# Instituto Tecnológico y de Estudios Superiores de Monterrey

*Campus Monterrey*

División de Graduados en Computación,
Información y Comunicaciones

Programa de Graduados en Computación,
Información y Comunicaciones

# A cost function for Internet with QoS parameters

## THESIS

Presented as partial fulfillment of the requirements for the degree of

Master of Science in Electronic Engineering
Major in Telecommunications

**Mauricio de Medina Aguilera**

August 1999

# Instituto Tecnológico y de Estudios Superiores de Monterrey

*Campus Monterrey*

División de Graduados en Computación,
Información y Comunicaciones

Programa de Graduados en Computación,
Información y Comunicaciones

# A cost function for Internet with QoS parameters

## THESIS

Presented as a partial fulfillment of the requirements for the degree of

**Master of Science in Electronic Enginnering**
**Major in Telecommunications**

**Mauricio de Medina Aguilera**

August 1999

# Aknowledgments

I want to thank Instituto Tecnologico y de Estudios Superiores de Monterrey (ITESM), for the opportunity to perform my studies of Master Deegre.

I am grateful to my thesis advisor, Cesar Vargas Rosales, Ph. D.; this work could not have been possible without his interest and collaboration.

I also want to thank Artemio Aguilar, M. Sc., and Leoncio Pardo, M. Sc., for their valuable comments.

Thanks to all my fellow students at CET.

Thanks to my entire family for their support and love.

# Abstract

Nowadays the number of users in the Internet is growing in an exponentially way and there are several applications that the Internet offers and these require to be handled in different ways to run efficiently and improve the use of the network resources.

We proposed a revenue function that allows the service provider to know their maximum revenue based on some parameters such as – the average number of packets, the average delay, the arrival rates, the capacity in the network links- and associate it with costs.

In order to obtain the revenue function, we established a network topology with two different classes of traffic, the delay permissible for each one and the capacity in the links of the network.

We used the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method to maximize the revenue function. We compare the optimized values versus non optimized values to observe their differences and how they rebound in the network performance and it is possible to control the network based on the prices.

# Table of Contents

# List of Figures

# List of equations

# List of Tables

# Chapter 1

# Introduction

Networks such as Internet are actually providing a wide variety of services, which are consuming widely differing amounts of resources.

It is correct that the technology is increasing, but so is the demand. Different kinds of traffic require different treatment on the network. The usage prices are not a good way to recover the cost of providing network capacity. Since the network costs are primarily the fixed cost of capacity, it makes more sense to charge users a fixed fee depending on the resources of their connection to the net. This is essentially the scheme that is used now. For a pricing system to be "incentive compatible" it is necessary that the use of higher quality service incur a higher cost to the user.

With the growing demand for multimedia, it is necessary to think about the allocation of multiple service qualities in an integrated network. For example, file transfers tolerate zero errors, but can tolerate substantial delay, on the other hand interactive video can tolerate some packet loss, but requires tight bounds on maximum delay and variation in delay.

In this thesis a cost function is proposed and analyzed to observe the behavior of the network under a pricing scheme.

Three stages need to be followed in order to provide good pricing schemes that optimize network performance. The first stage consists of defining the traffic classes, the QoS (Quality of Service) parameters to be satisfied as well as an objective function to be optimized which represents network net revenue, the optimization considers flat revenues and costs along the resources, services and routes.

The second stage is to consider the control of the traffic by varying the revenues and costs such that performance is maximized. The third stage is to consider different pricing schemes such as usage based [9], edge pricing [12], etc., and determine a pricing strategy such that dynamically controls network performance in an optimized form.

## 1.1 Objective

The aim of this work is to analyze the behavior of a network such that its processes give insight the behavior of the Internet by using QoS parameters to give a cost to the user and optimize performance of the network. We propose a cost function for the Internet Service Provider such that he knows the traffic that allows have the greatest profit under some scenarios.

We will gain insight on the issue of traffic control through pricing strategies.

## 1.2 Justification

Due to the growth of users at Internet and the inefficiency of giving a "good service," it is becoming important to increase the service and quality given by the network. As well as the revenues keeping prices low. Since different classes of traffic require different treatment from the network and it is necessary to optimize the resources of it, for this reason it is becoming important to have a function that allows the service provider to know the maximum utility.

## 1.3 Contribution

We have proposed a cost function in relation to the average number of packets, the average delay in the network, the number of links, the arrival and the departure rate in the network, the QoS necessary for different kinds of traffic and we have included a cost per link and per application.

## 1.4 Organization

The content of this thesis is organized as follow.

In Chapter two the concepts about Internet and the current pricing model are presented. In the Chapter three introduces the performance measures, the analytical model and the cost function, also the parameters and assumptions involved in our scenario. Chapter four presents the numerical results comparing a scheme using optimization of the cost function and without optimizing it and varying the variables involved. Finally, in Chapter five we present conclusions and suggest future work to continue with the issue of this thesis.

# Chapter 2

# Pricing the Internet

In this chapter we explain the characteristics of the Internet and we introduce the theoretical aspects of pricing in networks, we discuss the current pricing scheme and how the pricing can be made.

## 2.1 Internet Architecture and Mechanisms

The current Internet architecture is designed for point-to-point (or unicast) best-effort communication. Every packet header contains a source address and a destination address. Upon receiving a packet, a switch (or, equivalently, a router) consults its routing table to find, based on the packet's destination address, the appropriate outgoing link for the packet. The network makes no commitments about when, or even if, packets will be delivered. Sometimes the incoming rate of packets at a switch is greater than the outgoing rate, so queues build up in the switch. These queues cause packet delays and, if the switch runs out of buffer space, packets are discarded. The network does not attempt to schedule use; a source can send packets at any time, and the network switches merely exert their best effort to handle the load, [2].

However, there are efforts currently underway to extend this architecture in two ways. The first is to offer better support for multipoint-to-multipoint communications through the use of multicast. In the current Internet architecture, when a source sends a packet to multiple receivers, the source must replicate the packet and send one to each receiver individually. This results in several copies of the same packet traversing those links common to the delivery paths (i.e., those links that lie on more than one delivery path, where the delivery path is the route taken by the packet from source to receiver).

Nowadays the efforts extend the Internet's current service offerings to include a wider variety of qualities of service (QoS).

Actually the Internet has characteristics there is to take into account. First the overall bandwidth is quite limited which restricts the aggregate traffic load and prevents certain bandwidth-intensive applications from utilizing the Internet. Second, access to the Internet is severely restricted; this is in past because of the limited bandwidth available. Third, the Internet offers a single type of service (TOS); all packets are serviced on a best-effort, first-in-first-out (FIFO) basis. This single TOS severely limits the nature of applications that can be adequately supported. Fourth, there are no usage fees; users are not charged on the basis of how many packets they send, these fees are not based on the volume of traffic sent, [6], and are typically not passed back to individual users.

The current single class of best-effort service may not be sufficient to adequately support the requirements of some future video and voice applications. Moreover, offering all applications the same service is not an efficient use of bandwidth; providing several service priority levels and/or drop priority levels, [12].

## 2.2 Internet Technology

The Internet network uses packet-switching communications technology based on the TCP/IP protocols. While much of the traffic moves across leased lines from Network service providers, packet-switching technology is quite different from the circuit-switching used for voice telephony, where at the moment a telephone user dials a number, a dedicated path is set up between the caller and the called number. This path, with a fixed amount of network resources, is held open; no other caller can use those resources until the call is terminated. Packet-switching network, by contrast, uses "statistical multiplexing", this means that each circuit is shared by many users, and no open connection is maintained for a particular communications session. Data streams are little pieces called "packets". When a packet is ready, the computer sends it onto the network. When one computer is not sending a packet, the network line is available for packets from other computers. The destination does not need to be aware that a packet has been sent to it.

The TCP (Transmission Control Protocol) provides how to break up a data stream into packets and reassemble it; the IP (Internet Protocol) provides the necessary information for various computers on the Internet (the routers) to move the packet to the link on the way to its final destination.

The other distinguishing feature of Internet technology is that it is "connectionless". This means that there is no end-to-end setup for a session; each packet is independently routed to its destination. When a packet is ready, the host computer sends it to another computer, known as a router. The router examines the destination address in the header and passes the packet along to another router, chosen by a route-finding algorithm, [1].

## 2.3 The current Internet pricing model

We know that the Internet transport is already priced, although many users seem unaware of it. Pricing is on the basis of a fixed monthly subscription fee for a connection of a given bandwidth (called flat rate pricing). Most of the costs of providing the Internet are more-or-less independent of the network; i.e. most of the costs are fixed costs. If the network is not saturated the incremental cost of sending additional packets is essentially zero. In most cases the incremental usage of that bandwidth is priced at a flat rate of zero, [4].

Given that congestion is likely to be a serious problem in the near future of the Internet, and past proposals for its control unsatisfactory, it is beneficial to consider other pricing schemes that take into account incremental costs and allow some congestion control. Since the bandwidth is scarce, when the backbone or the capacity of the links becomes congested, one user's packet crowds out another's, resulting in dropped or delayed transmissions. Allocating scarce resources among competing uses is the central focus of economics. One of the fundamental principles of economics is that prices should reflect costs. In addition the pricing may discourages network usage by customers in times of high network load or congestion due to higher prices and it provides revenue for providers which may be used to extend or renew the current networking infrastructure and capacities.

As a general rule, users should face the prices that reflect the resource costs that they generate so they can make informed decisions about resource utilization. Some ways to do these are, [5]:

*The incremental costs of sending extra packets*

The price of sending a packet in an uncongested network should be close to zero; a higher price is socially inefficient since it does not reflect the true incremental costs. But if the network is congested there is more probability of delay and dropped packets so that some packets should be retransmitted.

*The social costs of delaying other user's packets when the network is congested*

When the network is congested due to large amounts of frames crossing it, this causes a significant delay in all users.

*The fixed costs of providing the network infrastructure*

In the case of a computer network like Internet, it is natural to think of paying for the network infrastructure with a flat access fee. Each party who connects to the network pays a flat price for network access distinct from the usage-based fee described earlier.

In general, these connection fees will vary, since different people and institutions value connection to the net differently. If the total amount that users are willing to pay exceeds the infrastructure cost, the fees could be assigned in a variety of ways, depending on the market conditions and the network providers' objectives.

*The incremental costs of connecting to the network.*

If network usage never reaches capacity, even at no costs for packets, then clearly there is no need to expand capacity. Usage prices that are based on congestion provide guidance about when to expand capacity. Consider the model with fixed capacity: Packet prices measure the marginal value of the last admitted packet. If the cost of expanding capacity to accommodate one more packet is less than the marginal value of that packet, then it makes economic sense to expand capacity.

*The incremental costs of using more links*

When the network has its capacity or router full or even is incapable of giving a QoS acceptable it is necessary to follow another path, where in some cases the route taken will be more expensive since the path taken uses more resources because it is using more links or using links dedicated to give another class of service or is full also.

## 2.4 Effects on and Inputs for Pricing schemes

Any pricing scheme or policy is affected by at least eight factors that encompass the economic, the technological, and the administrative situation, these form the market structure of the network, [11]:

- cost recovery
- network demand
- network load
- cost of working equipment
- type of services
- capacity expansion
- regulatory environment

A pricing scheme has to take into account the market in which communication services are offered. It needs to recover costs fully that are based on the provision of the network infrastructure and the equipment, its maintenance, its investment for future enhancements and extensions, and its offered type of services. Additionally, special legal restrictions for applying schemes have to be obeyed. Furthermore, a pricing scheme has to follow certain goals to be suitable as a monetary incentive. These economic goals include at least: (1) optimal efficiency to gain maximal welfare, (2) maximize the total utility of all users, (3) maximize the utility of a specific resource, (4) maximize profit, (5) fairness, and (6) recover costs fully, [8].

Pricing schemes or changes that are designed to recover costs fairly from this diverse population of users and to allocate network resources in an efficient manner require the capability to measure individual user traffic and to present users with prices and charges in a way that encourages efficient network use.

One incentive to reduce technical demands is a fair pricing scheme that takes into account the requested service class or QoS per streams.

## 2.5 Variability of prices

Prices may vary based on a number of different factors. These factors sometimes determine the charging unit, a certain period of time when the price is valid, or administrative issues. They can include different criteria, features or combination of them such as the usage of a particular resource, network access, Quality of Service (QoS) requirements, congestion level of the network , time-of-day, duration, destination, distance or number of hops to be traversed, requested service class, demand and supply, network providers and their cost structures, networking equipment/technology, tariffing acts, legal issues, per- packet, per-reservation, per flow, per-call, per-connection, per-day, week or month, [11].

The satisfaction that a network user derives from his/her network access depends on the nature of the application being used and the quality of service received from the network (in terms of bandwidth, delay, packet dropped, etc.); the network resources are used most efficiently if they maximize the total user satisfaction of the user community.

One component of network charges is the attachment fee; this is the fee charged for gaining access to the network and is independent of any actual or potential usage. One minimal requirement is that pricing should encourage the appropriate use of quality of service (QoS) signals (this means the signals sent by applications to the network requesting a particular quality of service). This is crucial for making the new QoS-rich network designs effective, and would enable them to achieve significant increases in network efficiency. An additional requirement is that pricing should discourage network usage during times of congestion, but not discourage it during relatively uncongested times somewhat vaguely like the telephone companies.

### *Pricing and type of service*

Different kinds of traffic require different treatment from the network. E-mail can be delayed without much loss; real-time video needs very rapid service. In order to provide appropriate treatment for different kinds of services, the network needs to know what type of traffic is actually flowing and the traffic required. For this, the routers must be capable of distinguish the type of traffic or the person who generates the data must indicate it.

With the growing demand for multimedia, it is necessary to think about how to allocate multiple service qualities in an integrated network. For example, file transfers tolerate substantial delay. Interactive video can tolerate some packet loss, but requires tight bounds on maximum delay and variation in delay.

The only way to measure the benefit of a pricing scheme is to measure the net satisfaction from the network. From the network point of view in order to provide customers an adequate service, the network needs a minimal description of the traffic type, the traffic descriptor, and the required QoS. These permit the network to allocate the necessary bandwidth and paths among the type of users as a function of the demand they place.

In this chapter we have explained the Internet structure and technology and how the pricing can be made, where the main objective is pricing the Internet without changing the current architecture. In the next chapter we introduce the network description and his traffic descriptors used to make our cost function.

# Chapter 3

# Model Description

In this chapter we introduce a cost function. This function takes into account parameters such as average number of packets, average delay allowed for each flow of data, values of permissible QoS for different kinds of traffic, type of data, cost and utility per class and link. We describe the topology of a network to be evaluated using pricing criteria to observe the behavior and performance of the network.

## 3.1 Background

*Network of transmission lines*

In data networks, there are many transmission queues that interact in the sense that a traffic stream departing from one queue enters one or more other queues, perhaps after merging with portions of other traffic streams departing from yet other queues. The difficulty is that the packet interarrival times become strongly correlated with packet lengths once packets have traveled beyond their entry queue. As an illustration of the phenomena that complicate the analysis, consider two transmission lines of equal capacity in tandem, as shown in Figure 3.1. Assume that Poisson arrival of rate $\lambda$ packets / sec enter the first queue, and that all packets have equal length. The first queue is $M / D / 1$ and in the second queue the interarrival times must be greater than or equal to $1/\mu$ (the packet transmission time). Furthermore, because the packet transmission times are equal at both queues, each packet arriving at or before the time the next packet arrives, so there is no waiting time in the second queue. Then, the interarrival times at the second queue are strongly correlated with the packet lengths.

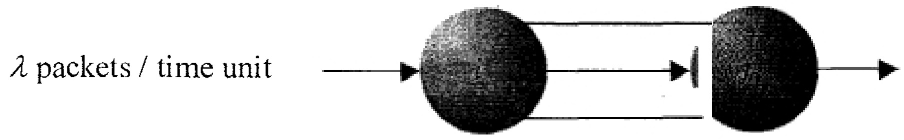


**Figure 3.1. Two equal capacity transmission lines in tandem.**

*The Kleinrock Independence Approximation*

We have seen the special case of two tandem queues that even if the packet stream is Poisson with independent packet lengths at their point of entry into the network, this property is lost after the first transmission line. To resolve the dilemma, it was suggested by Kleinrock, [7] that merging several packet streams on a transmission line has an effect akin to restoring the independence of interarrival times and packet lengths. It was concluded that it is often appropriate to adopt an $M / M / 1$ queueing model for each communication link regardless of the interaction of traffic on this link with traffic on other links. This is known as the Kleinrock independence approximation and seems to be reasonably good approximation for systems involving Poisson stream arrivals at the entry points, packet lengths that are nearly exponentially distributed, a densely connected network, and moderate to heavy loads, [1].

## 3.2 Types of application

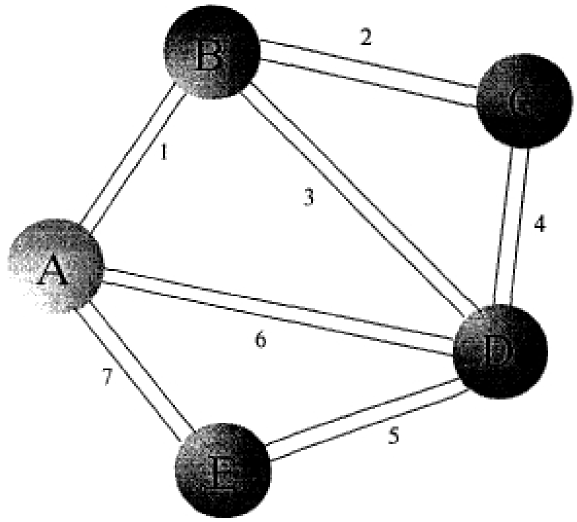
Suppose a network that has different classes of service which needs to be distinguished by the bandwidth requirement to provide an acceptable QoS. In our scenario we suppose that each router can distinguish each type of application, where the network will give the necessary bandwidth required. We constructed the objective function representing users of two different applications.

To compare the behavior, we have taken randomly different values of average customer arrival rate in packets per time unit ($\lambda$) and compare them with optimized $\lambda$'s obtained by solving a nonlinear optimization problem introduced later in this chapter.

## 3.3 Description of the network

The network used for the simulation has $N$ nodes that are connected by $L$ links. Let $\aleph$ be the set of nodes and $\mathcal{L}$ the set of links. Some of the nodes (known as end nodes) are directly connected to end stations while others are just used for transit between the end nodes, [14], see Figure 3.2.

The flow followed by the packets from source to destination is called path and these paths can take different routes. Routing tables located at each router determine the path followed by a session. These tables specify the links that a session must occupy to establish a connection from origin to destination. The links will accept sessions while they can provide the bandwidth required and assure the delay in the QoS contract. The QoS contract is where the service provider guarantees to the user the bandwidth and the necessary treatment for his/her application. Traffic sources are divided into classes and each class is provided with a different QoS guarantee, e.g., [10]. Since we argued that providing QoS per session is necessary to satisfy the diverse requirements of the applications, [16], we assumed that the network is capable of making QoS guarantees.



N = 5

L = 7

$\aleph = \{A, B, C, D, E\}$

$\mathcal{L} = \{1, 2, 3, 4, 5, 6, 7\}$

Figure 3.2. Connection between nodes

## 3.4 Performance measures

We assume that there are several packets crossing the subnet through paths fixed along the network and each flow following a path $p$ consists of a sequence of links $\ell \in \mathcal{L}$. Where $X_p$, in packets per time unit, is the arrival rate of the packet stream associated with the path $p$. Let $\Re_p$ be the set of links that path $p$ traverses, i.e., if path $p$ traverses links $(i, j)$ and $(j, k) \in \mathcal{L}$ then $\Re_p = \{(i, j), (j, k)\}$.

The total arrival rate to link $(i, j)$ is given by

$$\lambda_{ij} = \sum_{\forall p:\,(i,j) \in \Re_p} \chi_p. \tag{3.1}$$

We have considered that there are equal capacity transmission lines that receive Poisson arrivals with $\lambda_{ij}$ packets per time unit for link *(i, j)*, entering each queue. Services times are exponentially distributed with parameters $\mu_{ij}$ in link *(i, j)*, and an infinite buffer at each node. Assuming that all the sessions are independent of each other, arrival processes independent of other arrivals and departures, then Kleinrock's approximation, [7], can be used and we can model each link as an *M/M/1* queue. We have considered also several kinds of traffic at each path crossing the network, let *K* be the set of traffic classes in the network.

In order to measure the features of the network we need to establish the corresponding performance measurements such as the average number of packets and delay of each link $(i, j) \in \mathcal{L}$ and class $k \in K$. Based on these assumptions, the average number of packets in queue or service at link $(i, j) \in \mathcal{L}$ and class $k \in K$ is, [1],

$$N_{ijk} = \frac{\lambda_{ijk}}{\mu_{ijk} - \lambda_{ijk}}, \tag{3.2}$$

where $\mu_{ijk}$ is defined as the departure rate and it is the capacity of each link $(i, j) \in \mathcal{L}$ and class $k \in K$ and $1 / \mu_{ijk}$ is the average packet transmission time on link $(i, j) \in \mathcal{L}$ and class $k \in K$.

The average number of packets at link *(i, j)* is the sum of packets of all classes traversing such link,

$$N_{ij} = \sum_{k \in K} N_{ijk}, \tag{3.3}$$

The average number of packets in the network is

$$N = \sum_{(i,j)\in \mathcal{L}} \sum_{k \in K} \frac{\lambda_{ij_k}}{\mu_{ij_k} - \lambda_{ij_k}}, \tag{3.4}$$

The relationship among the arrival rate per link $\lambda_{ij}$ and the departure rate $\mu_{ij}$ can be observed in Figure 3.3, where a link is shown, and paths $p_1$, $p_2$ traverse this link, hence the total offered traffic to the link is the sum of the traffic offered by each of the paths.

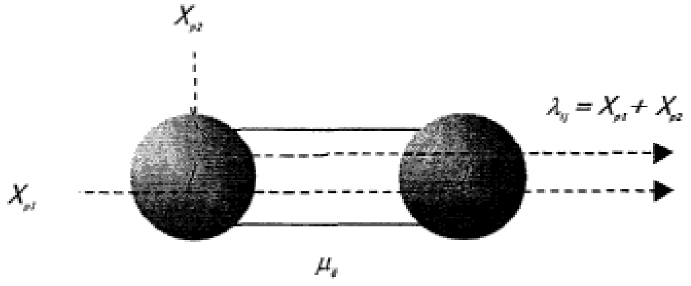


**Figure 3.3. Relationship between $\lambda_{ij}$ and $\mu_{ij}$.**

Then, using Little's Theorem with Equation (3.2), the average delay per packet in each link $(i, j) \in \mathcal{L}$ and class $k \in K$ is given by

$$T_{ij_k} = \frac{1}{\gamma_k} N_{ij_k}, \tag{3.5}$$

where $\gamma_k = \sum_p X_p$ is the total arrival rate to the system per class $k$. We have considered the average processing and propagation delay $d_{ij}$ at each link $(i,j)$ negligible.

The average delay per packet at link $(i, j) \in \mathcal{L}$ is,

$$T_{ij} = \frac{1}{\gamma_k} N_{ij}, \tag{3.6}$$

There are several services offered by Internet (voice, video, data, etc.) those services need different delay to work efficiently, this means that they need different classes of service. Resulting that the average delay per packet in link $(i, j) \in \mathcal{L}$ and class $k \in K$ in the system is

$$T = \sum_{(i,j) \in \mathcal{L}} \sum_{k \in K} T_{ij_k}. \tag{3.7}$$

## 3.5 Revenue Function

Our revenue function has been developed from the point of view of the service provider so that the latter can estimate the profit from the parameters mentioned above. Revenue, in our case, is the total utility minus the costs of giving a "good service". The utility is related directly to the amount of packets crossing different links from the route taken and the price per each link crossed.

The objective of developing a revenue function has been to find an optimal resource allocation for each user in such a way that the total utility of this link is maximized. From [3], [5], [13], [18], [19], we could notice that the most significant parameters —in order to give a good service— are delay, bandwidth and network load. In our model we consider infinite buffers, hence packet loss is not included in the cost function.

Taking into account equations (3.2), (3.4) and adding a cost per link $(i, j) \in \mathcal{L}$ and class $k \in K$, $c_{ijk}$ which represents the cost of keeping working the link with the features that allow the treatment of each class. We also have included the revenue of transporting a packet in link $(i, j) \in \mathcal{L}$ and class $k \in K$, $w_{ijk}$, and the QoS in relation to the delay needed in each link $(i, j) \in \mathcal{L}$ and for each class $k \in K$, $Q_{ijk}$. The parameters, $c_{ijk}$ and $w_{ijk}$ are associated with a price so as to obtain a relation that represents a profit or a loss.

In order to obtain the expected profit in the network, we multiply the average number of packets per link $(i, j) \in \mathcal{L}$ and class $k \in K$, Equation (3.4), by $w_{ijk}$ since we want to have a cost function that involves the total revenue of the network, ***R***, we need to exert a penalty if the provider of service does not accomplish the requirements of each type of application and it will be a loss in his profit. Given that the delay in link $(i, j) \in \mathcal{L}$ and class $k \in K$, $T_{ijk}$, is the QoS parameter we have subtracted it from $Q_{ijk}$.

In order to normalize the Equation (3.8) we have multiplied by $\gamma_k$, this normalization obeys Little's theorem, [1], and it is necessary to work with both parts of our cost function.

$$R = \sum_{(i,j) \in \mathcal{L}} \sum_{k \in K} N_{ij_k} w_{ij_k} + \sum_{k \in K} \gamma_k \sum_{(i,j) \in \mathcal{L}} c_{ij_k} (Q_{ij_k} - T_{ij_k}). \tag{3.8}$$

It must be noticed that we have summed the second term of Equation (3.8), that is the penalty, it will be negative when $T_{ijk} > Q_{ijk}$ which will decrease revenue when not satisfied, otherwise it will increase revenue.

The penalty contributes negatively when there are too many users in the network causing delay increases such that the QoS contract is not satisfied.

As it will be seen ahead, the optimization becomes important since it is of help to the service provider to obtain the values of different variables such as the penalty.

Then the problem has become an optimization problem of the parameters where we can notice that Equation (3.8) depends on the values of the variables involved specially of the arrival rates $\lambda s$.

In this chapter we have introduced the equations and assumptions of our scenario in a general way, we have described the performance measures that allow to obtain our revenue cost and how we got the model. Based in all these assumptions we will compare the results once the revenue cost be optimized versus non-optimized, the method used to maximize the function and their constraints will be mentioned and the results will be shown in the next chapter.

# Chapter 4

# Numerical Results

The cost function proposed in Chapter 3 is evaluated using one network that offers different services and two different classes of QoS.

The network analyzed tries to represent the behavior of the Internet using a pricing scheme, in this network we have proposed five routers without having a specific topology, but having different paths.

## 4.1 Network Model

In this model we have chosen five nodes with infinite buffer and four links with a finite capacity where not all the nodes are connected between them trying to represent a piece of the Internet.

We have proposed that the departure rate be greater than the arrival rate in order to have a good performance. The path followed by the flow is fixed where the network will accept the petition of service if it has enough bandwidth at each link of the path and if the delay lies within the parameters established.

In this scenario we have five flows which follow fixed paths (see Figure 4.1), the values of each arrival rate are different for each path $p$ and class $k \in K$. We have considered that the bandwidth necessary per session is maintained along their routes until the session is terminated.

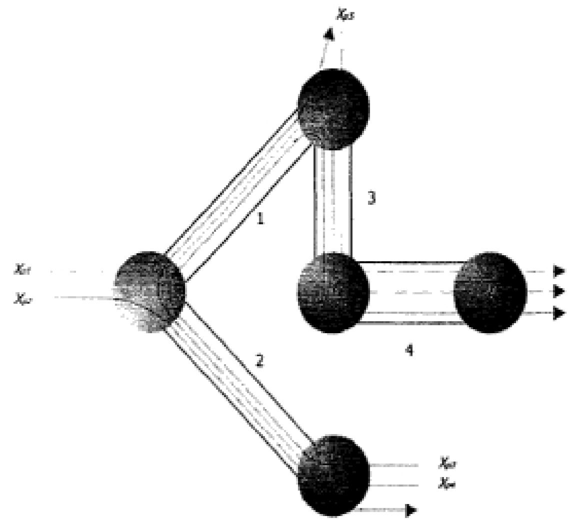


**Figure 4.1. Network used with his respective paths.**

At the beginning the network must know the type of traffic so that it can allocate the necessary bandwidth and the best path depending on the state network. The way how the packets can be distinguished is discussed in [4].

We have assumed that two classes of traffic will be transported through the network and for this, we have taken two values of QoS, one for each class, where all the links have the same QoS. These values are in relation to the delay necessary per type of service, and these are given in milliseconds. The values chosen were taken from the recommendation of ITU G.114 where this recommendation considers a good quality values between 0 and 200 msec., values of satellital quality lie among 200 msec., and 500 msec. For our purpose we have taken 30 msec., for class 1 traffic and 20 msec., for class 2 traffic. Then our parameter of QoS is the delay per packet and class in the system.

To do the analysis we have proposed a link capacity of 10 Mbps, and proceeded to calculate $\mu_{ijk}$, the average service rate of class $k \in K$ in link $(i, j) \in \mathcal{L}$.

Assuming all packets have the same length distribution, [17], then the mean length packet in bits is ( $L$ ), = (572)(8 bits) = 4576 bits, hence

$$\mu_{ij} = \left( \frac{10Mb}{4576} \frac{1}{\text{sec.}} \right) = 2185 \frac{bits}{\text{sec.}}$$

If we have two classes and three flows per link, $\mu_{ij_k}$ can be

$$\mu_{ij_k} = \frac{2185}{(3)(2)} = 360bps$$

To calculate the range of arrival rates per paths, $\lambda_{ij_k} = X_p L$, where $X_p$ is the arrival rate per path which allow to have three different flows in any link given the capacity of each link,

$$\mu_{ij} \rangle X_p L$$

$$\frac{10Mbps}{6L} \rangle X_p L$$

$$X_p \langle \frac{10x10^6}{6(4576)^2} = 364.53bps$$

In this scenario we assumed that all the links have the same capacity.

Based on the previous calculations the values of the arrival rates ($X_p$) for each path and class for the network in Figure 4.1 are shown in the Table, (4.1).

**Table 4.1. Values of arrival rate per class per time unit.**

| $k$ | $X_{1k}$ | $X_{2k}$ | $X_{3k}$ | $X_{4k}$ | $X_{5k}$ |
|---|---|---|---|---|---|
| *Class 1* | 200 | 200 | 370 | 130 | 225 |
| *Class 2* | 170 | 290 | 189 | 199 | 286 |

## 4.2 Optimization method

Once established the parameters involved in the revenue function the problem becomes an optimization problem. Since we are interested in improving the performance of the network it is necessary to maximize the revenue function defined in Equation (3.8).

In order to know where the network attains the maximum revenue with respect to $\lambda$, we need to solve the following optimization problem,

$$\begin{aligned} &\max_{\lambda} \quad R \\ &\text{subject to} \\ &\quad \lambda_{ijk} \leq \mu_{ijk}, \quad \forall\, k \in K, \forall (i,j) \in \mathcal{L} \\ &\quad \lambda \geq 0 \end{aligned} \tag{4.1}$$

where $\lambda$ is the vector that results from the concatenation of the arrival rates of each path flow, $\lambda_{ijk}$ is the arrival rate in link $(i, j)$ for class $k \in K$ and it is limited by the link capacity $\mu_{ijk}$, for all $(i, j) \in \mathcal{L}$ ,$R$ is the revenue in the network given in (3.8), and $\underline{0}$ is the vector of all zeros.

$R$ is a nonlinear function of the arrival rate vector $\underline{\lambda}$, as well as the $\lambda_{ij}$, $\forall (i, j) \in \mathcal{L}$, hence for each link $(i, j)$ and each class $k$, the average number of packets, $N_{ijk}$, and the average delay per packet, $T_{ij}$, are implicit functions of the same vector $\underline{\lambda}$. In order to solve this optimization problem we need to compute the gradient of an implicit nonlinear function of $\lambda$. We solve this computationally by the BFGS method, [15], where an estimate of the Hessian of the Lagrangian is updated at each iteration.

This algorithm is useful to minimize or maximize a nonlinear function of $n$ variables, $\min f(x)$ subject to the simple bounds $l \leq x \leq u$, where the vectors $l$ and $u$ represent lower and upper bounds on the variables and we chose it since we counted with the Matlab software which has this algorithm.

There are other forms to solve an optimization problem, such as, the finite differences which is derived from Taylor series expansions, exact and approximate methods where given a pattern and a text to search, it will return the starting position of the first exact match, should it exist; and simulated annealing methods where a function is evaluated until the optimal value is found, etc., but we chose the BFGS by the reasons mentioned above.

## 4.3 Basic Scenario

The network scenario described in Section 4.1 was evaluated varying the arrival rate for $X_{p1}$ class 1 in increments of 20 packets per second in order to compare the values of the arrival rates once they have been optimized.

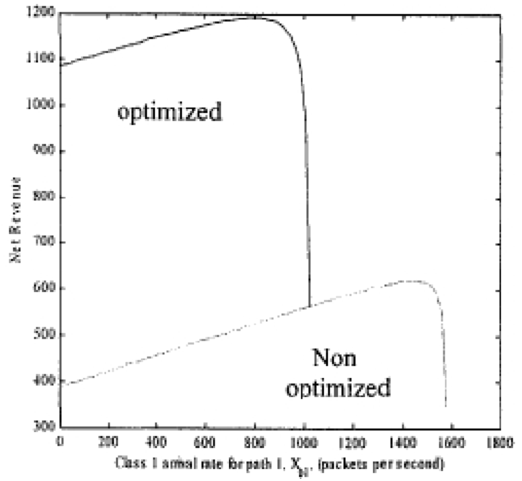


**Figure 4.2. Revenue as a function of external arrival rate for optimized and non-optimized cases.**

Figure 4.2 shows the revenue function *R* calculated using Equation (3.8) for optimized and non-optimized scenario. In the optimized scenario, the optimization problem in Equation (4.1) was solved first and then the external arrival rates at the solution were kept constant, except $X_{11}$, which was varied as explained, to obtain the figure. It can be seen that the optimization increases net revenue considerably over non-optimized cases. The non-optimized case was evaluated using the arrival rates in Table 4.1

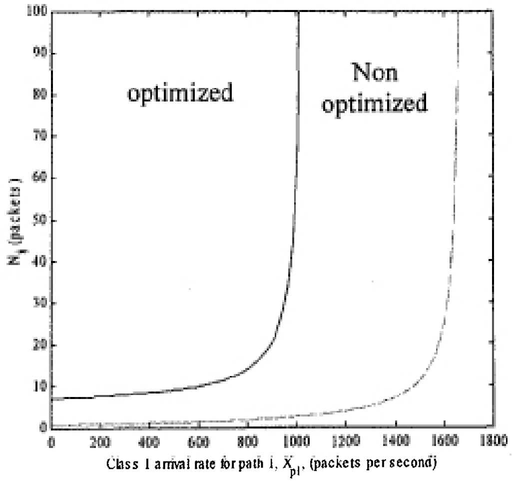


**Figure 4.3. Average number of packets in link 1 for optimized and non-optimized cases**

In Figure 4.3 we compare the average number of packets in link 1, $N_1$ calculated with (3.3). We notice that the average number of packets is greater using the optimized arrival rates in the interval from 0 to 1000 packets per second and it is possible to observe an asymptote in both cases, this is caused by the link arrival rate $\lambda_{ijk}$ being greater than the link capacity $\mu_{ijk}$ and the relation of Equation (3.2) being broken. Since the objective of the revenue function is to earn $w_{ijk}$ units for each packet accepted into the network, it will have more packets on the average than a non-optimized case.

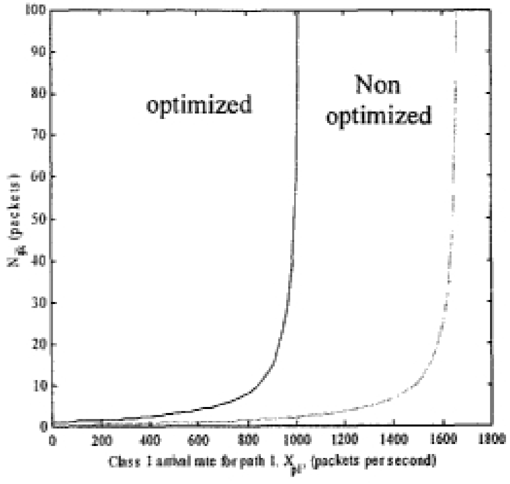


**Figure 4.4. Average number of packets for class 1 in link 1 for optimized and non-optimized cases.**

In Figure 4.4 we present the average number of packets for class 1 in link 1, $N_{11}$ calculated with (3.2) whose arrival rate is varied. We can see that $N_{ijk}$ is almost identical to $N_{ij}$ in Figure 4.3. The asymptotic behavior is also the same. This occurs since the average number of packets for other classes in link 1 are kept almost constant since their arrival rates are not varied.

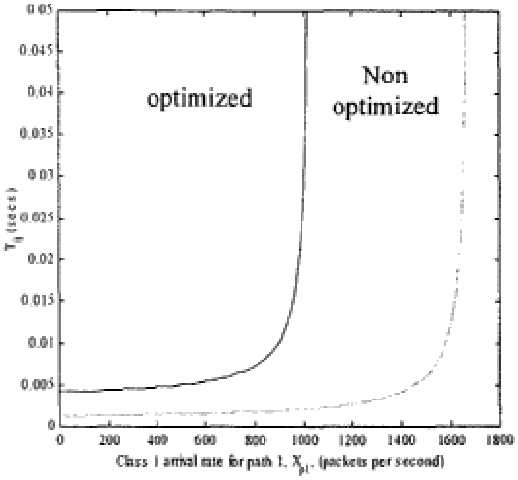


**Figure 4.5. Average delay in link 1 for optimized and non-optimized cases**

In order to observe how the delay is presented we plotted the same arrival rate versus the delay in link 1, $T_l$ calculated with (3.6), see Figure (4.5), and the delay with only the class belonging to class 1, $T_{ll}$ calculated with (3.5) (Figure 4.6), the delay using the optimized arrival rates is greater since we are accepting more packets into the network, but it is less than the QoS required.

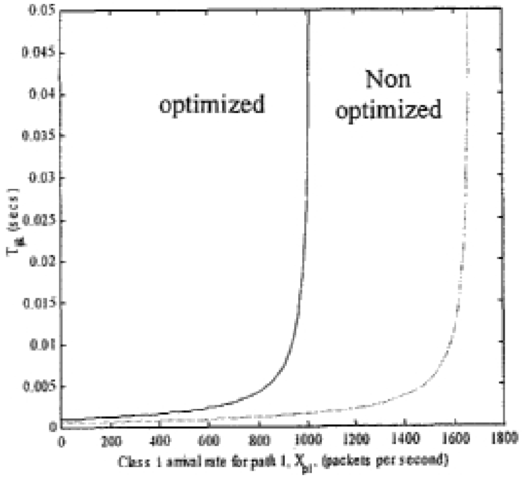


**Figure 4.6. Average delay in class 1 in link 1 for optimized and non-optimized cases.**

In order to observe the advantage of having the values of the arrival rate optimized and non optimized and how values allow to the service provider calculate the flow of each path and the rebounds in his / her revenue, we increase the value of $X_{p1}$ above the permissible value for link 1 class 1, see Table 4.2.

**Table 4.2. Exceeding the capacity in the link 1 class 1**

| *k* | $X_{1k}$ | $X_{2k}$ | $X_{3k}$ | $X_{4k}$ | $X_{5k}$ |
|---|---|---|---|---|---|
| *Class 1* | 1020 | 764.75 | 381.1624 | 764.75 | 764.75 |
| *Class 2* | 764.75 | 764.75 | 341.9318 | 764.75 | 764.75 |

**Table 4.3. Rebound in the relationship QoS - Delay**

| $Q_{ijk} - T_{ijk}$ | | |
|---|---|---|
| *Class 1* | *Class 2* | *Link* |
| -0.0224 | 0.0168 | 1 |
| 0.0264 | 0.0168 | 2 |
| -0.0224 | 0.0168 | 3 |
| -0.0224 | 0.0168 | 4 |

Since some values can saturate the link in a while determined the optimization is becoming very useful so as to know how to allocate the different arrival rate per application and class in an equitable form not allowing that the different arrival rates cause a delay that breaks the values specified in the QoS contract. Taking the values from Table 4.2 it can be observed that the arrival rate for the $X_{p1}$ class1 causes a relationship negative in the penalty in the links crossed by that arrival rate, see Table 4.3.

Link 2 is not altered since $X_{p1}$ does not cross that link.

We carried out another evaluation to check if it was possible to control the network by the variation of the parameters $w_{ijk}$ and $c_{ijk}$ so as to observe if the prices can manage the flow of the paths, we realized this experiment in link 1 since this link is traversed by three of the main flows, $X_{p1}$, $X_{p3}$ and $X_{p4}$. The results obtained are shown in the Table 4.2,

**Table 4.4. Optimized arrival rates by varying $c_{ij}$ in link 1.**

| $k$ | $w_{ij}$ | $c_{ij}$ | $X_{1k}$ | $X_{2k}$ | $X_{3k}$ | $X_{4k}$ | $X_{5k}$ | $Q_{1k}-T_{1k}$ | $R$ |
|---|---|---|---|---|---|---|---|---|---|
| Class 1 | 1 | 0.3 | 764.75 | 764.75 | 401.8055 | 764.75 | 764.75 | 0.0261 | 963.52 |
| Class 2 | 2 | 0.4 | 764.75 | 764.75 | 384.071 | 764.75 | 764.75 | 0.0168 | |
| Class 1 | 1 | 1.5 | 764.75 | 764.75 | 381.1624 | 764.75 | 764.75 | 0.0264 | 1191.3 |
| Class 2 | 2 | 2.5 | 764.75 | 764.75 | 341.9318 | 764.75 | 764.75 | 0.0168 | |
| Class 1 | 1 | 25 | 764.75 | 764.75 | 317.1008 | 764.75 | 764.75 | 0.027 | 5216.7 |
| Class 2 | 2 | 35 | 764.75 | 764.75 | 249.1008 | 764.75 | 764.75 | 0.0175 | |

In order to achieve this we fixed $w_{ij}$ and varied $c_{ij}$, where $X_{1k}$, $X_{2k}$, $X_{4k}$ and $X_{5k}$ reached the values permitted by the link capacity (10 Mbps) and this effect is the same in all the cases, the only value that presents a change is $X_{3k}$ and it is the flow that crosses all the links. When we decrease the cost $c_{ij}$ with respect to $w_{ij}$ this flow could accept more packets, on the other hand, when the cost increases, the arrival rate $X_{3k}$ decreases for this link. For the same link we fixed $c_{ij}$ and varied the values of $w_{ij}$ as shown in Table 4.3,

**Table 4.5. Optimized arrival rates by varying $w_{ij}$ in link 1.**

| $k$ | $w_{ij}$ | $c_{ij}$ | $X_{1k}$ | $X_{2k}$ | $X_{3k}$ | $X_{4k}$ | $X_{5k}$ | $Q_{ij}-T_{ij}$ | $R$ |
|---|---|---|---|---|---|---|---|---|---|
| Class 1 | 0.1 | 1.5 | 764.75 | 764.75 | 361.9253 | 764.75 | 764.75 | 0.0261 | 1175.1 |
| Class 2 | 0.2 | 2.5 | 764.75 | 764.75 | 312.0033 | 764.75 | 764.75 | 0.0171 | |
| Class 1 | 1 | 1.5 | 764.75 | 764.75 | 381.1624 | 764.75 | 764.75 | 0.0264 | 1191.3 |
| Class 2 | 2 | 2.5 | 764.75 | 764.75 | 341.9318 | 764.75 | 764.75 | 0.0168 | |
| Class 1 | 5 | 1.5 | 727.1647 | 737.0008 | 575.896 | 747.9416 | 670.9659 | 0.0255 | 1261.3 |
| Class 2 | 8 | 2.5 | 689.6553 | 725.1612 | 543.4627 | 764.75 | 671.7417 | 0.0147 | |

When the utility per link, $w_{ijk}$ is too small compared to the cost $c_{ijk}$ in the same link, the rate accepted by the network for the different $X_1$, $X_2$, $X_4$ and $X_5$ is kept constant and $X_3$ is the only one that varies, where its values for each class decrease.

We must mention that in both cases the relation $Q_{ij} - T_{ij}$ is positive, therefore we will not have a loss in revenue when the costs are varied in this form.

In the tables 4.4 and 4.5 we carried out the same variations in other link, (link 4), to observe if the effect would remain with no change regardless of the link.

The experiment was made in link 4 and the flows that cross that link are $X_{p1}$, $X_{p3}$ and $X_{p5}$, where the results can be observed in Table 4.4,

**Table 4.6. Optimized arrival rates by varying $c_{ij}$ in link 4.**

| $k$ | $w_{ij}$ | $c_{ij}$ | $X_{1k}$ | $X_{2k}$ | $X_{3k}$ | $X_{4k}$ | $X_{5k}$ | $Q_{4k} - T_{4k}$ | $R$ |
|---|---|---|---|---|---|---|---|---|---|
| Class 1 | 1 | 0.3 | 764.75 | 764.75 | 404.2586 | 764.75 | 764.75 | 0.0261 | 963.52 |
| Class 2 | 2 | 0.4 | 764.75 | 764.75 | 381.4706 | 764.75 | 764.75 | 0.0164 | |
| Class 1 | 1 | 1.5 | 764.75 | 764.75 | 381.1624 | 764.75 | 764.75 | 0.0264 | 1191.3 |
| Class 2 | 2 | 2.5 | 764.75 | 764.75 | 341.9318 | 764.75 | 764.75 | 0.0168 | |
| Class 1 | 1 | 25 | 764.75 | 764.75 | 317.9486 | 764.75 | 764.75 | 0.027 | 5216.3 |
| Class 2 | 2 | 35 | 764.75 | 764.75 | 249.1008 | 764.75 | 764.75 | 0.0175 | |

The results followed the same pattern when the cost was varied in this link, i.e., only $X_{3k}$ does not reach the maximum value so as to have all the arrival rates at their maximum capacity.

Table 4.5 presents the case when the cost is greater than the expected utility per service and it is a hypothetical case since $c_{ij}$ can never be smaller than $w_{ij}$.

**Table 4.7. Optimized arrival rates by varying $w_{ij}$ in link 4.**

| $k$ | $w_{ij}$ | $c_{ij}$ | $X_{1k}$ | $X_{2k}$ | $X_{3k}$ | $X_{4k}$ | $X_{5k}$ | $Q_{4k}-T_{4k}$ | $Q_{4k}-T_{4k}$ |
|---|---|---|---|---|---|---|---|---|---|
| Class 1 | 0.1 | 1.5 | 764.75 | 764.75 | 361.9253 | 764.75 | 764.75 | 0.0266 | 1175.1 |
| Class 2 | 0.2 | 2.5 | 764.75 | 764.75 | 312.0033 | 764.75 | 764.75 | 0.0171 | |
| Class 1 | 1 | 1.5 | 764.75 | 764.75 | 381.1624 | 764.75 | 764.75 | 0.0264 | 191.3 |
| Class 2 | 2 | 2.5 | 764.75 | 764.75 | 341.9318 | 764.75 | 764.75 | 0.0168 | |
| Class 1 | 5 | 1.5 | | | | | | | |
| Class 2 | 8 | 2.5 | | | | | | | |

The empty columns are due to the violation of the constraint $\mu_{ij} > \lambda_{ij}$ causing a division per zero in Equation (3.2).

We proceed to calculate the average delay per class $k$ in the connection taking the values of the tables 4.3 and 4.4 respectively in order to observe if the variation of the prices and costs can decrease the delay per flow such as the service provider can take a decision of how to put the variables $c_{ij}$ and $w_{ij}$ so as to manage the delay in the system.

**Table 4.8. Average delay per class $k$ in the connection varying $w_{ij}$.**

| *Flow* | *Class* | $w_{ij}=0.1, 0.2$<br>$c_{ij}=1.5, 2.5$<br>$T_k$ | $w_{ij}=1, 2$<br>$c_{ij}=1.5, 2.5$<br>$T_k$ | $w_{ij}=5, 8$<br>$c_{ij}=1.5, 2.5$<br>$T_k$ |
|---|---|---|---|---|
| 1 | 1 | 0.000471 | 0.000468 | 0.0011 |
| 1 | 2 | 0.000480 | 0.000474 | 0.0010 |
| 2 | 1 | 0.000157 | 0.000156 | 0.000388 |
| 2 | 2 | 0.000160 | 0.000158 | 0.000146 |
| 3 | 1 | 0.000232 | 0.000245 | 0.000413 |
| 3 | 2 | 0.000220 | 0.000218 | 0.000390 |
| 4 | 1 | 0.000314 | 0.000312 | 0.000300 |
| 4 | 2 | 0.000320 | 0.000316 | 0.000317 |
| 5 | 1 | 0.000314 | 0.000312 | 0.000256 |
| 5 | 2 | 0.000320 | 0.000316 | 0.000261 |

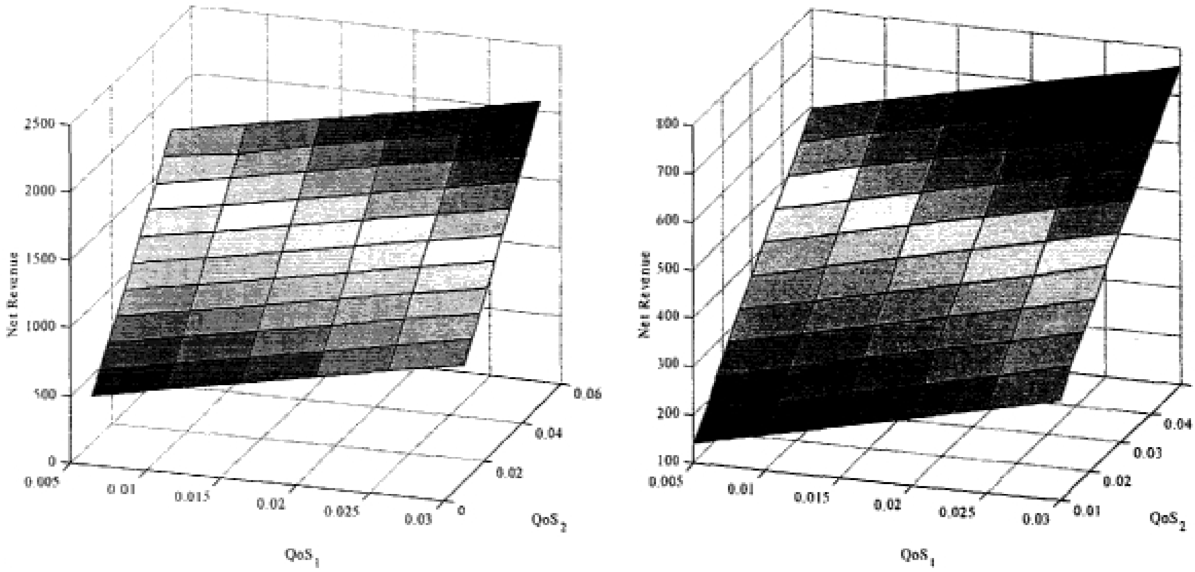
**Table 4.9. Average delay per class *k* in the connection varying $C_{ij}$.**

| *Flow* | *Class* | $w_{ij} = 1, 2$<br>$c_{ij} = 0.3, 0.4$<br>$T_k$ | $w_{ij} = 1, 2$<br>$c_{ij} = 1.5, 2.5$<br>$T_k$ | $w_{ij} = 1, 2$<br>$c_{ij} = 25, 35$<br>$T_k$ |
|---|---|---|---|---|
| 1 | 1 | 0.000466 | 0.000468 | 0.0011 |
| 1 | 2 | 0.000469 | 0.000474 | 0.0010 |
| 2 | 1 | 0.000155 | 0.000156 | 0.000386 |
| 2 | 2 | 0.000156 | 0.000158 | 0.000146 |
| 3 | 1 | 0.000262 | 0.000245 | 0.000413 |
| 3 | 2 | 0.000245 | 0.000218 | 0.000390 |
| 4 | 1 | 0.000310 | 0.000312 | 0.000300 |
| 4 | 2 | 0.000312 | 0.000316 | 0.000317 |
| 5 | 1 | 0.000310 | 0.000312 | 0.000256 |
| 5 | 2 | 0.000312 | 0.000316 | 0.000261 |

From Tables 4.8 and 4.9 the service provider can take a decision if he/ her can decrease the delay per flow such as he /her does not have problems with a possible loss in the revenue caused by the penalty of the function. It is possible observe that if $c_{ij}$ is diminished in relation with $w_{ij}$, the delay decreases although no in a significant form.

## 4.4 Quality of Service variation

Since the delay in each link is the only parameter than can cause a decrease in revenue and this penalty will take place if the delay at any link is greater than the necessary QoS per service we proceed to vary both qualities to observe and to compare the revenue when the QoS is varying.



**optimized**

**Non optimized**

**Figure 4.7. Revenue as a function of QoS for each class.**

In Figure 4.7 we present the relationship between the revenue calculated using (3.8), and the QoS for the optimized case (left) and the non optimized case (right), they were made varying the QoS for both classes and were compared by using the cost function at the moment of being optimized and we can see that when the QoS is too strict, the revenue takes smaller values, on the other hand, when the QoS is not so strict, the revenue reaches his maximum value.

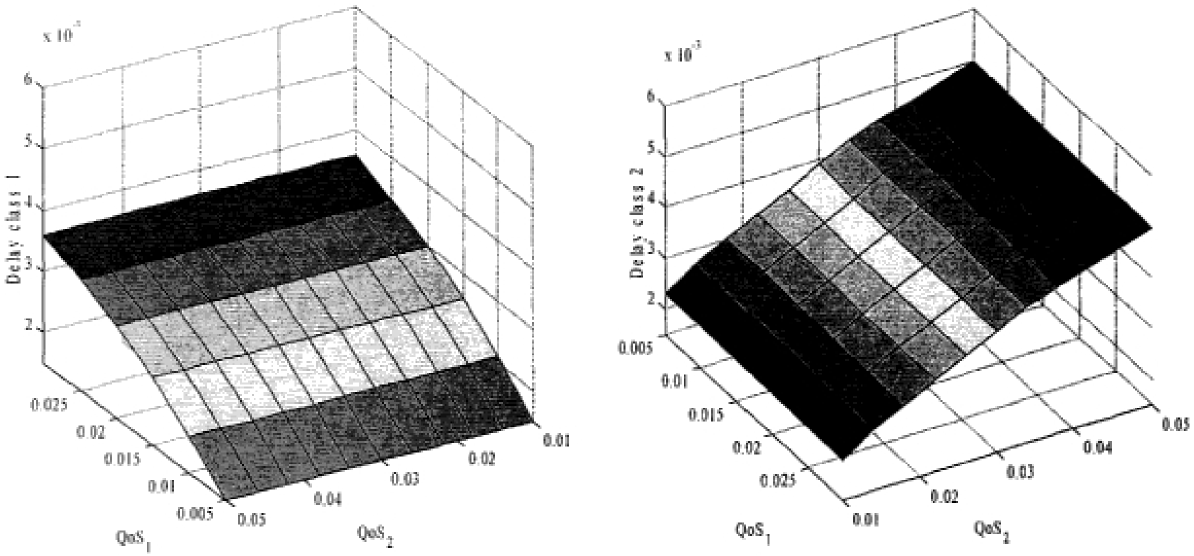


**Figure 4.8. Average delay in the network for each class.**

We separated the delay in the entire network by observing how the QoS rebounds in the delay per class. Since the QoS's are independent of each other, in Figure 4.8 we can observe that by varying only the QoS of class 1 the delay in class 1 can be affected. When both QoS are too strict the delay in the network is smaller since the traffic in the network is less. The same effect occurs when the QoS of class 2 is being varied at the delay in class 2.

In Figure 4.9, we made the same variations and we compared to the sum of arrival rates. The result was the same because when the QoS was too strict and the values of the arrival rates were smaller.

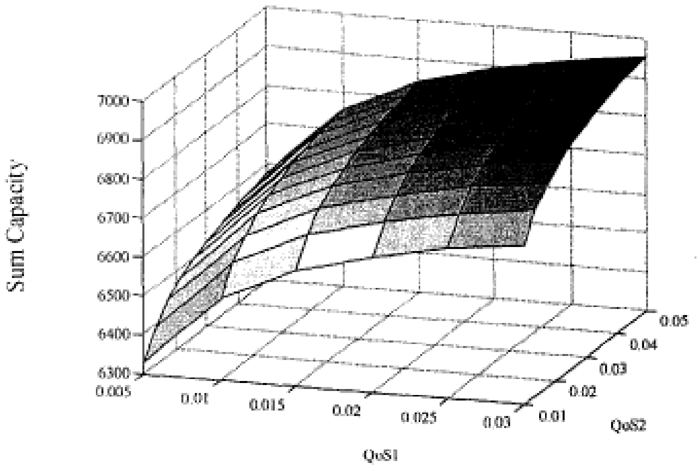


**Figure 4.9. Carried traffic in the network as QoS per class is varied**

Figure 4.9 is important, since it tells us the maximum amount of traffic that can be accepted in the entire network given QoS parameters such that revenue is maximized. This surface can be seen as the constrained traffic capacity of the network since more than that will not be achieved without violating some constraints in the optimization problem.

# Chapter 5

# Conclusions and Future work

In Chapter 4 we observed and compared the results using our revenue cost optimized and non optimized, where we can notice the next,

- The use of the revenue cost proposed tell us where exactly the level of traffic that produces the maximum revenue in the network is.
- Logically if the arrival rates are in their maximum capacity the number of packets increases and they can saturate the link capacity very quickly.
- It must be distinguished the type of traffic before establishing the connection so as to take the considerations of bandwidth and delay required for giving a good service.
- When we tried to handle the flow paths varying the cost and prices we observed that it was possible to manage any arrival rate, in our specific case ( $X_{3k}$) which was the arrival rate that used more links and the rest of the path flows reached the value permissible by the link capacity.
- The service provider must take care of the values of each QoS for each class since if he/she does not watch this relationship carefully he/she will have a severe lost in revenue.

During the formulation and evaluation of our model emerging some ideas emerged that can improve it and extend the work, these ideas are mentioned next,

- Consider other models for the nodes different from the M/M/1 model, this model needs to have a finite buffer capacity since the QoS depends on buffer size in a great measure.
- The arrival rates can have another class of distribution since recently researches have demonstrated that the flows in the Internet may be modeled like self-similar traffic processes.

- There must exist a parameter that helps measure the performance in the nodes such as blocking probability.

- It is necessary to establish a kind of routing that takes into account the parameters used in the revenue cost, such as- the value of QoS for each class, the cost and prices for using the links, and the delay and number of packets in it-.

# Bibliography

[1] Bertsekas D., Gallager R., *Data Networks*, Prentice Hall, 2nd. Edition, 1992.

[2] Black Uyless, *TCP / IP & Related Protocols*, McGraw Hill, 2nd. Edition, 1995.

[3] Clark D., "Combining Sender and Receiver Payments in the Internet", 1996.

[ 4] Cocchi Ron, Estrin Deborah, "A study of priority in multiple service class networks".

[5] Courcoubetis C., Vasilios A., "An approach to pricing and resource sharing for available bit rate (ABR) services", 1998.

[6] Courcoubetis C., Vasilios A., " An evaluation of pricing schemes that are based on effective usage", ICC 1998.

[7] Kleinrock L.,Communication Nets: Stochastic Message Flow and Delay, Mc Graw-Hill, 1964.

[8] Mackie-Mason J. K., Varian H.R., "Pricing the Internet".

[9] Mackie-Mason J. K., Varian H. R., "Some FAQs about Usage-Based Pricing", 1994.

[10] Murphy J., Murphy L., "Bandwidth allocation by pricing in ATM networks".

[11] Stiller B., "Accounting, Charging and Pricing of Communication Services", 1998.

[12] Shenker S., Clark D., "Pricing in Computer Networks: Reshaping the Research Agenda"

[13] Parris Colin, "A Framework for the Study of Pricing in Integrated Networks".

[14] Paxson V., Floyd S., "Why we don't know how to simulate the Internet", 1997.

[15] Reklaitis G., Ravindran A., *Engineering optimization: methods and applications*, 2 nd. Edition, 1983.

[16] Sairamesh J., Ferguson D., " An approach to pricing, Optimal allocation and Quality of Service provisioning in high-speed packet networks", 1995.

[17] Stallings W, *Local and Metropolitan Area Networks*, Macmillan, 4th. Edition, 1993.

[18] Veciana G., Baldick R., "Pricing Multi-service networks", 1994.