provisioning may be more affordable at the moment, if there is a market for QoS, the relative costs would be good to know. Businesses with critical network-based applications, such as large finance institutions, remote medical companies, as well as government groups and professional gamers [28], may all be willing to pay a premium for these guarantees. By better segmenting the market of potential users, a Network Provider (NP) can offer multiple services such that users are differentiated according to their desired service level and willingness to pay. Important questions regarding the profitability of QoS include the cost of deployment, pricing structure and potential profits.

1.4 Limitations

This paper has the following major limitations:

- The data used to determine QoE not 100% accurate. (This study focused on the QoE relative to the quality of the network based on packet loss ratio, while some of the data used in the analysis from another study was also affected by an introduced parameter of delay)
- Small sample size to calculate Packet-loss Robustness Factor (Bpl).
- When looking at change in traffic demand due to implementing QoS differentiation, a potential increase in traffic demand is not considered. The results received can therefore be viewed as worst case.
- When introducing a new service, a competitive response is expected, resulting in supply and demand changes for a service. These responses are unpredictable in nature, and are not accounted for in this study. It is possible that this may impact the results given in this paper.

1.5 Outline

The remainder of the paper is organized as follows: An overview of relevant technologies and discussion on which technologies are likely to be used in future optical networks is given in Chapter 2. Chapter 3 presents essential studies leveraged in this paper. Chapter 4 considers whether (and how much) a NP can potentially improve profits by upgrading from a single service to t least two service classes with QoS differentiation. In Chapter 5, a relationship between a user's willingness to pay for a network service and the quality of that service is established. The results obtained are used in chapter 6, where a suggested price factor is provided to price one service class in relation to the other. A brief overview of which factors do and do not affect cost of deployment is given in Chapter 7. Finally, concluding remarks and suggestions for future work are given in Chapter 8.

Chapter 2

Background

. This chapter presents the background and arguments for important optical networking technologies used in the following chapters. It also provides arguments for which technologies are likely to dominate the future, and therefore needs to be provisioned for.

2.1 Quality of Service Differentiation

Quality Of Service (QoS), as used in this paper, refers to the quantitative parameter describing the quality of a service whose purpose is the delivery of data packets.

When speaking of QoS differentiaton, it is meant: the ability to provide different priority to different users, applications, or data flows. QoS differentiation is assumed to be provided on per-class architecture. Traffic is grouped into two service classes and given appropriate network resource priority according to the service class. The use of only two service classes has been argued in [23], and are defined here as follows:

- *Guaranteed rate service*: The users of this service are given priority in admission to the system as long as there is available capacity. An alternative name for this type of service is High Priority service.

- *Best-Effort (BE) Service*: A network service where the network does not provide any guarantees of data delivery. BE-users share their given capacity equally and their bit rates and delivery time depends on current traffic load. An alternative name for this type of service is Low Priority service.

2.1.1 QoS Per-Class Architecture

In the per-class architecture, there are two models for QoS differentiation: relative QoS and absolute QoS. In the relative QoS model, the guarantees of one class are defined relatively in comparison to the other class(es). For example, a high-priority service class is guaranteed to experience a lower PLR than a low-priority service class, but no upper bound loss probability is provided. The actual experienced PLR therefore depends on the traffic. In the absolute QoS model, a bound for QoS parameters is provided for the service classes with guaranteed traffic. In the case of a two-class architecture, the remaining class is considered BE. This paper focuses on absolute QoS, as this can provide a hard guarantee for the PLR in a Guaranteed service class. Absolute guarantees are essential to support applications with delay and bandwidth constraints, such as multimedia and mission-critical applications.[58] Moreover, the absolute QoS can be desirable from an Internet Service Provider (ISP)'s perspective, as the guarantees can be passed on to the customer with or without a Service Level Agreement (SLA).

2.2 Why Bufferless Optical Networks?

The next generation of optical networks are likely to be bufferless. This section gives a brief history and expected future of Network Provider (OPS).

2.2.1 Evolution of OPS Networks

Due to the ever increasing demand for bandwidth, optical networks are increasingly becoming the transport medium of choice due to their increased single cable bandwidth over electrical transport mediums, lower cost and complexity [36, 52]

As discussed in section 2.4, research has shown that internet traffic is inherently “bursty” in nature. Circuit-switched architectures are not very suitable for supporting data traffic, mainly due to their low bandwidth utilization of data transmissions with a short duration relative to the set-up time [8]. Packet switching - a communication method which groups data into packets - is more effective since it allows sharing of available network bandwidth. A packet consists of control information and the user data (payload). In a typical packet switched network, a packet is buffered and queued at each intermediate network node as a packet traverses from source to destination. The Internet backbone is becoming more optical, but there is currently no technology that can store light for a sufficient amount of time (optical buffer), and consequently there are no fully optical switches available [31].

2.2.2 Point-to-Point WDM

A temporary solution, Point-to-Point WDM, performs switching in the electronic domain. This technique may be regarded as the first generation Optical Internet [52]. Fig. 2.1 represents one view of how optical switching networks may evolve. This entails converting the optical signal into the electronic domain after they are de-multiplexed, switching the electric signal by an electronic switching module, and converting the signal back to the optical domain. The final optical signal is multiplexed back to optical fibers. The conversion from an optical signal to an electronic and back to the optical domain is often called an Optical Electrical Optical conversion (O-E-O conversion). An O-E-O conversion severely limits the maximum achievable bandwidth as well as increases the cost of the switch. The reason O-E-O conversions are expensive has been the need for many discrete, single-function optical, components required for each O-E-O conversion. These components include lasers, modulators, wavelength lockers, detectors, attenuators, WDM multiplexers and de-multiplexers [21].

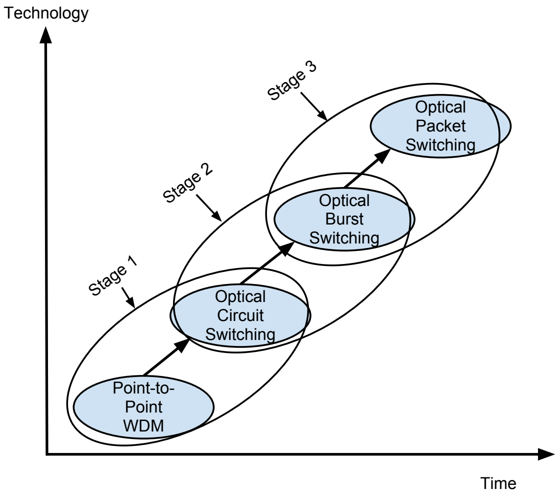


Figure 2.1: Potential evolution of optical switching techologies

2.2.3 Optical Circuit Switching

While waiting for integrated optoelectronic technology to advance to allow for optical RAM, Optical Circuit Switching (OCS), which does not require a buffer, could be implemented as an all-optical alternative. However, statistical multiplexing cannot be used to share resources (wavelengths), as a lightpath needs to be established from a source node to a destination node, reserving a dedicated wavelength on each node link on the physical path. Bandwidth will hence not be effectively utilized when there is little traffic transmitted over the reserved wavelength. Additionally, since the number of wavelengths available is limited, not every node can have a dedicated lightpath to every other node in the network. This would result in additional switching using O-E-O conversion to compensate, or the use of longer transmission paths.

2.2.4 Optical Burst Switching

To advance, multiple authors have suggested that OBS provides an attractive alternative [56, 58, 12]. OBS is feasible in the near future [8], and combines the best of the coarse-grained circuit-switching and the fine-grained packet-switching architecture while avoiding their shortcomings [8]. In OBS, data is aggregated into data bursts. A data burst is a train of packets moving together from an ingress node to an egress node, switched together at intermediate nodes [23]. A burst consists of a header, and data. The header, called a Control Burst, is transmitted separately in advance of the payload, which is called the Data Burst. When the Control Burst has been successfully transmitted, it reserves the bandwidth along a path for the corresponding Data Burst. By adjusting the offset time (time between control- and data-burst) at the source to be larger than the propagation time of the control packet, the need for intermediate buffering is eliminated [8].

Networks using OBS have already been deployed [15], and have shown that OBS may be key to next-generation optical networks. With OBS in mind, this paper focuses on QoS differentiation in bufferless optical packet switched networks.

2.2.5 Other hybrid optical network architectures

In addition to OBS, there have been several other all-optical network architectures proposed over the last few years. These approaches employ more than one switching paradigms and can be viewed as hybrid approaches. Examples include: Optical Burst Transport Network (OBTN), Overspill Routing in Optical Networks (ORION) and Optical Migration Capable Networks with Service Guarantees (OpMiGua) [46]. These approaches all require each node to contain at least one switching matrix, which establishes transparent optical lightpaths between input and output fibers [46]. Hence, regardless of which approach is adopted in the future, it is likely to be bufferless before technology has advanced enough to provide optical RAM for all-optical packet switching.

2.3 Existing QoS Schemes

Existing IP-based QoS schemes rely on packet switching and require buffers to isolate different traffic service classes [23]. The complexity of such schemes is usually high, and would require the use of an electronic buffer. Consequently, an O-E-O conversion is necessary to utilize these QoS schemes, compromising data transparency and rendering them ineffective for the prospected future bufferless OPS networks.

2.4 OPS Network Traffic Arrival Views

There has long been a divide in the research community about whether self-similarity or Poisson based arrival models are more suitable for modeling internet traffic. Earlier measurements suggested that the Internet traffic has a self-similar pattern [27, 43, 14], which means that the traffic appears bursty on many or all time scales . However, the suitability of these results were later challenged by arguing that traffic in newer networks were better modeled by traditional Poisson-based (more generically: Markovian) models when measuring over longer time intervals [24]. More recent studies based on more recent internet core traffic have confirmed that the Internet is losing its previous self-similar predominant nature [50].

It has been shown that there can be significant deviations in blocking probability (of a packet or "burst" of packets in a switched network) between different arrival models [41]. The traffic modeling presented in this thesis is described in chapter 6.1.

2.5 Packet Loss as Sole Measure of QoS

In a circuit or buffer-switched network, most packets will go through switches without being buffered. This translates to the largest contribution to the delay in the network coming by far from the transmission delay between switching points. The actual node delay time has been found negligible compared to the transmission delay [9]. End-to-end transmission time is likely to be dominated by transmission delay in (buffered) connecting networks, hence no service differentiation in the network architectures described here with respect to delay will be needed [6]. Packet loss, however, is a major concern. Some real-time services like VoIP, online gaming and video streaming require very low PLRs. The packet loss in OPS networks are often dominated by contention [55], as discussed in secton 2.6.

It is important to note that the packet size influences the IP PLR parameter. While a range of packets may be appropriate due to the variety of applications available, this is rather complex since the probability of losing a long/large packet is normally higher than losing a short/small packet [6]. For the purposes of this paper, the mean packet length is set to 472 bytes with cut-off at 40 and 1500 bytes [55].

2.5.1 Acceptable Packet Loss

The amount of packet loss which is acceptable for a consumer depends on the application. The International Telecommunication Union (ITU) recommends $1 * 10^-3$ in Y.1541 [47] as a bound on the network performance between user network interfaces. One should however note that this objective is partly based on studies showing that high quality voice applications and voice codecs will be essentially unaffected by 10^-3 PLR [47]. The ITU does provide a provisional recommendation with a smaller upper bound for PLR than $1 * 10^-3$, namely $1 * 10^-5$. However, the $1 * 10^-3$ PLR objective supports TCP with the limitations of widely deployed legacy settings, or assumes that some bottleneck will be encountered beyond the end-to-end user path. Since a hypothetical core network can not control the entire end-to-end network path, this paper follows the recommendation of $1 * 10^-3$ as an upper bound for acceptable packet loss.

This upper bound is given for end-to-end service, while the optical network architectures presented in this paper are more likely to be deployed only as a core network due to bandwidth and performance considerations. If a core OPS network were to fully exploit the ITU recommended upper bound, there would be no tolerance for packet loss for neighboring networks required for end-to-end connectivity. The surveys reviewed in Chapter 5 show that users are generally a bit more lenient to packet loss than the ITU recommendation.

2.6 Contention Resolution Schemes

In optical packet switched networks, contention occurs when a data packet destined for wavelength busy transmitting another packet. In electrical packet-switched networks, this is resolved by storing the packet losing the contention in electronic RAM. Due to the lack of optical RAM in current optical networks, alternative approaches for contention resolution is needed.

2.6.1 Wavelength Conversion

Wavelength conversion technique resolves contention by assigning bursts of traffic to a different wavelength in the same fiber, if the destination wavelength in the destination port is busy. This is done with the help of a wavelength converter. In the following chapters, wavelength conversion is assumed used in the network unless otherwise stated.

2.6.2 Fiber Delay Lines

Fiber delay lines provide sequential buffering by sending routing the contention losing optical packet through optical delay lines of different lengths [19]. However, in order to implement a large buffer capacity, large amounts of delay lines are required. Due to the sheer bulk of delay lines, they are generally regarded as impractical [6].

2.6.3 Deflection Routing

The Deflection routing contention resolution approach [11] transmits contending packets to other nodes than their preferred next-hop node. The packet is then routed to its destination using a (generally) less optimal route. How effective this approach is depends heavily on the network topology and the traffic pattern [55]. More importantly, deflection could cause packets to arrive with a significant delay, as well as out of order.

2.6.4 Intelligent Packet Loss Combating Mechanisms

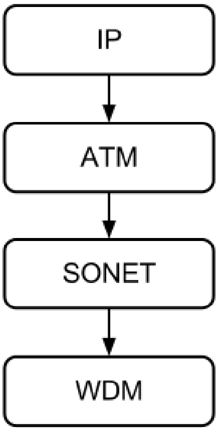
Packet loss combating mechanisms differ from contention resolution schemes previously presented in that they do not attempt to reduce the number of packet losses when a contention occurs [39]. Instead, they attempt to reduce the PLR by intelligent network behavior.

One such scheme is the Network Layer Packet Redundancy Scheme (NLPRS) [38] which has been shown to be a viable approach to reduce PLR to an acceptable level for asynchronous OPS [39]. The NLPRS has the added advantage that it can be extended to Protected Network Layer Redundancy Scheme (P-NLPRS) [40], see section 3.1.

2.7 WDM Layer (Traffic)

Wavelength Division Multiplexing (WDM) is technology which multiplexes a number of optical signals onto different wavelengths in a single optical fiber. This technique greatly increases the capacity of a strand of fiber. When a large number of channels are operating over a single optical fiber, the technique is generally referred to as Dense Wavelength Division Multiplexing (DWDM). WDM networks are being widely deployed in core networks. Using optical switches, a new optical layer has been introduced: The WDM layer is capable of supporting several different higher-layer services, such as Synchronous Optical Networking/Synchronous Digital Hierarchy (SDH/SONET) connections, Asynchronous Transfer Mode (ATM) virtual circuits, and IP -switched data traffic [59]. Currently, the typical protocol stack for optical networks is shown in Figure 2.7.

To understand why this protocol stack is in place, it is necessary to know



that SDH/SONET, multiplexing protocols for optical data transfer, were originally designed to support voice communications [36]. Due to its protocol neutrality, it is often chosen for transporting ATM cells, designed to handle both large amounts of network data traffic, as well as voice.

This protocol stack has several disadvantages [51]:

- Fully optical IP processing is not possible.
- 22% of the traffic in an IP over SDH/SONET over WDM network, is used for overhead.
- There is much redundant functionality in the layers.
- Different layers compete for protection in the event of failure.

The issues regarding protection and restoration are regarded as being one of the most pressing, necessitating a simplification of the protocol stack [29]. As IP networks data traffic has far exceeded that of voice traffic, it is possible to merge layers into a single layer to eliminate these disadvantages.

A move to the Generalized Multi-Protocol Label Switching (G-MPLS) architecture is foreseen, a control plane designed to encompass various switching techniques like time-division (SDH/SONET) and wavelength switching [36, 30], hereby providing a single routing control plane for these switching techniques.

A new, simplified protocol stack is shown in Figure 2.2.

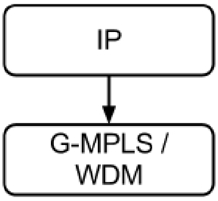


Figure 2.2: A simplified protocol stack for future optical netowrks

2.8 Network Survivability

A very important aspect of modern networks is network survivability: “The capability of a system to fulfill its mission, in a timely manner, in the presence of threats such as attacks or large-scale natural disasters.”[32]

Because of continuously increasing bit rates, an unrecovered network failure becomes a significant loss for ISP. Because of this, no ISP is likely to accept unprotected networks. As cable cuts are very frequent [59], physical redundancy as well as restoration protocols are required to reroute failed connections. The function of rerouting failed connections is referred to as *Restoration*.

Although well defined restoration techniques already exist in upper electronic layers like SDH/SONET, ATM, IP, these are slow relative to lower layers and may require intensive signaling [29]. By performing the restoration in the optical layer, the outage time can be decreased by exploiting fast rerouting of the failed connection.

Another technique for network survivability used in WDM networks is predesigned protection, where some resources like wavelengths or fiber strains are reserved for recovery from failures.

2.8.1 WDM Pre-designed Protection Techniques

There are two main protection techniques against single-link failures in optical WDM networks: Link protection and path-based protection.

With *link protection*, the nodes adjacent to the link failure reroutes the affected traffic around the failed link.

With *Path-based protection*, end-to-end paths are protected which is crucial to provide end-to-end guarantees in circuit-based switching. This is accomplished by transmitting two identical copies of each packet are transmitted from a source to a destination node over two node- and link disjoint paths.

Path- and link-restoration schemes have been extensively researched in circuit-switched transport networks. It has been shown that path-based protection mechanisms usually lead to better resource utilization than link protection [44, 59]. 1+1 and 1:1 path protection are amongst the most common path protection techniques [40].

1+1 path protection reserves a node and link-disjoint backup path and wavelength and transports the same data over both paths. Link-disjoint means that there are no common fiber cables used for the primary and backup path [40]. On the occurrence of either link or node failure on either path, the data will still arrive successfully at the destination through the remaining path. Compared to 1:1 protection, 1+1 protection provides instantaneous recovery as no switching is required between the primary and backup path. In 1:1 path protection, one dedicated backup path is preallocated for one primary path. At the event of failure, the failure must be localized, and a switch from the primary to the backup path must be done [3].

2.9 Usage Based Internet Access

This section provides a reasoning for the usage-based pricing scheme employed in section 4.2.1 and Chapter 6.

Most Internet Service Provider (ISP)s charge a (monthly) flat fee and differentiate the products based on the speed of the connection.The higher the speed, the more they charge. This means that there is little correlation between the amount a consumer uses their connection (bits transmitted) and the price they pay. For instance, for two selected Asymmetric Digital Subscriber Line (ADSL) products belonging to a major ISP in Norway, it is possible to provide an example of the more expensive product actually becoming the cheapest, given a sufficiently large amount of data transmitted and if looking at the price per transmitted bit. Given that traffic can be modeled with a Poisson process, see section 2.4, and that future optical access networks are expected to provide sustainable data rates with a passive subscriber distribution of (at least) 1:64 [18], it is safe to assume that very few users will constantly use their entire network bandwidth.

Table 2.1 presents an example with a fast, more expensive product transmitting 100 Gigabyte (GB), and a slower, more reasonable product transmitting 50GB of data.

Table 2.1: Example of ADSL prices

Speed	16 Mbps	5 Mbps
Price	399,- NOK	349,- NOK
Bits transmitted	819200 Mb (100GB)	409600 Mb (50GB)
Price/Megabit (Mb)	0.0004871	0.00085205
Time to transmit	14.22 hrs	22.76 hrs

Table 2.1 shows that if a user 1 transmits 100 GB, that user end up paying 57.16% less (per bit) than a user 2 who transfers 50 GB of data. This indicates that the more traffic you transmit, the less you pay per bit of data. This indirectly encourages the user to transmit as much as possible. However, if demand exceeds capacity, everyone will experience packet loss, and the service will be less enjoyable for all. These types of scenarios has been described as a Commonize Costs–Privatize Profits Game, where a group of individuals acting independently on their short-term interests, will over harvest a common resource.

On the contrary, metering traffic could encourage a minimalist data usage pattern, lowering total capacity usage for all, allowing for more users to divide available capacity and (in a perfect competition market) lower the price for all due to less required capacity upgrades from the ISP.

One reason why ISPs can provide a fixed pricing scheme is because the marginal cost of a single data packet is almost zero. There is not really a problem with this fixed pricing as long as bandwidth demands are below available capacity. However, it is only because of big technological advances like WDM that capacity has been able to keep up with demand [36].

2.9.1 Current Events

In the USA, there is a growing trend of changing billing practices to reflect how much data the customer is using.

Federal Communications Commission (FCC) chairman Julius Genachowski has long been saying that he supports usage-based pricing for Internet service to "help drive efficiency in the networks" [2].

Currently, many US based ISPs are implementing bandwidth caps, a step on the way towards full usage-based pricing[26, 45]. When analyzing potential profits and determining a fair price for a new product or service, it is beneficial to base the price on the amount the service has been used.

To enable an optimal analysis, price based costing is used in chapter 4 when analyzing how it can be profitable to implement two service classes instead of one.

2.10 Pricing Strategies

There are generally three approaches to establish a price for a good or service: Price-based costing and Competitive analysis. Because this paper considers introducing two new services simultaneously, the pricing of these becomes slightly more complex.

2.10.1 Cost-Based Pricing

This pricing strategy, also known as cost-plus pricing, considers how much the good being sold cost to make. By finding the break even point, i.e. the point at which cost or expenses and revenue are equal, a profit margin can be added to end up with the price. Because this paper introduces service differentiation, there are two prices to consider. These prices are related to each other in that they should priced according to their quality.

2.10.2 Price-Based Costing

This includes what the customers are willing to pay for the product, how much is it worth to them compared to other goods? If the customers are not willing to pay more than what it costs to make the product, it is probably not profitable to make the product. On the other hand, they may be willing to pay more than what would be generated by simply adding a profit margin.

In the scenario in this paper, it is natural to assume that the customers' willingness to pay more for an Internet service will depend on the perceived quality of that Internet service.

2.10.3 Competitive Analysis

This strategy involves looking at other companies and their prices for the same product. There are currently no commercial companies offering service differentiation in the manner presented here. It is then natural to look at substitutions. Given the vast number of ISPs and the multitude of countries and markets in which these ISPs compete in, a pricing strategy based on competitive analysis is considered beyond the scope of this paper.

Chapter 3

Quality of Service differentiation in OPS networks

This chapter covers the most central papers suggested for implementing QoS differentiation in this thesis.

A 1+1 path protection scheme for optical packet switched networks presents the Protected Network Layer Packet Redundancy Scheme (P-NLPRS), which provides 1+1 path protection in OPS networks. As optical networks are vulnerable to packet loss due to contention and link failure, a form of end-to-end path protection is required to guarantee a certain service provision. 1+1 path protection is simply two identical copies of every network packet transmitted from an ingress node to an egress node over two disjunct node- and link paths in a network.

3.1 A 1+1 Path Protection Scheme for Optical Packet Switched Networks

A quantitative study performed by the authors of [40] has proven that 1+1 path protection can be provide 33% to 91% of the cost of traditional 1+1 path protection. In a network with m paths, P-NLPRS achieves path protection by grouping $m - 1$ data packets, addressed to the same egress node, with one newly created redundancy packet at an OPS ingress node. These m packets are transmitted over m disjoint paths. If the the number of redundancy and data packets arriving are greater than the total of $m-1$ data packets, lost data packets can be reconstructed from successful data and redundancy packets using the FSRaid/NLPRS specification. Hence, P-NLPRS can provide functionality equal to 1+1 path protection if there is a single failure on a node or link on one of the m paths, also known as single element failure. As a side note, a degree of QoS can be achieved by adding redundancy packets to high priority traffic only.

Compared with traditional 1+1 path protection, in a network with 4 node and link-disjoint paths, between each node pair, P-NLPRS provides 1+1 path protection at only 33% of the cost. However in a more realistic network, analysis show that 1+1 path protection can be provided between 64% and 97% of the cost. The cost reduction depends on the node connectivity (Number of node- and link disjoint paths between the nodes in a node pair). The more connected the network is, the (generally) more cost savings can be achieved with the P-NLPRS.

3.2 Quality Of Service in Asynchronous Bufferless Optical Packet Switched Networks

In order to provide sufficient QoS for real time and critical applications, a form of service differentiation is needed. "Quality Of Service in Asynchronous Bufferless Optical Packet Switched Networks"[53] presents , an extension of Preemptive Drop Policy (PDP). In an OPS where contention resolution is not used and multiple packets are routed towards the same wavelength at the same time, one packet will be transmitted and the remaining packets will be discarded/lost. PDP is a method to provide service differentiation in an asynchronous bufferless OPS by means of preemption. Preemption allows for QoS differentiation by allowing high priority traffic to preempt low priority based on a probability p. By utilizing PDP, the remaining packets will be converted to idle wavelengths on the same fibre, if there are any. This provides superior PLRs to other QoS differentiation mechanisms like Wavelength Allocation (WA) and Intentional Packet Dropping (IPD), at a slightly increased cost of complexity [42]. Although [42] focuses on relative QoS guarantees, all preceding QoS differentiation schemes may be extended to provide absolute guarantees as well, as shown in [39] [58]. An extension of PDP is the Adaptive Preemptive Drop Policy (APDP). APDP can provide absolute QoS by adjusting the preemption probability if the PLR rises or falls past a predetermined QoS requirement. It is shown that APDP scales very well with the number of available wavelengths, which are expected to keep increasing as DWDM technology advances.

Chapter 4

Model Formulation

4.1 Implementing Service Guarantees

Consider a Network Provider (NP) that owns its own network resources and offers equal best-effort capacity to all its customers. The NP is considering implementing APDP in order to provide QoS differentiation (i.e. different service guarantees for different classifications of users). The service system has a finite processing capacity C which supports two services which will also be referred to as classes. It has been argued by the authors of [6], and in section 2.5, that the use of only two service classes are optimal in core networks. Supporting only two service classes reduces complexity of the control structure while still achieving the link utilization benefits of statistical multiplexing. One “Guaranteed Rate” service with minimum packet loss, is here denoted as class 1. A “BE” service class with medium to low packet loss and a low need for buffering, will be denoted as class 2. In the rest of this paper, various parameters will be tagged with subscripts 1 or 2 to denote the two respective classes. An important feature is that both service classes are delivered over common resources, i.e capacity is not split with a fraction dedicated to each service. For simplicity, the system assumes that the demand for capacity of class 1 users is less than that of the total capacity in the system at all times.

This is a reasonable assumption as a NP is not likely to sell service to more class 1 than it can guarantee service to under optimal conditions.

4.2 Model Formulation

When upgrading a network to include service differentiation, there is a nonrecurring cost related to upgrading the infrastructure and acquisition of additional software. This cost will here denoted as C_D. As the technique for providing QoS differentiation, APDP is designed to operate in an asynchronous optical bufferless network. The magnitude of C_D depends on the already existing network architecture. In order to retain a similar quality BE service, there may also be additional costs related to increasing network capacity. This cost will be denoted C_N. It is important to note that $C_D + C_N$ is a nonrecurring one-time investment. In the case that a company uses its capital to fund this cost, the company must weigh this investment towards other potential profits from other investments if the company decides to fund this themselves (e.g bank interest rate). In the case that the company requires a loan, there may be significant interest rates associated. This is all included in the $C_D + C_N$.

It should be clear that a business' revenue before deployment of service differentiation, R', must increase with more than the total cost, $C_D + C_N$, in order to obtain a profit from implementing service differentiation i.e: ($\delta R > C_D + C_N$). The new revenue after implementing service differentiation is denoted by R. The revenue growth must be a direct consequence of deploying service differentiation. This equates to that the revenue after implementing service differentiation must be larger than the costs $C_D + C_N$ as well as the non-differentiated service revenue, R', in order for the investment to be worthwhile:

$$R > R' + C_D + C_N \tag{4.1}$$

It is the goal of the service provider to obtain maximum revenues by pricing the two service classes optimally.

4.2.1 A Usage Based Pricing Scheme

As elaborated on in section 2.2, this approach can provide both the NP and its customers a price based directly on the actual use. If a more traditional fixed pricing scheme is desired by an NP, usage statistics from end-users in each service class can be collected averaged to set a base for fixed price.

In a network without service differentiation, a NP provides a single service class, denoted as S' to its customers with a certain guarantee of quality as defined in an SLA. In a scenario where the user pays a price per bit transmitted, each bit, denoted as q', has a price $p' > 0$. The NP's total revenue from the non-differentiated service is then the number of bits sold multiplied by the price:

$$R' = q' * p'$$

Now, that NP considers two service classes, class 1 which is a G-type class and class 2, a BE-type class. I.e class 1 traffic is given priority over class 2 traffic. Each bit is denoted by q_i and price per bit denoted by p_i, where the subscript $(_i)$ denotes the class.

Revenue is now given by:

$$R = q_1 * p_1 + q_2 * p_2 \tag{4.2}$$

For the NP to obtain a profit:

$$q_1 * p_1 + q_2 * p_2 > C_D + C_N + q' * p' \tag{4.3}$$

Under the assumption of perfect competition, supply is determined by marginal cost. Firms will produce additional output as long as the cost of producing an extra unit of output is less than the price they will receive.

When introducing service differentiation, there is a throughput penalty due to increased packet loss [42, 53]. This could result in both service classes being affected by a greater PLR than in a single-service scenario.

Chapter 5

Quality of Experience

This chapter will quantify a relationship between willingness to pay and quality as perceived by users.

5.1 Willingness to Pay

In order to price a product optimally, it is valuable to know what a potential customer is willing to pay for the product. This must be weighed towards the cost of making the product, in order to see if the product can become profitable. The Willingness To Pay (WTP) can be used to more accurately establish a margin on top of the cost, as opposed to simply adding a guesstimated margin.

Looking at micro-economic theory, the basic law of demand states that if demand increases and supply remains unchanged, then it leads to higher equilibrium price and higher quantity. If demand decreases and supply remains unchanged, then it leads to lower equilibrium price and lower quantity. The price one is willing to pay also depends on the *utility*, where an increase in utility results in an increased demand [10]. *Utility* represents a consumers satisfaction when consuming/using a good or service. Assuming an increase in QoS results in an increased utility, it can be assumed that a customer is willing to pay more for quality, disregarding other factors (e.g income, utility of unrelated goods). In the scenario in this paper, it is natural to assume that the customer is willing to pay more for an Internet service depending on the perceived quality of that Internet service.

5.2 Quality of Experience

To measure the perceived quality of an internet service, and therefore WTP, one may use Quality of Experience (QoE). QoE is a subjective measure of a customer's experience with a service or application. QoE is defined by the ITU as "The overall acceptability of an application or service, as perceived subjectively by the end-user." [37].

Although QoE is by nature subjective, there are several methods to quantify it. This paper will analyze two recent subjective studies, as well as make use of a mathematical model to quantify a user's QoE with regards to PLR. The Mean Opinion Score (MOS) is a test recommended by the ITU to obtain the human user's view of the quality of the network. It is widely accepted as the standard for evaluating speech quality. The MOS test procedures consist of users rating speech quality in a five-grade scale, from 1 (lowest) to 5 (highest). See table 5.1.

Table 5.1: Mean Opinion Score Classification

Values	Mean Opinion Score
5	Very Satisfied
4	Satisfied
3	Fair
2	Dissatisfied
1	Not Recommended

The IQX hypothesis (exponential interdependency of quality of experience and quality of service) presented in [17] proposes a generic formula where QoE and QoS parameters are connected through an exponential relationship. They state that QoE can be expressed as a function of n influence factors I_j:

$$QoE = \Phi(I_1, I_2, ..., I_n)$$

Further, they focus on the single influence factor $I = QoS$ to derive the relationship $QoE = f(QoS)$. As discussed in section 2.5, PLR will be used as the sole parameter for measuring QoS. This paper will focus on a subset of this, the relationship $QoE = f(PLR)$. This relationship will then be used to portray a person's willingness to pay for a service, $WTP = f(PLR)$.

5.3 QoE-PLR Relationship

In general, it seems reasonable to assume that a user's sensibility to quality is more perceivable the higher the QoE already is. A very small decrease in quality (PLR) will strongly decrease the QoE at a very high or almost perfect QoE (in this case, connection quality) because it is more noticeable. On the contrary, if the QoE is already very low, this miniscule decrease of quality will go almost unnoticed, since the QoE is already so low. An analogy to this is that of a new, spotless car versus that of a rusted, dented old car. A minor scratch on the new car could largely decrease its value, whereas it might not even get noticed on the dented old car.

5.4 Subjective Studies

This section presents two papers where empirical data is extracted and fitted functions to. In section 5.5, these functions are used to find a generic relationship between QoE and PLR.

5.4.1 A Pilot Study to Assess QoE

"A Pilot Study to Assess Quality of Experience Based on Varying Network Parameters and User Behaviour" [33] assesses QoE for users who are web surfing in a controlled environment. It is interesting both because they conduct the test at a relatively high PLR of 5-10% and because it's the test subjects are current internet users (2011), with their ensuing demands. Only the empirical data will be used from this study, as their proposed analytical model is both based on network delay measured in milisecond(s) (ms), as opposed to solely PLR, and largely untested. Test subjects were asked to evaluate their experiences while browsing at five minute intervals and placing a score according to the MOS scale, as described in table 5.1.

Figure 5.1 shows obtained MOS values dependent on the measured PLR for the conducted experiments in [33] and for the fitted equation $a * exp(-b * plr) + c$. Each point represents a user evaluation of their experience measured in a discrete MOS value for a given packet loss probability. The experiment was conducted with a limited discrete set of PLRs, several test subjects have given the same score for multiple PLR values. Consequently, one plotted point may represent overlapping points portraying more than one test subject's experience at a given PLR.

As PLR approaches 1, the QoE in terms of MOS approaches its minimum of one. A point representing the minimum QoE vs the maximum PLR has been added to represent that a customer will not be satisfied (MOS = 1) with 100% PLR (i.e no internet service).

The unknown parameters in the model function $a*exp(-b*plr)+c$, are retrieved by means of nonlinear regression. The optimization toolbox of Matlab was used to obtain an optimal fitting function. The R-square can be strong indicator of a good fit if it covers the set of data, hence the fitting function with the highest R-square value was selected (R-square value closest to 1). R-square is defined as "the ratio of the sum of squares of the regression", indicating how large of a proportion of variance is accounted for by the model. A value of 1 means that the fit explains 100% of the total variation in the data about the average, while a value of 0 means that the fit explains 0% of the total variation about the average. Equation 5.1 shows the obtained fit for the data in [33], graphed in Figure 5.1.

$$4.17 * exp(-11.35 * x) + 0.9374 \tag{5.1}$$

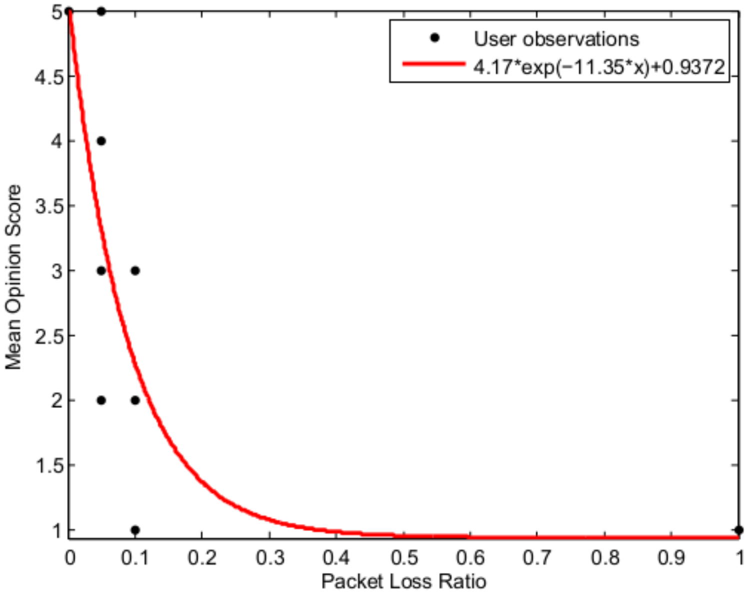


Figure 5.1: MOS as a function of PLR from [33]

The goodness of fit can be measured using several different metrics. The Sum of Squares Due to Error (SSE) measures the total deviation from the response values of the fit to the response values. It should approach zero to indicate that the model has less random errors. Root Mean Squared Error (RMSE) is an estimate of the deviation of the random component in the data. It should also approach zero to make for a fit more useful for prediction. The R-square measures how large of a proportion of variance is accounted for by the model. A value of 1 means that the fit explains 100% of the total variation in the data about the average, while a value of 0 means that the fit explains 0% of the total variation about the average.

Equation 5.1 yields SSE $= 14.95$, RMSE $= 0.8063$ and R-square $= 0.5022$. This is not a particularly good fit from an objective viewpoint. This could be due to a limited sample size as well as both the ignored latency and limited range of PLR. On the other hand, it is reasonable that the obtained, fitted equation outputs data that could very well have come from a subjective test. That a PLR of about 30%

during web surfing equating an MOS score of 1, "not acceptable", is plausible, especially when viewing from a perspective of WTP. A PLR of 30% would result in many site requests timing out, and large latencies on those that do not. The claim that few or no people are willing to pay for this, especially considering there are many other products available, seems very accurate according to MOS results.

Limitations

This paper has a sample size of only 12 participants, making the margin of error large. Furthermore, the study partly relies on inducing a delay of up to 100ms in addition to inducing PLR. This delay has been ignored in this paper and will theoretically result in a slightly more concave curve than what is shown in Figure 5.1. Additionally, the paper tests for a narrow, limited PLR.

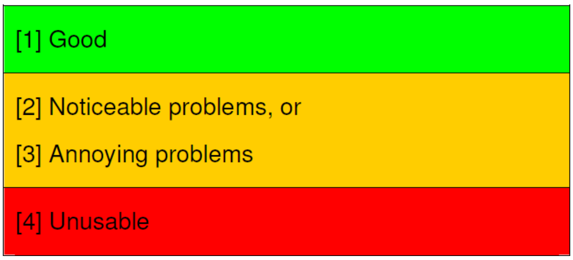
5.4.2 Assessing Network Quality of Experience

Assessing Network Quality of Experience [25] attempts to find a relationship between the two domains QoE and QoS.

They test a small but diverse range of popular Internet applications at various induced PLRs and latencies. This section will only consider the PLR portion.

Figure 5.1 shows obtained MOS values dependent on the measured PLR for the conducted experiments in [33] and for the fitted equation $a * exp(-b * plr) + c$. Each point represents a user evaluation of their experience measured in a discrete MOS value for a given packet loss probability. The experiment was conducted with a limited discrete set of PLRs, several test subjects have given the same score for multiple PLR values. Consequently, one plotted point may represent overlapping points portraying more than one test subject's experience at a given PLR

In addition to observing and recording the behavior of the application at different PLRs, a panel of expert evaluators gave their views of the experience using a generic 'traffic light' scale, see Figure 5.2. Each observation is marked as a point in Figure 5.3. As the experiment was conducted with a limited discrete set of PLRs and a wide range of applications, several applications have been given the same QoS score for corresponding PLR values. Consequently, one plotted point may represent overlapping points portraying than one application's evaluated user experience at a given PLR.



[1] Good

[2] Noticeable problems, or

[3] Annoying problems

[4] Unusable

Figure 5.2: "Traffic light" scale from [25]

The unknown parameters in the model function $a*exp(-b*plr)+c$, are retrieved by means of nonlinear regression. As before, the optimization toolbox of Matlab was used to obtain an optimal fitting function with the R-square value closest to 1. Equation 5.2 shows the obtained fit for the data in [25], graphed in Figure 5.3:

$$1.942 * exp(-16.98 * x) + 1.014 \tag{5.2}$$

Equation 5.2 yields SSE = 19.6, RMSE = 0.6261 and R-square = 0.2429. This is not a particularly good fit either, from an objective viewpoint. This could be due to a limited sample size as well, but perhaps more importantly, the limited range of tested PLRs. There is a large frequency of the highest score, 3, in the data, likely the result of the tests being conducted at too low PLRs. This could translate to a much higher density of scores at the extreme scores, resulting in a worse fit. This aside, the curve does resemble that of Figure 5.1 with QoE reaching the lowest score (unusable) at a PLR around 0.3. A PLR of 30% would result in most services being slow, and many services being completely unusable. The claim that few or no people are willing to pay for a service of this quality, especially considering there are many other products available, is supported. Furthermore, Figure 5.3 decreases faster per unit of PLR compared to that of Figure 5.1. This is as expected, because some of the applications which observations are plotted

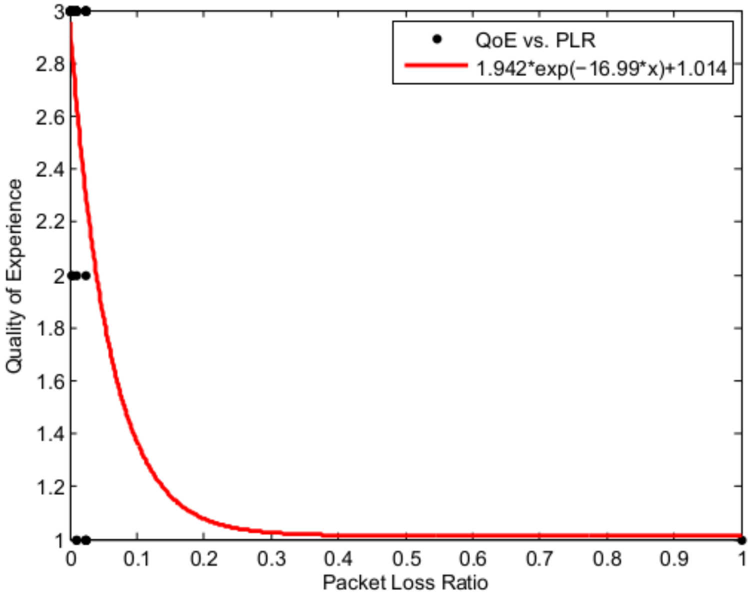


Figure 5.3: QoE as a function of PLR according to [25]

from the second survey in Figure 5.3 are more QoS sensitive than those from the first survey, plotted in Figure 5.1.

5.5 Objective Methods

Objective methods utilize algorithms and formulas that model or measure the QoE quantitatively, automatically and with repeatable results. These results highly correlate with subjective MOS scores when measured against each other. There is a considerable amount of research on voice quality prediction based on objective methods, split into two subgroups: Intrusive vs non-intrusive.

Intrusive vs. Non-Intrusive

Intrusive (signal based) methods compare the original (voice) signal to that which has been sent/processed through the network to estimate final (voice) quality. As a result, it is difficult to expand on these methods to include non-voice based applications, Therefore, these methods are unsuitable for generalizing QoE.

Non-intrusive methods like ITU's E-Model [4] and the IQX hypothesis [17] are based on network/application parameters.

The e-model is found to correlate well with subjective MOS scores [20], and while the Internet Engineering Task Force (IETF) recommends a more up-to-date method defined in RFC3611, the improvement lies in the calculation of the equipment impairment factor. This input variable is set to 0 for the purposes of this paper, rendering the updated E-model equal to the one recommended in [4].

5.5.1 E-Model

The E-model is a mathematical model which uses quantitative, measurable parameters to estimate quality. It has proven useful as a transmission planning tool and can be used by transmission planners to help ensure that users' will be satisfied by end-to-end transmission performance [4]. By using the e-model, it is possible to provide a prediction of the expected quality as perceived by the user.

The output of the e-model is a scalar quality rating value, R, which ranges from 0-100 and has a direct correlation with the overall quality. An estimated MOS can be calculated from the R-value, which, to allow comparison to the fitted models, have been done in this paper. It should be noted that the maximum obtainable 'R' value is 93.2, which only translates to a MOS of 4.41. This number is proposed by the ITU to convey the reduced quality predictions under ideal conditions. It has been shown that test subjects are averse to voting a perfect MOS score of 5 even under ideal conditions. Rather, statistics show that the average upper bound MOS for a subjective test is around 4.5 [13], which serves as an upper bound in the e-model.

The E-model is intended for estimating voice quality, but can be manipulated to estimate performance of a wide variety of applications by changing the input parameter Packet-loss Robustness Factor (Bpl). Bpl is defined as "The robustness of a codec to random packet loss" but could just as well describe the robustness of an application to packet loss. The model has the added advantage that by setting most/all input parameters to 0, PLR can be isolated so that $MOS = f(PLR)$. Additionally, it can model both bursty and random packet loss by increasing the burst parameter the Burst Ratio (BurstR). The Bpl parameter is normally found by using an intrusive QoE method, Perceptual Evaluation of Speech Quality (PESQ), to produce a set of MOS scores corresponding to packet loss rates ranging from 0 to 20%. The resulting PLR vs MOS curve is used as a reference model, and Bpl parameters are adjusted in the E-model until the output closely resembles that generated from the PESQ scores [1].

Likewise, this can be done with other applications and actual MOS scores as well. The objective here is to find a Bpl suitable to represent a wide enough range of applications such that calculating user QoE with the E-model and that the estimated Bpl can represent the users' QoE of a network as a whole. Subjective empirical data will be used to find an accurate Bpl for a wide range of applications.

The E-Model is modeled by the following formula at the given boundary conditions, Figure 5.4 shows the E-model with default parameters, and Bpl = 0.

$$R = 93.2 - (95 * (plr/(plr/7)))$$

$$WTB = \begin{cases} \text{For } R < 0: & WTP = 1 \\ \text{For } R > 100 & WTP = 100 \\ \text{For} 0 < R < 100: & WTP = 1 + (0.035) * R + (.000007) * R * (R - 60) * (100 - R) \end{cases}$$

To make the E-models output more closely resemble those of Figure 5.1 and Figure 5.3, the Bpl is adjusted to 6. This is shown in Figure 5.5.

With the Bpl set to 6, it can be seen that the curve resembles that of Figure 5.3 and 5.1, belonging to the empirical data. This model shows a maximum close to 4.5, as opposed to the empirical data. This is due to the empirical data studies used a discrete scale in their surveys, rather than continuous data as recommended by the ITU. The MOS reaches the minimum score of 1 at around 30% PLR. At 30% PLR, even browsing the web would be slow and frequently time out. Many other QoS sensitive applications like online gaming could be completely unusable.

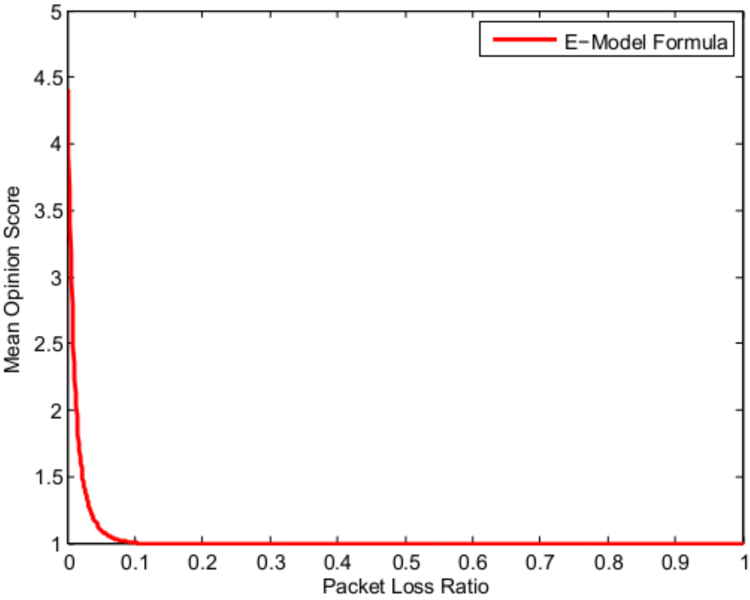


Figure 5.4: MOS as a function of PLR based on the E-model with Bpl = 0

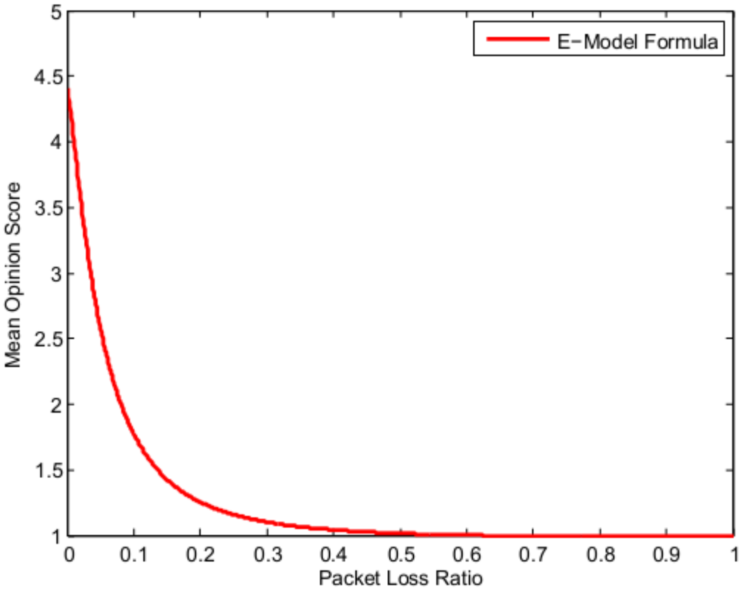


Figure 5.5: E-Model with Bpl = 6.

It is unlikely that users' would be satisfied in any applications, all in all making this model plausible.

5.5.2 Limitations

The E-model cannot judge the impact of variable packet delay and has limited tested and confirmed accuracy with non-random packet losses.

5.6 Mapping MOS to Willingness To Pay

The MOS score by itself does not reveal how much more (or less) a customer is willing to pay for a product. By rescaling the E-model formula with a set $bpl = 6$ from a MOS scale to a QoE scale of 0 to 1, the WTP factor can be expressed as a function of packet loss, i.e: $WTP = f(PLR)$.

As this will be used to calculate the profitability of two service classes and compare that to a single-service network, the WTP price can be quantified as a percentage of the maximum price a customer is willing to pay at perfect service/zero packet loss (or a price a company sets to recover costs of providing a service), P_{max}. An absolute number is not required for this purpose, as both the price before and after QoS differentiation deployment can be written as a fraction of P_{max}.

$$P_{WTP} = P_{max} * WTP(plr)$$

By converting the MOS scale of from 1-4.5 to a QoE of 0-1 scale, mapping of Willing to Pay vs PLR can be seen in See Figure 5.6. The price is now mapped as a percentage a customer is willing to pay correlated to the maximum price they are willing to pay.

Further, by changing the scales of Figure 5.3 and 5.1 from their respective scales to a common scale of 0-1, all plots can be viewed in the same figure for comparison. As shown in Figure 5.6, WTP(PLR) is very closely fitted to the plots of the empirical data.

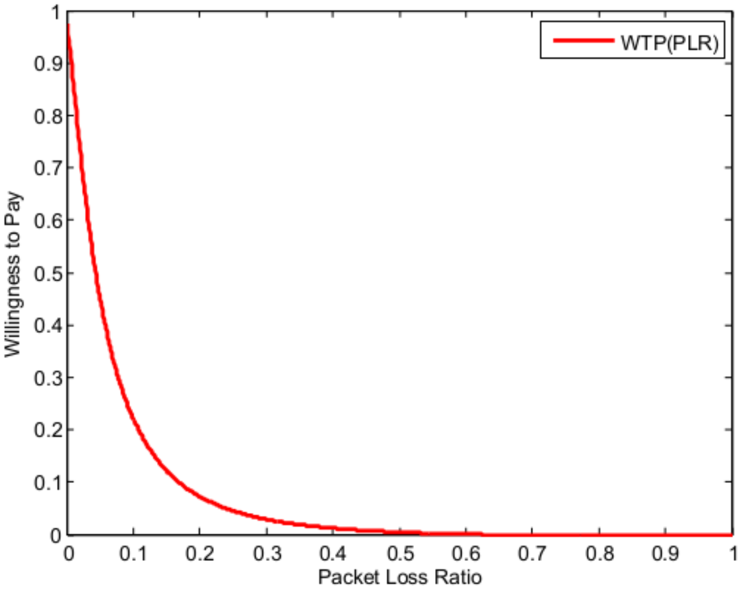


Figure 5.6: WTP(PLR) used for estimating price

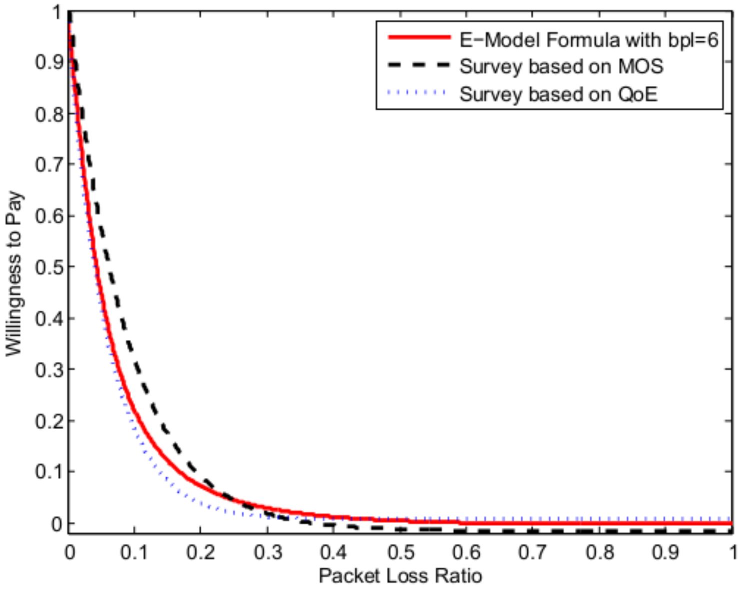


Figure 5.7: Comparison of WTP(PLR) model to empirical data

Chapter 6

Analysis

Assuming users are willing to pay more for a service of higher quality, as established in section 5.1, this chapter investigates optimal pricing structures by comparing pricing before and after implementing service differentiation in an asynchronous, optical packet-switching network. Switching time in a bufferless network is miniscule and will consequently be ignored for this analysis. The arguments for ignoring switching time and other parameters in this analysis to describe the QoS are made in section 2.5.

It is natural to give network priority to the class that generates the highest revenue. It is hence assumed that $\mu_1^{-1} p_1 > \mu_2^{-1} p_2$, (i.e a bit processed in class 1 generates more revenue than a bit processed in class 2), where μ_i^{-1} is the mean service time (time a packet is in the system) for a given service class. In other words, class 1 will have a higher price than class 2.

6.1 Switch Architectures and Arrival Models

To setup the analytical models to compare the before and after differentiated service scenarios, we consider a switch architecture and the arrival models for an asynchronous, bufferless, optical network. The analytical model employed to evaluate the PLR for the single service class scenario is provided in Eq. 6.1, whereas the analytical model and simulated results for the differentiated service scenario are detailed in [53]. In the data leveraged within this paper for both scenarios, the traffic arrival rate is chosen for an asynchronous, bufferless, optical switch with full wavelength conversion.

The packet lengths for traffic for all service classes in both service class scenarios are independent and identically distributed with mean packet length L = 472 [53, 6]. For a discussion, see section 2.5. A list of common parameters and their initial values for both service scenarios is given in Table 6.1. The initial values are used for the remainder of this chapter, unless stated otherwise.

Table 6.1: Network Parameters and Initial Values

Parameter	Description	Initial value
N	Number of wavelengths	128
C	Arrival rate	2.5Gbps
L	Mean packet length	472 Bytes
A	System load	0.8
S_i	Relative share of traffic of a class	0.2
s	Switching time	0

6.1.1 Service Differentiation by APDP

In “Quality of Service in Asynchronous Bufferless Optical Packet Switched Networks” [53], a strictly non-blocking asynchronous, bufferless, optical switch is considered. Packets may arrive at any instant because of the asynchronous nature. As discussed in 2.5 and 2.6, the major concern of switches of this kind is packet loss due to contention. This paper will only focus on the results found with full wavelength conversion. It is further assumed that all connection requests arrive according to an independent Poisson process with constant arrival intensity for each service class. This simplification was done to ensure manageable results, and has been used in other works to calculate performance metrics [8, 53, 56] for optical network architectures. However, an on/off state model with exponential holding times has been adopted to model a finite population of sources and dependency between the service time and the packet inter-arrival time. This ensures a closer model to that of a real switch. A more detailed description is available in APDP

Figure 6.1, borrowed from [53], shows the PLR as a function of the system load A for a switch with full wavelength conversion. It is important to note that [53] denotes High priority traffic as 0 and low priority traffic as 1. This is in contrast to the class 1 and class 2 used in this paper to denote high and low priority traffic, respectively.

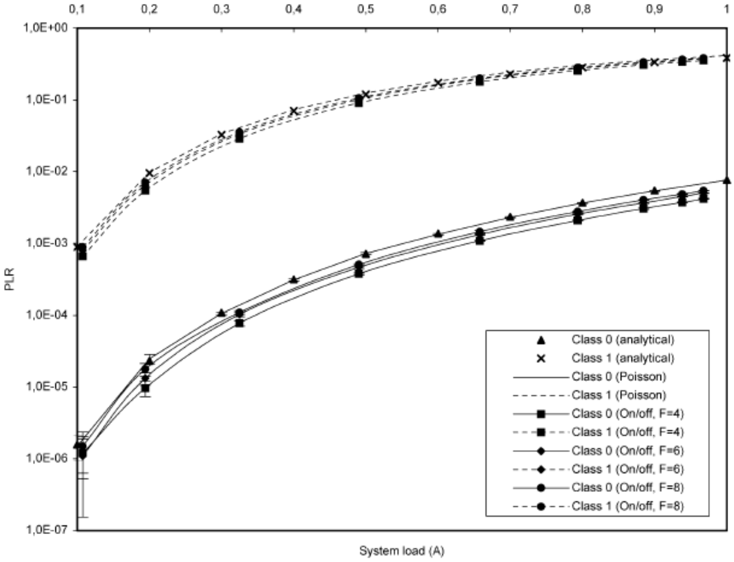


Figure 6.1: PLR as a function of the system load (A), from [53]

6.1.2 Single Service Class, Best Effort Scenario

For the single service class, a constant arrival rate λ with an infinite population of sources is assumed. This differs slightly from the differentiated service scenario which is modeled using an on/off state model with a finite population of sources.

However, the on/off arrival model approaches the Poisson arrival model as the number of input sources increases. This paper assumes a total input of 128 wavelengths (in APDP, 16 wavelengths per 8 fibers), thus the on/off model generated closely follows the Poisson model [53].

To model the single class with full wavelength conversion, the Erlang loss model can be adopted [41].

$$P_b = \frac{\frac{(\lambda h)^N}{N!}}{\sum_{i=0}^{N} \frac{(\lambda h)^i}{i!}} \tag{6.1}$$

Where:

- N = Number of wavelengths
- λ is the arrival rate
- μ^{-1} is the mean service time (time a packet is in the system)

The mean service time is defined as: $\mu^{-1} = L/C$ while the arrival rate is defined as $\lambda = A * N * \mu$, where $\mu = C/L$.

6.2 Fixed System Load Scenario

In a fixed system load scenario, a constant portion of total network capacity is utilized in the network before and after implementing service differentiation. Let $C_N = 0$ such that there are no investments in traffic resources. The NP will profit when,

$$q_1 * p_1 + q_2 * p_2 > C_D + q' * p' \tag{6.2}$$

It is necessary to set a traffic distribution between the two service classes in order to accurately estimate the PLR for each service class in the differentiated service scenario. Let S_1 be the share of the total data sold to class 1 and S_2 be the share of the total date sold to class 2. Then, q_1 will be equal to S_1 multiplied to the total data and q_2 is likewise S_2 times the total data. By assuming that the total traffic does not change with the deployment of service differentiation, i.e. $q_1+q_2=q'$, and that $q_1=0.2*q'$ (relative share of class 1 traffic is 20%), $S_1=0.2$ and therefore $S_2=0.8$. Then equation 6.2 then becomes,

$$p_1 * 0.2q' + p_2 * 0.8q' = C_D + p'q' \tag{6.3}$$

By dividing by q', a relationship between the number of bits sold and the cost of deployment is revealed:

$$p_1 * 0.2 + p_2 * 0.8 - p' = \frac{C_D}{q'} \tag{6.4}$$

where $\frac{q_1}{q'} = 0.2$ and $\frac{q_2}{q'} = 0.8$. Eq. 6.4 shows that when $q' \to \infty$, $C_D \to 0$. Naturally, the more data sold over time, the more the cost of deployment is recovered and the closer to Return on Investment (ROI).

6.2.1 Price-Based Costing

While it can be assumed that the marginal cost of one data packet equals (almost) zero after a given period of time, and a provider can accordingly price internet services, it is usually not in the service providers' interest to do so based on that information alone. Seeking to maximize profits, they would be interested in how much the service is worth to the customer, thereby knowing how much they can charge.

By expressing the price for each class in Eq. 6.3 in terms of the WTP factor as a function of PLR multiplied to the minimum price a provider needs per unit of data for all services (i.e. class 1 plus class 2) to cover the costs of service differentiation deployment, $P_{min} \leq p_1 + p_2$, it is possible to estimate how much an internet provider may charge a user for service in either class. The price charged to each class, based on customer willingness to pay is then:

$$p_1 = P_{min} * WTP(plr_1) \tag{6.5a}$$

$$p_2 = P_{min} * WTP(plr_2) \tag{6.5b}$$

It should be noted that the price P_{min} is denoted here as a minimum, as opposed to a maximum as in chapter 5. This is because, while it is the maximum value of the service a provider will provide, hence the maximum price a customer would be willing to pay for the service, it is redefined for analysis as the minimum price the company will need to cover costs of the service.

Lets assume that a service provider currently charges a price p' per unit of data. If service differentiation is implemented, it may be expected that less data is sold for the BE class due to customer migration to class 1. Additionally, there will be a slight decrease in quality for BE users due to re-distribution of the bandwidth with dedicated bandwidth for class 1 plus the throughput penalty for implementing service differentiation [42]. This means that the revenue of service class 1, $p_1 * q_1$ must cover both that of the deployment cost, C_D per total bytes sold q', and the decreased revenue (relative to the original revenue before differentiation) from service class 2, as shown in equation 6.4 rearranged below,

$$p_1 * 0.2 > \frac{C_D}{q'} + (p' - p_2 * 0.8) \tag{6.6}$$

It is important to note that q' will be the total data sold for the given time period to fully cover the costs of deployment. Expressed in terms of minimum total price, P_{min}, willingness to pay as a function of PLR, $WTP(plr_i)$, and plugging in the total and relative shares of network traffic, q' and q_i, equation 6.6 becomes,

$$0.2 * P_{min} * WTP(plr_1) > \frac{C_D}{q'} + (p' - 0.8 * P_{min} * WTP(plr_2)) \tag{6.7}$$

Solving for P_{min} gives,

$$P_{min} > \frac{\frac{C_D}{q'} + p'}{0.2 * WTP(plr_1) - 0.8 * WTP(plr_2)} \tag{6.8}$$

A service provider considering whether to implement service differentiation would need to know the cost to upgrade their service, estimate the PLR that class 1 will experience, and the PLR that class 2 will experience under a known system load, A, and refer to the studies presented above for an estimate of the customer

WTP factors for either class based on the PLR estimates. Furthermore, they need to determine the time period for a Return on Investment to cover the total costs of deployment, $C_D$and how much data on average they will sell to their total customers in that ROI period. Finding the total bytes sold, q' will be discussed in detail further below. Knowing how much they currently charge, p', and using all the estimated values, they can use the above equation to find a minimum total initial price for the implementation of service differentiation to be worthwhile.

Equations 6.5a and 6.5b will then provide the prices that the provider should charge each class of customers. The last variable that is important to discuss is q', the total data sold. More specifically, in order for the model to work, q'needs to be defined as the data sold in the time frame expected to recover 100% of the invested cost, multiplied by the total ROI time, and by the unit of data sold per unit of time (i.e. average Gigabytes per month for all customers, or capped Gigabytes per month at worst case). For example, if a company has a customer pool of 500,000 users, each using an average of 100 Gigabytes per month, and an executive decision is made to have an ROI of 5 years (60 months), then q' will be $500,000 * 100 * 60$, or $3 * 10^9$ Gigabytes to be sold.

An example of using the customizable model to find the price for each class is as follows: A network provider wishes to implement service differentiation. Suppose it will cost about \$10 million to upgrade infrastructure and implement differentiated service for 2 classes: Guaranteed service and a Best Effort service. The provider is planning for a ROI of 5 years (60 months) and would like to know whether they can reasonable price the services accordingly for service differentiation. Originally, the provider charged \$0.50 per gigabyte of data to 500,000 customers and plans to partition 20% of the data traffic to the Guaranteed class, and 80% to the Best Effort class. The tables below outline various PLR estimates for each service class and the customer WTP for each PLR value under different system loads (calculated using the model from section 5.6.)

In a high traffic situation, in which the system load A is at 80%, the PLR for the service classes might be estimated for each service class as follows in table 6.2.

Table 6.2: WTP in percent of P_{max} with $A = 0.8$

WTP(plr)	WTP
$WTP_1(0.004)$	0.9322
$WTP_2(0.4)$	0.0123

Knowing that the backbones in the internet typically run at 10%-15% of their total capacity [35], ($A = 0.20$), it is necessary to look at lower system loads as well. Table 6.3 presents the WTP factors for PLRs corresponding to an example of a typical system load, 20

Table 6.3: WTP in percent of P_{max}, $A = 0.20$

WTP(plr)	WTP
$WTP_1(0.00002)$	0.9739
$WTP_2(0.009)$	0.8726
$WTP'(3.553 * 10^{-47})$	0.9741

Using the equation 6.8, the company will require a total price of at minimum \$2.83 cumulative per Gigabyte at a system load of 80%. Class 1 would be willing to pay about \$2.67 per Gigabyte under this situation, and class 2 would be willing to pay \$0.04 per Gigabyte (accruing a total of \$2.71 per Gigabyte). At a system load of 0.2, the max price is \$0.83 per Gigabyte. Class 1 would be willing to pay \$0.81 per Gigabyte, and class 2 would be willing to pay \$0.72 (accruing a total of \$1.53 per Gigabyte. Under 80% system load, if the company charged the price either class is willing to pay, the company may not reach the ROI goal and implementation of service differentiation may not be in their best interest (p_1 plus p_2,\$2.71, is less than the estimated P_{min} of \$2.83). On the other hand, at lower system loads (based on existing networks), the company will earn almost twice the amount of revenue per unit of data than required to meet the ROI for service differentiation (p_1 plus p_2,\$1.53,is 1.8 times the estimated P_{min} of \$2.83).

Therefore, it may be an attractive investment.

As argued in section 5.1, it can be seen from Table 6.3 that at lower PLRs, there is a smaller difference in the WTP factors as the users are generally satisfied with the service provided in both classes. The decreased difference may also be in part due to a smaller throughput penalty, which tends to decrease with lower system loads [42]. However, it is most likely a very minor reduction, as the use of PDP to provide QoS differentiation has not been shown to reduce throughput significantly even at higher system loads [42]. In total, the decreased difference in WTP translates to a lower minimum price (P_{min}) necessary to compensate for throughput penalty and cost of deployment over a given time period. Increasing the prices for service in either class (or charging more than P_{min} in total) may be necessary to cover the network costs, C_N, as well. A brief overview of required equipment, is given in 7.

The model presented intends to be customizable for each network provider's situation. Careful analysis is necessary to determine whether implementing service differentiation is worthwhile for the provider given the time to realize a ROI, the network parameters, and whether they can charge the prices to the customer classes necessary to meet the minimum price requirement. Based on the findings, under certain circumstances, service differentiation may be profitable on asynchronous, optical packet-switching network.

Chapter 7

Capital Expenditure of Implementing QoS Differentiation

Much of today's infrastructure is based around SDH/SONET, multiplexing protocols originally designed for circuit mode communications [36]. There are associated costs with implementing and upgrading to service differentiation using APDP for example. The actual cost depends on the pre-existing architecture. By looking at likely components for an OCS or OBS network which utilizes glsdwdm, a NP can look at their current architecture and see what needs to be upgraded to estimate its' CAPEX. This chapter gives an overview of the capital expenditure (CAPEX) involved with deploying two service classes considering both software and hardware.

7.1 Next Generation Optical Networks

Optical WDM networks can be constructed consisting of the following main components [30, 54]:

- (Single mode) Optical Fibers, which allow for high-bandwidth transportation
- Erbium Doped Fiber Amplifiers (EFDAs) to extend the range of an optical fiber by amplifying the weakened optical signals
- Add-Drop Multiplexors (ADMs) and/or Reconfigurable Optical Add-Drop Multiplexer (ROADM)s, which allow for adding or dropping of wavelengths
- Optical Cross Connects (OXC), a fiber interconnection device to route signals from an input port to the desired output port
- Switch Control Units, (SCUs) which create and maintain a forwarding table and are responsible for configuring the OXC.
- A mechanism which forwards packet by labels instead of addresses, (e.g G-MPLS). This allows for pre-configured Label Switched Path (LSP)s, which eliminate the need for a switch to "read" the label of each packet header. This can be used to set up light paths used in OCS and OBS [34].

Some of these components may already exist in the NPs current network. A SDH/SONET-based network may already contain EFDAs and optical fibers of sufficient quality to support OCS or OBS. Hardware CAPEX in such a network would mostly come from upgrading the nodes.

7.2 APDP Implementation Requirements

The CAPEX of implementing APDP are caused by both increased software/scheduling complexity and hardware complexity.

The following is required to implement APDP [42]:

- In order for the switch to know which packets/bursts to prioritize (and which to preempt), the output wavelength state information must include information about the service class of the packets/bursts.
- To minimize bandwidth usage, APDP should only discard the latest packet arrival of lowest priority when performing a preemption. This can be accomplished by including information about when the switched packets arrived to minimize bandwidth.
- To minimize bandwidth waste in downstream nodes, the part of the preempted packet that has already been transmitted needs to be erased. This requires additional hardware.

The adding and dropping of wavelengths in order to provide preemption, can be performed by a ROADM. To remove fragments of preempted packets that have already been transmitted, a checksum may be calculated and compared to the checksum in the header, as described in [54] The output wavelength state information can be included in the packet forwarding mechanism G-MPLS [30]. It can then be concluded that generally, if upgrading from a point-to-point optical architecture, there will be both hardware and software costs. Upgrading from an already bufferless IP-over-WDM architecture, the only added cost is implementing new software.

7.3 Path Protection

As opposed to traditional 1+1 path protection, where two physical fiber links are needed between each node in a network, Protected Network Layer Redundancy Scheme (P-NLPRS) has proven to provide the same functionality at a fraction of the cost, depending on the cost function applied [40]. This efficiency improvement can result in a NP needing to deploy less fiber capacity than if using traditional 1+1 path protection techniques. Given that the cost of deploying new optical fibers can cost as much as $70,000 USD per mile, [48] significant cost reductions can be made. The actual efficiency improvement depends on the network topology, more specifically the node connectivity [40]. Generally, backbone networks are have mesh-based topologies and metro networks have ring-based topologies [5]. Mesh-based topologies generally have higher node connectivity than ring-based. Since many metro networks are currently migrating to a mesh topology [5], a reduction in the cost of deploying new fiber is feasible for most (future) networks. In realistic network scenarios, between 64% and 97% cost savings can be achieved [40]. This reduction in cost, must however be compared to the added cost of deploying [40]. Additionally, P-NLPRS might be deployed solely to combat contention, which results in the cost of deployment being more justifiable than if QoS was not implemented.

Chapter 8

Conclusion

As the capacity of data networks increases, service providers can benefit from knowing how to price their services according to how much a user is actually willing to pay. A model calculating the the WTP factor as a function of PLR has been presented based on customer satisfaction studies. It was used to develop a framework that suggests a pricing strategy for implementing service differentiation. It determines a minimum price to cover the costs of deployment given various network parameters, customer and business parameters, and a ROI time period. The framework addresses the question of whether implementation of service differentiation on asynchronous, optical packet-switching networks is worthwhile and suggests how much a network provider may increase or effectively price their services to improve revenue over time. Finally, a brief cost analysis of implementing QoS differentiation is provided, covering both hardware and software costs to implementing and upgrading existing infrastructure.

8.1 Future Work

This section suggests future work to expand and improve on the frameworks and discussion in this paper. Some suggestions include:

- To improve the mapping between WTP and PLR, an empirical study should be performed with a higher number of test subjects as well as a larger PLR range. Additionally, all network parameters should be controlled as PLR varies.

- To improve the mathematical framework for pricing service differentiation on bufferless optical networks by graphing and modeling. Graphing may help determine when, at certain parameters, the investment would become cost-effective with varying parameters.

- Expand analysis with an increased network capacity scenario which negates the throughput penalty introduced by QoS differentiation.

- The cost analysis should be expanded to include operating expenditure (OPEX) - in particular, energy requirements.

Bibliography

[1] ITU-T SG 12. Estimates of ie and bpl parameters for a range of codec types. Technical report, ITU, 2003.

[2] Paul Ausick. Usage-based internet charges get backing from fcc chairman. *24/7 Wall street*, 2012.

[3] A. Banerjee, L. Drake, L. Lang, B. Turner, D. Awduche, L. Berger, K. Kompella, and Y. Rekhter. Generalized multiprotocol label switching: an overview of signaling enhancements and recovery techniques. *Communications Magazine, IEEE*, 39(7):144–151, 2001.

[4] J. A. Bergstra and C. A. Middelburg. Itu-t recommendation g.107 : The e-model, a computational model for use in transmission planning. Technical report, ITU, 2003.

[5] Joseph Berthold, Adel A. M. Saleh, Loudon Blair, and Jane M. Simmons. Optical networking: Past, present, and future. *J. Lightwave Technol.*, 26(9):1104–1118, May 2008.

[6] S. Bjørnstad, S. Stol, and D.R. Hjelme. Quality of service in optical packet switched dwdm transport networks. *Proceedings of SPIE*, 4910:63–74, 2002.

[7] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss. An architecture for differentiated service, 1998.

[8] M. Yoo C. Qiao. Optical burst switching (obs) - a new paradigm for an optical internet. *Journal of High Speed Networks*, Volume 8, Number 1/1999:69–84, January 1999.

[9] F. Callegati, G. Corazza, and C. Raffaelli. Exploitation of dwdm for optical packet switching with quality of service guarantees. *Selected Areas in Communications, IEEE Journal on*, 20(1):190 –201, jan 2002.

[10] Karl E. Case and Ray C. Fair. *Principles of economics.* Prentice Hall, Englewood Cliffs, NJ [u.a.], 3. ed edition, 1994.

[11] Y. Chen, H. Wu, D. Xu, and C. Qiao. Performance analysis of optical burst switched node with deflection routing. In *Communications, 2003. ICC'03. IEEE International Conference on*, volume 2, pages 1355–1359. IEEE, 2003.

[12] Yang Chen, Chunming Qiao, and Xiang Yu. Optical burst switching: a new area in optical networking research. *Network, IEEE*, 18(3):16 – 23, may-june 2004.

[13] M. Chu and H. Peng. An objective measure for estimating mos of synthesized speech. In *Seventh European Conference on Speech Communication and Technology*, 2001.

[14] Mark E. Crovella and Azer Bestavros. Self-similarity in world wide web traffic: evidence and possible causes. *IEEE/ACM Trans. Netw.*, 5(6):835–846, December 1997.

[15] Ireland DCENR. About exemplar network. `http://www.dcenr.gov.ie/Communications/Communications+Policy/Exemplar+Network/About+Exemplar.htm`. Retrieved May 2012.

[16] Matt Dellinger. Fiber-optic transatlantic cable could save milliseconds, millions by speeding data to stock traders. `http://www.popsci.com/technology/article/2011-04/new-transatlantic-cable-will-speed-information-exchange-price`, 2011.

[17] M. Fiedler, T. Hossfeld, and P. Tran-Gia. A generic quantitative relationship between quality of experience and quality of service. *Network, IEEE*, 24(2):36–41, 2010.

[18] O.S. Gebizlioglu, H. Kuwahara, V. Jain, and J. Spencer. Optical transport and access network technologies [series editorial]. *Communications Magazine, IEEE*, 50(1):98 –99, january 2012.

[19] B. Glance and R.W. Wilson. Optical delay line, November 22 1994. US Patent 5,367,586.

[20] T.A. Hall. Objective speech quality measures for internet telephony. In *Voice over IP (VoIP) Technology, Proceedings of SPIE*, volume 4522, pages 128–136, 2001.

[21] Infinera. Photonic integrated circuits: A technology and application primer. `http://www.infinera.com/pdfs/whitepapers/Photonic_Integrated_Circuits.pdf`, 2005.

[22] D. Isenberg. Research on costs of net neutrality. `http://isen.com/blog/2007/07/research-on-costs-of-net-neutrality.html`, 2007.

[23] A. Kaheel, T. Khattab, A. Mohamed, and H. Alnuweiri. Quality-of-service mechanisms in ip-over-wdm networks. *Communications Magazine, IEEE*, 40(12):38 – 43, dec 2002.

[24] T. Karagiannis, M. Molle, M. Faloutsos, and A. Broido. A nonstationary poisson view of internet traffic. In *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, volume 3, pages 1558–1569. IEEE, 2004.

[25] Jeremy Klein, Jonathan Freeman, Jane Lessiter, Eva Ferrari, and Mark Wilds. Assessing network quality of experience, 2009.

[26] Richard Lawler. Comcast rethinks bandwith caps-trials two new policies that involve higher 300gb monthly limits. *Engadget.com*, 2012.

[27] W.E. Leland, M.S. Taqqu, W. Willinger, and D.V. Wilson. On the self-similar nature of ethernet traffic (extended version). *Networking, IEEE/ACM Transactions on*, 2(1):1–15, 1994.

[28] Lei Liang, Zhili Sun, and H. Cruickshank. Relative qos optimization for multiparty online gaming in diffserv networks. *Communications Magazine, IEEE*, 43(5):75 – 83, may 2005.

[29] G. Maier, S. De Patre, M. Martinelli, and A. Pattavina. Resilience schemes in wdm networks, 2001.

[30] E. Mannie. Generalized Multi-Protocol Label Switching (GMPLS) Architecture. RFC 3945, 2004.

[31] D. Miller. Device requirements for optical interconnects to silicon chips. *Proceedings of the IEEE*, 97(7):1166–1185, 2009.

[32] A.J. Mohammad, D. Hutchison, and J.P.G. Sterbenz. Poster: Towards quantifying metrics for resilient and survivable networks. In *Proceedings of the 14th IEEE International Conference on Network Protocols (ICNP 2006). Santa Barbara, California, USA*, pages 17–18, 2006.

[33] Le Thu Nguyen, Richard Harris, and Jusak Jusak. A pilot study to assess quality of experience based on varying network parameters and user behaviour, 2011.

[34] X. Niu, W.D. Zhong, G. Shen, and T.H. Cheng. Connection establishment of label switched paths in ip/mpls over optical networks. *Photonic Network Communications*, 6(1):33–41, 2003.

[35] A. Odlyzko. Data networks are lightly utilized, and will stay that way. *Review of Network Economics*, 2(3):210–237, 2003.

[36] M.J. O'Mahony, C. Politi, D. Klonidis, R. Nejabati, and D. Simeonidou. Future optical networks. *Journal of Lightwave Technology*, 24(12):4684–4696, 2006.

[37] Focus Group on IPTV standardization. Definition of quality of experience (qoe). Technical report, ITU Telecommunication Standardization Sector (ITU-T), 2007. `http://ties.itu.ch/ftp/public/itu-t/fgiptv/readonly/Previous_Meetings/20070122_MountainView/il/T05-FG.IPTV-IL-0050-E.htm`.

[38] H. Øverby. Network layer packet redundancy in optical packet switched networks. *Optics Express*, 12(20):4881–4895, 2004.

[39] H. Øverby. *Quality of service differentiation, teletraffic analysis and network layer packet redundancy in optical packet switched networks.* PhD thesis, Norwegian University of Science and Technology, 2005.

[40] H. Øverby. Cost evaluations of the protected network layer packet redundancy scheme. *J. Opt. Netw.*, 5(10):747–763, Oct 2006.

[41] H. Øverby. Traffic modelling of asynchronous bufferless optical packet switched networks. *Computer communications*, 30(6):1229–1243, 2007.

[42] H. Overby, N. Stol, and M. Nord. Evaluation of qos differentiation mechanisms in asynchronous bufferless optical packet-switched networks. *Communications Magazine, IEEE*, 44(8):52 –57, aug. 2006.

[43] V. Paxson and S. Floyd. Wide area traffic: the failure of poisson modeling. *IEEE/ACM Transactions on Networking (ToN)*, 3(3):226–244, 1995.

[44] S. Ramamurthy, L. Sahasrabuddhe, and B. Mukherjee. Survivable wdm mesh networks. *Journal of Lightwave Technology*, 21(4):870, 2003.

[45] Janko Roettgers. At&t's new bandwidth cap is bad news for netflix. *Gigaom.com*, 2011.

[46] J. Scharf, A. Kimsas, M. Kohn, and G. Hu. Obs vs. opmigua-a comparative performance evaluation. In *Transparent Optical Networks, 2007. ICTON'07. 9th International Conference on*, volume 3, pages 294–298. IEEE, 2007.

[47] Telecommunication standardization sector of ITU. Y.1541 : Network performance objectives for ip-based services. Technical report, International Telecommunication Union, 2012.

[48] Cisco Systems. *Introduction to DWDM Technology.* Cisco Systems, 2001.

[49] B. Teitelbaum and S Shalunov. Why premium ip service has not deployed (and probably never will). `http://qos.internet2.edu/wg/documents-informational/20020503-premium-problems-non-architectural.html`, 2002. Internet2 QoS Working Group Informational Document.

[50] György Terdik and Tibor Gyires. Does the internet still demonstrate fractal nature? In *Proceedings of the 2009 Eighth International Conference on Networks*, ICN '09, pages 30–34, Washington, DC, USA, 2009. IEEE Computer Society.

[51] P.N. Tran. Wdm network planning and management, 2010.

[52] R.S. Tucker, R. Parthiban, J. Baliga, K. Hinton, R.W.A. Ayre, and W.V. Sorin. Evolution of wdm optical ip networks: A cost and energy perspective. *Journal of Lightwave Technology*, 27(3):243–252, 2009.

[53] Harald Øverby and Norvald Stol. Quality of service in asynchronous bufferless optical packet switched networks. *Telecommunication Systems*, 27:151–179, 2004. 10.1023/B:TELS.0000041007.13864.1c.

[54] V.M. Vokkarane and J.P. Jue. Prioritized burst segmentation and composite burst-assembly techniques for qos support in optical burst-switched networks. *Selected Areas in Communications, IEEE Journal on*, 21(7):1198–1209, 2003.

[55] S. Yao, B. Mukherjee, SJ Yoo, and S. Dixit. A unified study of contention-resolution schemes in optical packet-switched networks. *Journal of Lightwave Technology*, 21(3):672, 2003.

[56] Myungsik Yoo, Chunming Qiao, and Sudhir Dixit. Qos performance of optical burst switching in ip-over-wdm networks. *Selected Areas in Communications, IEEE Journal on*, 18(10):2062 – 2071, oct 2000.

[57] Qiong Zhang, Neal Charbonneau, Vinod Vokkarane, and Jason Jue. Tcp over optical burst-switched networks with controlled burst retransmission. *Photonic Network Communications*, 22:299–312, 2011. 10.1007/s11107-011-0329-8.

[58] Qiong Zhang, V.M. Vokkarane, J.P. Jue, and Biao Chen. Absolute qos differentiation in optical burst-switched networks. *Selected Areas in Communications, IEEE Journal on*, 22(9):1781 – 1795, nov. 2004.

[59] D. Zhou and S. Subramaniam. Survivability in optical networks. *Network, IEEE*, 14(6):16–23, 2000.

Appendix A

WTP Calculator

This appendix lists relevant MATLAB code for WTP calculation based on PLR. A zip file containing the files may be found by searching for the thesis title at: `http://daim.idi.ntnu.no/soek/index.php`

A.1 WTB Calculator

WTBcalculator.m calculates a willingness to buy for a service, provided a packet loss ratio for that service.

- The input, plr, must be in percent from 0-100.
- The output, WTP, is given in percent of a maximum price a customer is willing to pay for a hypothetical perfect service or the minimum price a company requires to cover all service costs.