Pricing in a Multiple ISP Environment with Delay Bounds and Varying Traffic Loads

by

Sameh Gabrail

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2008

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

In this thesis, we study different Internet pricing schemes and how they can be applied to a multiple ISP environment. We first take a look at the current Internet architecture. Then the different classes that make up the Internet hierarchy are discussed. We also take a look at peering among Internet Service Providers (ISPs) and when it is a good idea for an ISP to consider peering. Moreover, advantages and disadvantages of peering are discussed along with speculations of the evolution of the Internet peering ecosystem. We then consider different pricing schemes that have been proposed and study the factors that make up a good pricing plan. Finally, we apply some game theoretical concepts to discuss how different ISPs could interact together. We choose a pricing model based on a Stackelberg game that takes into consideration the effect of the traffic variation among different customers in a multiple ISP environment. It allows customers to specify their desired QoS in terms of maximum allowable end-to-end delay. Customers only pay for the portion of traffic that meet this delay bound. Moreover, we show the effectiveness of adopting this model through a comparison with a model that does not take traffic variation into account. We also develop a naïve case and compare it to our more sophisticated approach.

Acknowledgements

First, I would like to express my sincere gratitude to my supervisor Professor Ravi R. Mazumdar, for taking me on to be his student. I am thankful for his constant guidance, support, and encouragement. I have learned a lot from him not just academically but on a personal level as well. He has taught me how to approach a problem and how to work my way to solving it with confidence. He has given me insight on how to make this thesis more realistic. As with regards to his teaching, he has a sincere desire to teach us the newest topics in the literature. His class is unique in stimulating and challenging the thoughts of the listeners. It is truly a privilege to work under his supervision.

Secondly, I would like to thank my parents for their love, support, and encouragement to carry on with this work. I am indebted to them for all their sacrifices for me through out the years to become who I am today.

Furthermore, I would like to thank Professor Catherine Rosenberg along with all my fellow graduate students in our group for sharing their research work and preparing informative presentations that benefited me personally.

Finally, I would like to thank my Lord Jesus Christ for making all of this possible. I acknowledge that in Him are hidden all the treasures of wisdom and knowledge. (Colossians 2:3)

Dedication

I would like to dedicate this thesis to my Lord and Savior, Jesus Christ, and to my parents, Faten and Yousry.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Pricing the Internet has always been a hot topic. After all, what good is a network if you can not make profit out of it! Internet pricing has taken many turns since the early stages of the Internet itself. There are many arguments nowadays as to whether the current pricing schemes are adequate or not.

After investigating and studying various pricing schemes for the Internet, it turns out that coming up with a good pricing plan is a non-trivial task. Some questions come to mind such as what does the Internet structure look like? What factors should be taken into consideration when designing a good pricing scheme? Once we placed a good and efficient pricing scheme, how does one use this model in a multiple ISP environment such that all ISPs maximize their profit in a fair manner?

Andrew Odlyzko in [12] argues that pricing the Internet follows the same pattern of evolution as other communication technologies. He says that the history of snail mail, telegraph, and telephones present a consistent pattern in their pricing evolution. First quality rises which then leads to price decrease, then usage starts to elevate, hence total profits increase, and finally pricing models become much simpler. Odlyzko is in favor of the flat rate pricing scheme, because it meets customers' likings for being simple and predictable. Moreover, it stimulates and encourages more usage.

On the other hand, Jörn Altmann and Karyen Chu in [1], argue that a problem with flat-rate pricing is that it is very stressful for the network. The result would be an overall reduction in network performance. They continue to elaborate the fact that users do not face the true marginal cost of their usage. Nevertheless, they also state some disadvantage of other pricing schemes such as pure per-minute pricing plans. In their paper [1], they acknowledge the fact that QoS should be included in coming up with good pricing schemes, and they propose such a pricing scheme. Furthermore, Burkhard Stiller et al in [17], state the same argument of the need for pricing and charging for more than a single service class. They say that flat-rate pricing should be revised and they propose another pricing scheme, which they call the cumulus pricing scheme.

Peering between local ISPs is also a hot issue for several reasons. Transit costs are reduced when peering, ISPs have more control over routing, path redundancy occurs, and QoS is enhanced due to less latency. Although peering is becoming more popular, it cannot be a substitute for transit services. We will discuss various peering issues in this thesis and what challenges are faced by ISPs to establish peering.

Much research has been conducted on how Internet Service Providers (ISPs) interact in a multiple ISP domain. Researchers have often used Game Theoretical principles to ensure fair dealings among these providers while maximizing their profits at the same time. Many models have been suggested, some proposed that ISPs be the sole players of the game, whereas others included customers as well. Some models introduced different hierarchies for players, and others introduced the leader-follower game. In this thesis we will take a look at some of these proposals, choose one, and extend it to meet our needs.

The thesis is organized as follows: Chapter 2 investigates some background and related work. It begins with basic definitions and terminology for the Internet peering ecosystem. It shows the hierarchy in the Internet and defines different classes of players. Also, peering issues, such as deciding when to consider peering, and the evolution of the peering ecosystem are discussed. Next we discuss what requirements should be taken into consideration when placing a pricing plan. Furthermore, various pricing plans are presented along with a description of each and how these pricing schemes compare with the requirements. We then discuss how to apply a pricing scheme in a multiple ISP environment.

In Chapter 3, we introduce the formulation of our pricing model. Assumptions are laid along with the game framework which consists of the players of the game and their utility functions. A detailed explanation of the formulation of the game is also presented.

In Chapter 4, numerical results are given in the form of tables and figures. Four scenarios of the game as different combinations of customers and ISPs are discussed. At the end of this chapter, we develop a naïve scenario to see if things could be done in an easier way. Furthermore, we show a comparison between our model and a model developed by [14], and explain how our model is more suitable.

The thesis ends with Chapter 5 giving conclusions and some suggested future work. Appendix A, gives a pseudocode for our program for the scenario of two ISPs serving two customers. Finally, Appendix B presents a brief explanation of Nash equilibrium used in our program.

Chapter 2

Background and Related Work

2.1 THE INTERNET PEERING ECOSYSTEM

The Internet is a diverse body consisting of millions of networks, which further comprise many network devices such as routers, switches, servers, workstations, etc. These networks are operated by a number of network operators, content providers, and end users that are called "Internet Players" by [11].

2.1.1 Basic Definitions and Terminology

The Internet Peering Ecosystem is defined by [11] as a population of networks connected by means of a protocol stack, which is provided by the Internet Engineering Task Force. The network operators interact and interconnect in different business affairs.

The Global Internet Peering Ecosystem consists of different Internet Regions. An Internet Region is usually defined as a part of the Internet that is governed by certain geographical boundaries. For an Internet Service Provider (ISP) to be in contact with the rest of the Internet, it requires transit or peering relationships within an Internet Region. Transit ISPs sell access to the rest of the Internet, whereas peering is a connection between two ISPs in which they both provide access to one another's customers free of charge.

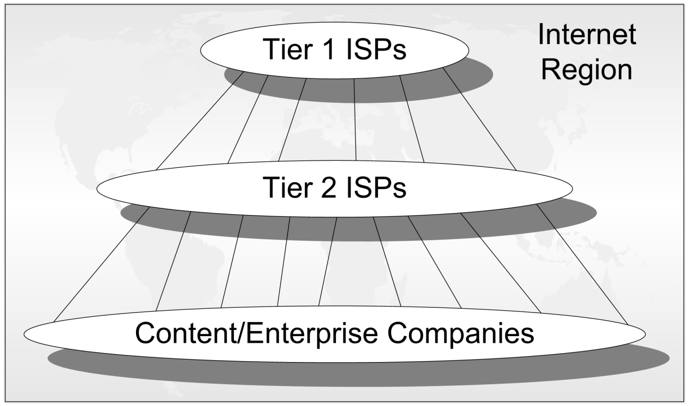


Figure 2.1: Classes of Internet Players within an Internet Region.

In a Peering Ecosystem three classes of Internet players are defined: Tier 1 ISPs, Tier 2 ISPs, and Content Providers/Enterprise Companies (shown in Figure 2.1). These definitions are given in detail in the next few sections according to [11].

Tier 1 ISP

Tier 1 ISPs are considered to be at the top of the hierarchy of the Internet model. These ISPs have access to the entire Internet Region through peering relationships; hence they do not pay transit fees. All other ISPs located in this Internet Region need to purchase transit from those Tier 1 ISPs in order to have a connection to the Internet Region and to the outside world as well.

Tier 1 ISP peering is done only to provide accessibility to their customers, because they do not pay transit fees. Figure 2.2, which was taken from [11], gives an example of a Tier 1 region in the US.

Since some Internet Regions may become very large, therefore to distribute peering load among multiple points, an Interconnect Region is introduced. Figure 2.3 shows an example of eight interconnection regions within one big Internet Region, which is the entire United States.[1]

[1]Figure 2.3 was taken from [11].

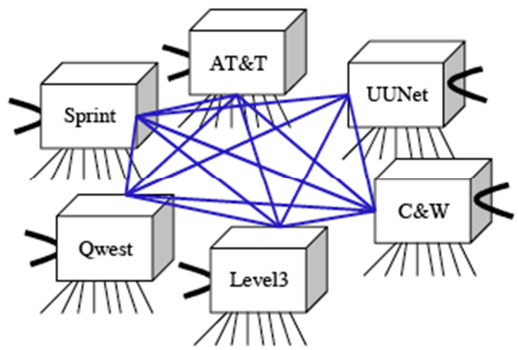


Figure 2.2: An example of a Tier 1 Interconnection Region in the US.

Figure 2.3: Tier 1 Interconnection Regions within an Internet Region (US).

Tier 2 ISP

This type of ISPs buys transit from Tier 1 ISPs and then resells it to either other Tier 2 ISPs or to Content Providers/Enterprise Companies or regular customers.

Tier 2 ISPs usually peer only in Interconnect Regions where they sell services, which generate a diverse community of peering Tier 2 ISPs within each Interconnect Region. As a result there is a vast number of peering policies and peering becomes a local routing optimization problem for each Tier 2 ISP.

There is great cooperation between Tier 2 ISPs based on similar interests. They all tend to buy transit and therefore are interested in peering relationships.

Tier 2 ISPs are often referred to as the "middle man". They do not like anyone to know that they buy transit, because of some competition with Tier 1 ISPs over really big customers. Generally speaking, Tier 1 ISPs do not like to peer with Tier 2 ISPs, because they consider them as potential customers.

Tier 2 ISPs have many incentives to peer such as:

- ***Trim down Transit Costs***

  When ISPs choose to peer, provided that a large enough volume of traffic is being exchanged, it is beneficial, because this reduces the transit costs. The reason there is a requirement that a large volume of traffic is exchanged is to compensate the initial setup costs of a private exchange between the two ISPs. This will be discussed later in detail.

- ***Enhanced Quality of Service***

  Less delay occur when traffic flows directly between ISPs that are in a peering relationship, rather than first sending the traffic to a transit provider that forwards it to the destination ISP. Moreover, probability of packet loss decreases when peering.

- ***Tighter Routing Control***

  It has been noticed that some ISPs prefer to have some control over the route that its traffic traverses. By peering this requirement is fulfilled.

The Content Provider

Content Providers are those companies that run an Internet service but do not sell transit. Practical examples of Content Providers are eBay, Amazon.com, and Staples.com. Examples of Enterprises include Agilent, Hertz, Avon, and General Electric.

Content Providers purchase transit to deliver their content to their online users. They are very similar in their function to the Tier 2 ISPs in the Peering Ecosystem. However, they do not sell transit, do not focus on operating a network, and do not manage peering relationships. Furthermore, the ordinary user would also fall under the content provider's category.

In general there are no peering relationships among content providers. This is due to the fact that they either compete with each other or they have no interest in one another's business.

2.1.2 Advantages and Disadvantages of Peering

Before we list some of the advantages and disadvantages of peering (mostly stated by [10]) we would like to note that peering cannot be a replacement of transit by any means. This is, because transit provides access to the entire Internet routing table, while peering just provides access to each other's customers [10].

Advantages of Peering

Some of these advantages have been discussed before but we include them here again just for the sake of completion.

- ***Reducing Costs***

  This was discussed before and it is further elaborated in [10] by comparing the transit costs and the peering ones with respect to the volume of traffic. Figure 2.4 shows the tradeoff and the breakeven point. It is obvious that as the amount of traffic increases both the transit and the peering costs decrease. However, transit costs decrease slightly and in far away steps, whereas peering

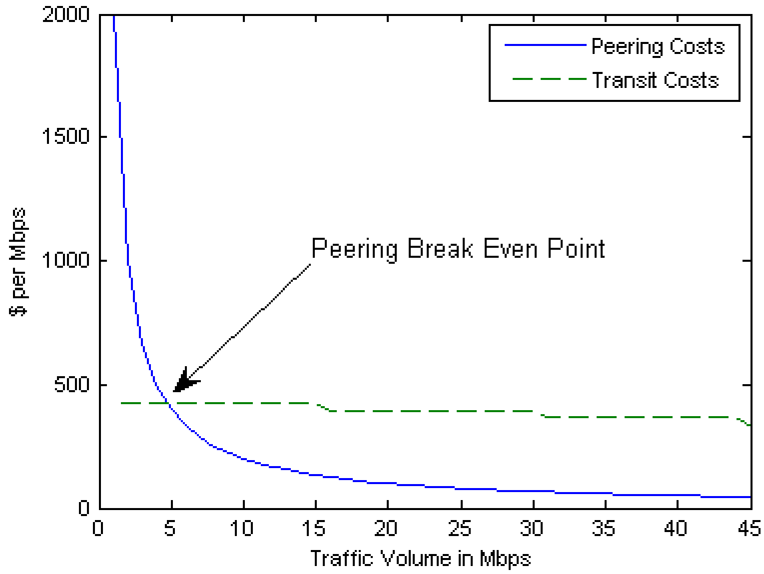


Figure 2.4: Transit Costs and Peering Costs vs. volume of traffic.

costs decrease exponentially, because ISPs are able to cover their initiation setup costs for peering.

- ***Low Latency Path***

  The lowest delay path is provided between ISP customers. Peering improved performance by as much as 40-50 milliseconds as mentioned in [10].

- ***More Control over Routing***

  More flexibility is given to ISPs to route around congested paths. This further enhances the quality of service to the customers.

- ***Providing Redundancy***

  Transit services could serve as a backup plan in the case of a peering session failure.

Disadvantages of Peering

- ***Lack of Expertise***

  Network expertise is needed for peering unlike just buying transit from a higher Tier ISP.

- ***Administrative Startup Costs***

  Peering needs a lot of negotiations between legal departments of two ISPs, which becomes tedious.

- ***Not Always Granted***

  Some ISPs have unpublished peering conditions that allow only the largest ISPs to peer with them.

- ***Slow Process***

  It may take several months to get a peering relationship up and going.

2.1.3 Evolution of the Peering Ecosystem

As a case study for the U.S. Internet Peering Ecosystem, the author of [11] has found out that there are some dramatic changes that happened to the Peering Ecosystem over the past few number of years. In this section we will take a brief look at those changes and their causes.

The Economic Breakdown of the Telecom Sector in 1999/2000

As a result of that, many Tier 1 ISPs went bankrupt, which in turn caused Internet players to make sure that their businesses and financial status is well. From that time, the economics of peering has been a hot topic.

Used Networking Equipment Market Grows

After the bankruptcy of many companies discussed above, used networking components became available and were much cheaper than their new counterparts. This in turn led the existing ISPs at the time to reduce their capital cost of peering.

Peer-to-Peer File Sharing Networks Grow in Popularity

Peer-to-peer networks have shown a dramatic increase in popularity lately. This has lead to enormous amount of traffic being exchanged among users of the Internet.

Since users leave their PCs on 24/7 to download files such as music or videos, this has become a heavy burden on their service providers. Therefore, ISPs had greater incentives to adopt the peering mechanism since it is cheaper on their part.

As a side note, it was shown [11] that peer-to-peer traffic volume grows when cable companies peer with each other. This is, because software such as Kazaa tends to fetch files across the recently peered network route. Its protocol determines which sharer is more local and automatically selects that sharer to download from. A result of all of this, is that when cable companies start to peer, they should expect a significant increase in their traffic volume. This could be seen from Figure 2.5 taken from the Toronto Internet Exchange (TorIX) usage statistics over a year[2]. This figure shows how usage increases over time.

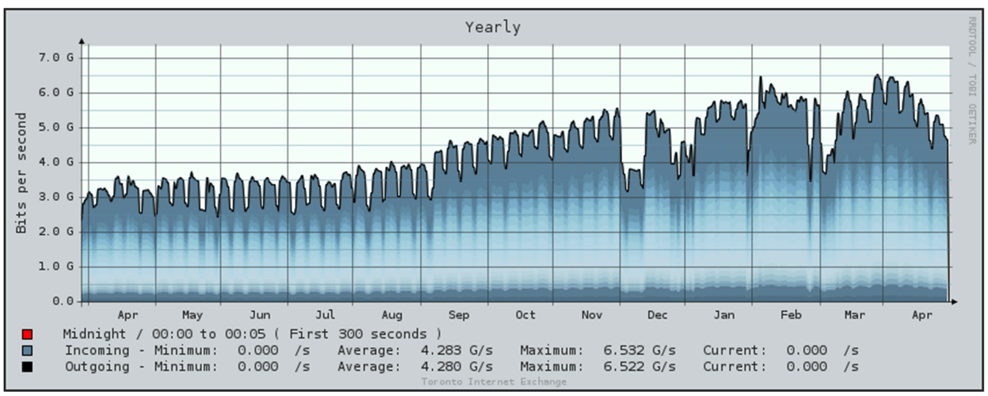


Figure 2.5: Toronto Internet Exchange (TorIX) aggregated switch traffic over a period of a year.

2.2 PRICING SCHEMES

Pricing schemes for the Internet have taken so many forms since the birth of the Internet itself. It turns out that coming up with a pricing plan is a non-trivial task, because a lot of factors must be taken into consideration. The author of [13] gives a nice presentation of the general requirements for a good pricing plan. These requirements include three major groups, which are shown in Figure 2.6. Balancing of these requirements determine the quality and suitability of Internet pricing.

We briefly discuss these requirements below.

[2]This figure was taken from http://www.torix.net/peers.php

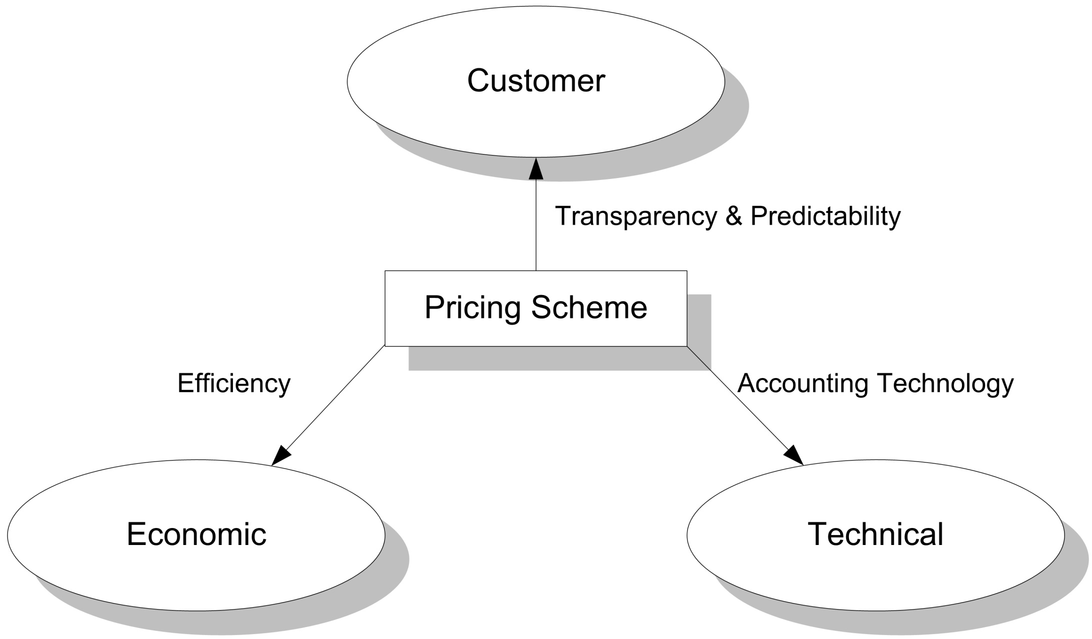


Figure 2.6: General Requirements for a good pricing scheme.

- **Customer:** A preference for customers is a major role in determining a pricing scheme. Transparency and predictability play an important role towards the customers' likings.

- **Economic:** An ISP's goal is to run a network in the most efficient manner. It does so by maximizing its network utility function or the total revenue. This in turn plays an important role in placing prices, because it might indicate the status of the network in terms of congestion. For instance suppose a customer is misbehaving by overusing the network resources, then the ISP may punish this user by raising the prices.

- **Technical:** Accounting and the way an ISP collects data about the customers' usage is a major factor in pricing. This is a very complex problem due to the huge amount of packets traveling through the network. Several proposals have been submitted to reduce this complexity by carefully choosing parameters, classes, and accounting locations.

In the following sections we will take a look at some of the current proposed pricing schemes and see how they conform to the above-mentioned requirements.

2.2.1 Flat-Rate Pricing

Customers are charged a flat rate for unlimited access to the Internet per month. This pricing scheme is most common in North America. This pricing plan applies to both dial-up access service and broadband service. A typical flat rate fee would be around US $20.00 per month for dial-up access and US $10.00-200.00 for broadband access [1]. QoS for broadband access differs by the initial contract between the ISP and the customer, which typically depends on the connection speed (access bandwidth). Users cannot switch to higher connection speeds on demand; they must purchase a higher bandwidth from the ISP, which is again a flat rate.

It has been shown by the INternet Demand Experiment (INDEX)[3] that users prefer flat-rate pricing plans the best. This is, because users already know what their bill will be at the end of the month (pricing stability). Furthermore, they can leave their PCs on 24/7 downloading all sorts of files. On the other hand flat-rate pricing plans are inefficient economically, because users do not face the true marginal cost of usage which leads to over-usage of the network resources. As a result there is a potential of reduction of overall performance of the network due to congestion.

To conclude this discussion, flat-rate pricing satisfies the customer and technical requirement, because it is acceptable to the customer and at the same time easy to monitor by the ISP. However, it does not satisfy the economic requirement due to its inefficiency in handling network resources.

2.2.2 Usage-Based Pricing

There are two main charging schemes within the usage-based pricing. The most common one is the minute-based pricing plan. Customers are charged on a per minute basis. This is most common in Europe where customers have free usage of some minutes, but if they exceed it they are charged per minute. Typical rates vary between 1 and 4 cents within Europe, independent of whether the access technology is telephone or DSL [1]. There is also another type of usage-based pricing which is byte-based. In this scheme the user is charged according to how many bytes of data he/she has downloaded or uploaded.

[3]We give a detailed explanation of INDEX in a subsequent section.

Usage-based pricing encourages users to disconnect from the network when they are not using the service which helps improve the service on the networks side. However, this is inconvenient to the customers for two reasons. First, they can not predict their bill at the end of the month (pricing instability). Second, they are not motivated to spend more time on the Internet, which limits revenues from other sources such as advertisements and retailing. Furthermore, it limits the evolution of the Internet to be a multi-service/multi-purpose network [1].

2.2.3 Diff-Serv Pricing

Regular Diff-Serv

Traffic is classified according to required QoS. The better QoS a customer requires the more he/she has to pay. Such discrimination in prices maximizes the ISP's profit. This kind of pricing is common in other industries such as airplane travel where the price of tickets differ [7].

This type of pricing could be somewhat acceptable to customers, because you pay for exactly what you want. As for the ISP's side it is also beneficial, because not only is the network better off in terms of congestion, but also the ISP's profit is maximized. As for the technical requirement it becomes very complex to monitor the packets of each user.

A Variation of Diff-Serv

A variation of the well-known Diff-Serv pricing scheme has been proposed in [1] and we give a brief explanation of it in this section. The basic idea proposed is to combine both flat-rate pricing plans with usage-based pricing plans in a Diff-Serv manner. To elaborate, users would have unlimited access to the Internet with an initial bandwidth agreed upon in the contract. However, the users can also access higher bandwidths on demand in which they are charged according to a usage-based scheme. This proposal is based on the famous INDEX experiment which took place at the University of California (Berkeley). This experiment was aimed at analyzing user demand for the Internet as a function of QoS, budget, various pricing schemes, and application. The different experiments are:

- ***Minute Pricing Experiments***

  Subjects are charged on a per minute basis for connecting to the Internet at each of five different bandwidths which are 16, 32, 64, and 128 kbps. All subjects have free of charge unlimited access to the Internet at an 8 kbps bandwidth. Subjects have the option of selecting any of the bandwidths at any time. This experiment runs for seven weeks with variation in prices throughout, except that the first week is free of charge to give the subjects an idea of their usage.

- ***Byte Pricing Experiment***

  Subjects still have the 8 kbps bandwidth for free and unlimited access and a 128 kbps bandwidth that is charged according to the number of bytes transmitted. The experiment again runs for seven weeks with per-byte prices changing every week.

- ***Minute-Byte Pricing Experiment***

  Pricing plan is based on different combinations of both minute based and byte based plans. Subject are given the option of selecting a pure per minute plan, a pure per byte plan, or a combination of the two. Subjects also have the option of choosing from six different bandwidths within each pricing plan.

- ***Flat-Rate Buy-Out Option Experiment***

  This experiment is similar to the minute pricing experiment; however, customers have the option of buying out a certain bandwidth with its corresponding price. If the subject were to choose to buy-out a certain bandwidth then this bandwidth could not be changed for the rest of the week. The highest bandwidth's price is chosen randomly and then the prices of the lower bandwidths are taken to be a fraction of the highest one.

In this portion of our discussion we just give the final results of the INDEX project for more details of the experiment please refer to [1]. In Figure 2.7, Altmann and Chu show the average bytes transmitted per user per day for the different experiments, which are from left to right, the free trial weeks, two minute based experiments, the byte based experiment, the minute-byte based experiment, and the flat-rate buy-out experiment. As it is obvious from Figure 2.7 that when the flat-rate buy-out option was available the usage increased significantly to almost reach that of the free weeks' trial!

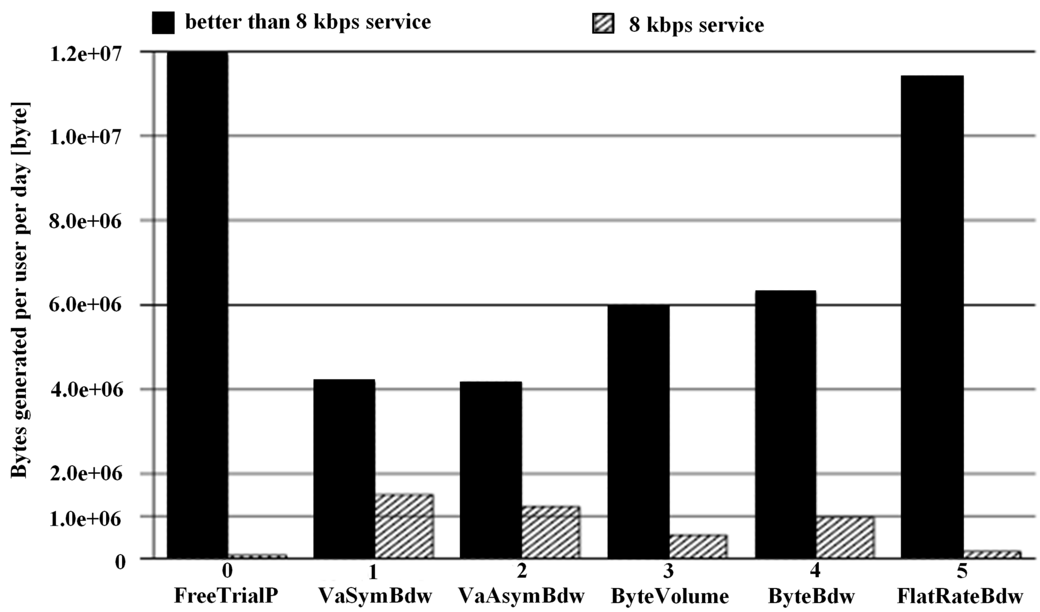


Figure 2.7: Transmitted bytes per user per day vs. different pricing schemes.

More results of INDEX have shown that subjects have spent more money per week during the flat-rate buy-out phase than in the other experiments. Moreover, the results show that subjects prefer the option of having higher non bought-out bandwidths at their demand whenever they need them. Putting all these results together enabled the author of [1] to come up with the proposal previously explained. Some concluding remarks about this proposal are found in the next three paragraphs.

From the customers' perspective, the flat-rate buy-out option is considered a very good pricing scheme, because not only do they have unlimited access to the Internet, but they also have access to higher bandwidths any time they need it. So it combines the benefits of the flat-rate pricing plans (stability of prices) and flexibility of acquiring different QoS for access on demand.

From the ISPs' prospective, this plan provides additional revenues besides the ones they would get from purely flat-rate plans. This was the main advantage of the regular Diff-Serv plan. Furthermore, this pricing plan reduces over-usage of resources due to the user-based component, which fulfills the economic requirement.

Finally, in terms of the accounting technology, it will definitely be more difficult than pure flat-rate pricing. However, it would probably be less or of equal complexity to the Diff-Serv pricing scheme.

2.2.4 Congestion Pricing

The basic idea behind this scheme is to have customers pay for the congestion they make by sharing in the social costs. Different approaches have been proposed and we include some in this section that have been described in [7].

Smart Market

This is a kind of auction scheme. At each congested node a bid value for each incoming packet indicates the amount that the owner of this packet is willing to pay to have his packet go through the node. Then the node (router) holds an auction and admits the packets with the highest bids. For more information on this scheme refer to [9]. This type of pricing scheme is considered very efficient economically, but can yield unstable charges along with complex technical accounting difficulties [13].

Vickrey Auction

This scheme is very similar to the smart market; however, the winner pays the second highest bid instead of the highest one. Since bidders are no longer afraid of making high bids, therefore their true preferences appear. Although this scheme may prevent congestion effectively, but it was determined that it can not work in a large network. In a large network with multiple congested routers, a packet that wins at a given router might lose at the next, because the bid value at the first router is subtracted from the total bid value. If a user wishes to program the packet to give a new bid at each router then a feedback signal to the user must occur to provide information about the current status of the packet. Unfortunately this is not feasible, because this feedback signal will congest the network even more! For more information about this scheme refer to [18] and [8].

Edge Mechanism

The idea here is to shift the pricing decisions to the edge of the network. This allows operators to charge for expected congestion instead of the actual one. A scheme called the split-edge pricing framework [2] argues that instead of users making individual payments to owners of congested routers, pricing is applied only to users at the network's end. Each ISP finds out the cost of traffic through its own network and offers various service categories to its neighbors at a certain price. This pricing scheme is very good for its simplicity and transparency to the user [13].

Statistical Approaches

Routers mark packets in case of congestion using two bits in the IP header which are known as the Explicit Congestion Notification (ECN) mechanism. An ISP can make use of this by aggregating the ECN mark and at the receiver the number of marks can be counted then a charge would be applied. This scheme might be difficult to implement, because various networks use the ECN field for different purposes which invokes a compatibility problem. Furthermore, some security concerns may arise if a rogue router tampers with these two bits. We would like to note that this scheme is somewhat similar to the cumulus pricing that is discussed in the next section.

To conclude this section of congestion prices one has to take the customers' opinion to this scheme into consideration. As stated before, users generally prefer stable QoS levels at predictable prices. Moreover, many congestion pricing schemes assume small numbers of congested nodes. Although this assumption maybe valid for predictable networks such as road networks, it cannot be easily applied to the Internet which has highly unpredictable congestion scenarios.

2.2.5 Cumulus Pricing

It is a flat-rate pricing scheme in principle where rates vary over long periods of time. The author of [13] proposes this scheme that is based on cumulus points that serve as a feedback mechanism. First, an initial agreement is set between the

ISP and the customer. This agreement includes certain QoS requirements by the user such as expected bandwidth, delay, probability of packet loss, etc. Next, the ISP monitors the user's usage and after each month the user either cumulates red points for exceeding the agreed upon bandwidth or cumulates green points for the opposite case. The more the difference between the actual usage and the agreed upon bandwidth the more cumulative points are given. A decision is taken after some specified number of billing periods (a billing period = 1 month) which could be after one year for example or if the number of points exceed a certain maximum (threshold). At the end of the year the cumulative points are added for the user in which a green point cancels a red one. If the user has green points he/she may get a refund or cumulate these points for the next year. If the user has red points a decision may be to pay an equivalent amount of money to the number of red points and perhaps increase the bandwidth requirement for the next year. Figure 2.8, which was reproduced from [13], gives a nice illustration of a typical user who agreed on a bandwidth of x and receives points according to the actual usage till a threshold is reached in the month of June when a decision should be made. It is observed that a green point cancels a red one from the month of March. A formal model for this scheme is described in [13].

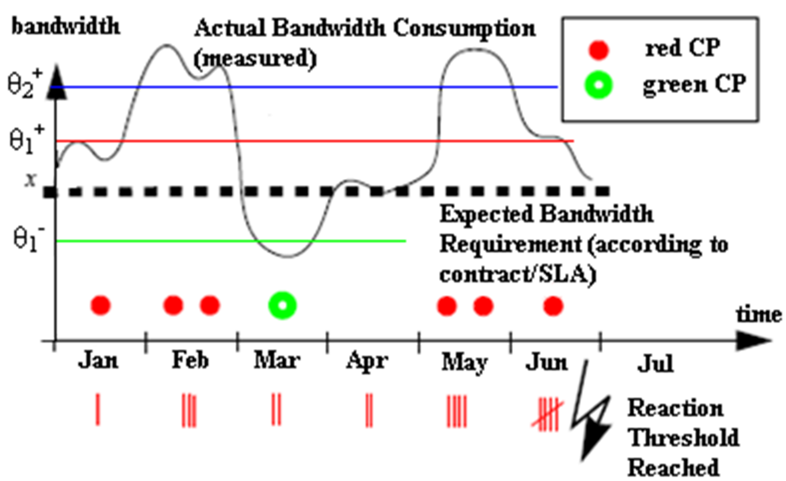


Figure 2.8: Illustration of the Cumulus Pricing Scheme.

Final Remarks: When comparing the cumulus pricing scheme to our requirement types defined earlier, we find that this scheme does a good job. First, it is a flat-rate scheme, so it has all the benefits of this scheme in terms of the acceptance by the majority of customers (price stability).

Second, it gives variability in the QoS, which again according to the INDEX experiment is highly preferred by users. Third, it is transparent to the customer and provides feedback every billing period (a month) and does not annoy the customer, because the decision is made after a long period of time when the decision threshold is met.

Fourth, it is an implementation of the edge pricing scheme, because the only interaction for the customer is with the ISP in negotiating the contract and receiving cumulus points.

Fifth, it is economically efficient, because accumulated points at the end of a period are either rewarded for under-use or penalized for over-use. This in some sense should keep the network in a non-congested state.

Sixth, it is technically feasible, because after all, it is up to the ISP to specify the way it wants to monitor the users' usage. Therefore, a simple method of taking rough measurements or perhaps even estimations of user's activity would fulfill the accounting technical feasibility issue.

Finally, it would be a good idea to research some of the metering policies and their complexities. Also, more focus should be on how to prevent fraudulent users who try to cheat the scheme. Furthermore, studying the different trade-offs of design issues for ISPs should be given more attention.

2.3 PRICING IN A MULTIPLE ISP DOMAIN

So far we have discussed the basic building blocks of the Internet and we have seen the concept of peering and of purchasing transit. We also talked about different pricing schemes that are being applied currently or under research. It is now time to bring all these issues together and study how ISPs interact with one another in the real world and how they place their prices to attract customers in a competitive market.

This research topic is becoming more popular these days, because of current demands for change of the Internet itself. As discussed in the literature nowadays, communication networks are trying to converge to one big network by combining the wireless cellular communication world with the IP world. This has been a dream for 3G networks, but now this dream has been postponed till when 4G is deployed.

Whether this is all feasible in the near future or not, researchers have already started researching how billing can be performed in a multiple ISP environment. In this section we would like to take a brief look at four papers discussing how Game Theory can be used to provide means for implementing pricing models in such an environment where ISPs compete for customers.

The first paper is by Shakkottai and Srikant [15]. In this paper the authors place a pricing model for the Internet consisting of a hierarchy of ISPs, such as Tier 1, Tier 2 and Local ISPs, along with the possibility of peering through Private Exchange Modules. Some game theoretical concepts are first discussed. Then based on the Internet model placed; they study how different local ISPs interact at the same level in the hierarchy and between different levels in the hierarchy. Then they take a look at the case when there are private exchange modules. After that, QoS is considered along with other pricing schemes in the model. In all of the discussion of this paper a threat strategy is placed based on Nash reversion to ensure the cooperation of different ISPs. The authors prove that if a large number of ISPs are present, this will lead to price wars. Eventually some ISPs would buy out others until we have a smaller number of players to ensure that a threat strategy is in place. For more details on this research, we refer the reader to [15].

We then look at [3]. The authors of this paper study two types of games. The first one is with an ISP as a player and a customer as the other player of the game. They show that a leader-follower game in this case may lead to a non Pareto optimal solution which may also be unfair. They argue that by cooperation, and with the help of some kind of government regulations to sustain it, both players would be better off than with the leader-follower game. The second game is between two ISPs competing for customers, which is a nonzero sum game. They show by numerical methods that in this game a Nash equilibrium exists and that this proposal is applicable to other pricing problems such as the Paris Metro pricing. For more information on this paper please refer to [3].

In [6], the authors propose a fair revenue-sharing scheme based on weighted proportional fairness. They argue that this scheme leads to higher profit for all providers while encouraging improvements and upgrades for bottleneck links. Moreover, they point out that noncooperative pricing strategy leads to unfair distribution of profit among providers. It is to be noted that in this paper the price of a route controls the number of customers using this route, which is modeled by a demand function. Therefore, ISPs dynamically adjust their prices to control traffic demands on the links of their network. Please refer to [6] for more details.

The fourth and final paper to be discussed in this section is by R. Malhamé et al [14]. We would like to note that our pricing model, which is given in the next chapter, is based on the model implemented in [14]. For this reason, we briefly go through this paper indicating the differences between the two models, whereas in the next chapter we give an in depth proposal of our model.

One of the motivations for [14] is that the profit of ISPs, which are the major players of the Internet, has been neglected in many pricing schemes. Therefore, the pricing scheme tries to maximize the profit of ISPs while maintaining a level of quality of service for the customers. The structure proposed is of a third party (TP) provider that oversees the ISPs. So the function of the TP is to negotiate a unit flow price with a customer and to choose ISPs along the path of the customer's traffic from the source to the destination. Since the TP knows the structure of the Internet, therefore it asks those ISPs that can deliver the customer's traffic to provide a certain delay guarantee (statistically not deterministic) specified by the customer. It is agreed that a customer pays only for the percentage of traffic that meets the agreed upon maximum tolerable delay. The TP takes a fixed percentage of the profit and distributes the remaining portion to the ISPs according to a sharing system that reflects the QoS they proposed to offer.

Some of the assumptions made by [14] are as follows: First, the game between ISPs is a noncooperative game. Second, QoS is statistical and not guaranteed, because deterministic QoS is wasteful in bandwidth requirement and it becomes too complicated in a multiple ISP environment. Third, this is a Stackelberg game in which the TP is considered the leader and the ISPs form the group of followers. Fourth, a perfect information game is assumed in which all the players know everything about each other's strategies. Fifth, only one data flow passes through the ISPs, i.e. only one customer is allowed to send a single class of packets from the source to the destination. Sixth, fixed routing is assumed from the source to the destination. Finally, each ISP is represented by a single bottleneck node along the route from source to destination in which an M/M/1 queue represents this node.

In our model, we took the same basic framework of the game and made some extensions to the model. First, we added more than one customer to the game, hence we have different traffic classes with varying packet lengths. Second, as a result of varying packet lengths, each ISP is now represented by an M/G/1 queue instead of an M/M/1 queue. Finally, since we are dealing with a more general scenario, we can no longer use probabilities to measure the portion of traffic that meets the specified delay criterion by the customers. Instead we use the Pollaczek-Khinchine formula to give average delays in which we compare with the maximum delay tolerated by the customer. As a result of all this, the utility functions of the TP and ISPs have been altered. Furthermore, the cost paid by a customer is not only a function of the rate of traffic λ but also a function of the maximum tolerable delay of that customer along with the coefficient of variation of its traffic.

To conclude this section, we would like to point out the reason behind choosing [14] as a base for our work. This pricing scheme meets the three criteria discussed before in Section 2.2 which are customer's preference, economic efficiency for the ISP's network, and simple accounting and billing technique. This pricing scheme meets the customer's expectations of paying the same predictable amount of money over the billing period while maintaining a fixed QoS. On the other hand, the network only allows the amount of traffic that it can handle efficiently (by maximizing its profit) before the agreement is signed between the customer and the TP. Moreover, billing is not a big issue due to the fact that QoS is monitored statistically and not in a deterministic fashion. Therefore, this pricing scheme serves the purpose and by this point we are ready to present our model in the next chapter in depth.

Chapter 3

Model Formulation

3.1 INTRODUCTION

So far in our discussion of Internet pricing models in a realistic environment of multiple ISP domains, we have reached a point where we chose a model [14] that suits our needs. In this chapter we introduce our expansion and variation of this model to include more than one customer with traffic that is varying in packet length. To do so, there was a need to re-define the utility functions of the players along with adding some variables and slightly changing the way the game is played.

In Figure 3.1 we present a diagram of how the game is setup. In this game we have two customers, C_1 and C_2, the third party (TP), and two Internet service providers, ISP_1 and ISP_2. This figure aims to show the relationship between these different players of the game. Dashed lines represent control flow, whereas solid lines represent data flow.

The two customers are requesting their traffic to flow from point A to point B. They specify the maximum amount of tolerable delay they require. This request is forwarded to the TP. Then the TP takes action by contacting different ISPs along the path asking them to grant access to its customers. Each ISP responds to the TP with the initial amount of bandwidth they are willing to offer the customers. Each ISP has the option of adding extra bandwidth at additional cost on the ISP's part, provided that this would maximize its payoff. The TP then performs all the optimization calculations, which will be discussed later, and agrees with the customers on a certain rate of traffic for each one of them along with the unit flow price. This unit flow price is a function of the maximum tolerable delay, coefficient of variation of traffic, and the rate of traffic. After attaining a portion of the profit, the TP distributes the rest of the profit among the ISPs according to a sharing mechanism which takes their contributions into consideration. After all these negotiations are accomplished, traffic starts to flow from point A to point B. It is in the customers' best interest to maintain traffic rates at the same negotiated rate, this can be done by sending dummy traffic.

To give a numeric example, we use the values of the scenario that will be mentioned in Section 4.2.1. Taking a look at Figure 3.1, the example proceeds as follows. First, customer 1 and customer 2 notify the TP that their desired maximum delays are 3 msec and 4 msec, respectively. They also, specify that they want their traffic to move from point A to point B. Second, TP does some measurements for the traffic coming from the two customers and negotiates with them the unit flow price of traffic. It turns out that customer 1's traffic has packets that are exponential in nature, whereas customer 2's traffic has deterministic packet lengths. The unit flow price agreed upon is of the form $\exp(-\frac{\lambda_j}{0.75})$, where λ_j is the rate of traffic arrival from customer j $(j = 1, 2)$. Third, the TP passes on all the previously mentioned values to ISP_1 and ISP_2, which lie on the path from the source to the destination. Fourth, both ISPs reply back to the TP with the initial amount of fixed bandwidth ($\mu_1 = 1.1$ packets/msec and $\mu_2 = 1.2$ packets/msec) that they are willing to offer along with a value for the maximum extra bandwidth ($\Delta\mu_1 = 1$ packet/msec and $\Delta\mu_2 = 1$ packet/msec) that they can add. There is a price that an ISP pays for each unit of added bandwidth. This value is denoted by $p_1 = 0.075$ and $p_2 = 0.055$. Now that all the information is available, a non-cooperative game is played by the ISPs for each possible value of λ_j's. Since the TP is the leader of this game, it will choose the value of λ_j to maximize its profit. Once λ_j's are determined, the decision variables of the ISPs are also found along with their associated profits. Finally, the costs for customer 1 and 2 are calculated. For the numerical results, please refer to Table 4.2.

The structure of this chapter is arranged as follows, in Section 3.2 we give detailed information about the assumptions made in the model. In Section 3.3, we go through a number of definitions for the framework of the game followed by a description of the utility functions of the players of the game. Finally, in Section 3.4 a series of steps are presented to highlight the procedure by which the game flows.

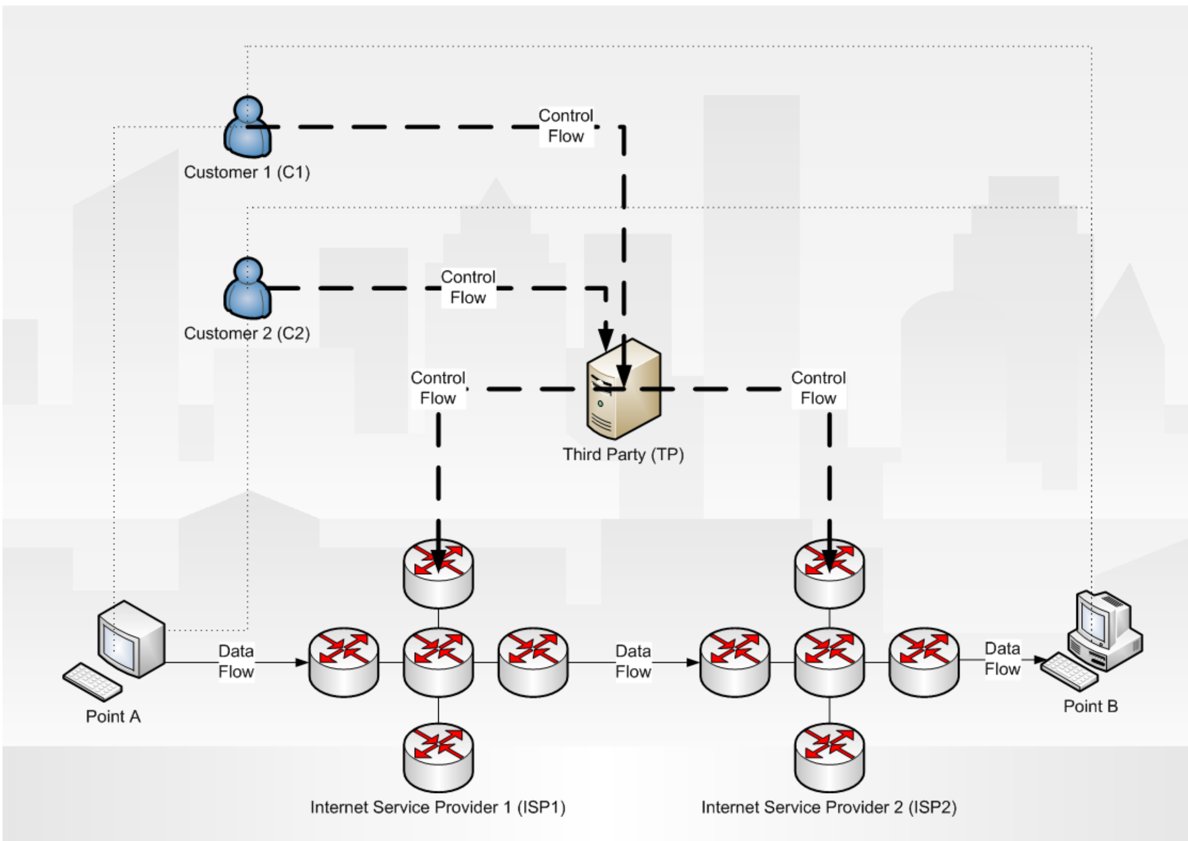


Figure 3.1: The Relationship between the Players of the Game

3.2 ASSUMPTIONS

In this section we present the assumptions made related to our model. First, we do not take routing into consideration. We just place the fact that customers need their traffic to get from the source to destination. The TP is assumed to know the structure of the Internet and the different autonomous systems with each region belonging to which ISP. Hence, it is the TP's responsibility to contact different ISPs and find out which ones are willing to deliver the customers' traffic from the source to the destination. On the macro-level, the TP deals with routing between ISPs, whereas on the micro-level, each ISP deals with its own routing issues.

Second, the design of the model enables customers to pay only for the percentage of traffic that is actually delivered from source to destination with the agreed upon QoS. In this case the QoS is the maximum tolerable delay for the customer's traffic. The model takes this into account using a decision variable in the utility function of the TP and the ISPs. This decision variable is set to zero if the maximum tolerable delay criterion is not met, whereas it is set to one if the criterion is met. Furthermore, we would like to point out that in our model a customer is really a group or a class of customers. In other words, they are not just a single user.

Third, as mentioned before, QoS is statistical not deterministic (guaranteed). This saves on wasted bandwidth resources along with giving a rather easier pricing scheme in terms of monitoring for billing purposes. Moreover, if deterministic QoS had been used, it would have been much more challenging in determining end-to-end delays within a multiple ISP domain.

Fourth, the system is decentralized by using a control scheme over QoS that is enforced by the share of the total profit each ISP gets.

Fifth, each ISP is represented by a single bottleneck node along the route from the source to the destination. An M/G/1 queue is used to model this node.

Sixth, a perfect information game is assumed. Every player knows the extra bandwidth unit buying cost by each ISP, the original bandwidth offered by each ISP, and the unit cost versus traffic relationship between the customer and the TP. Furthermore, ISPs are transparent to the customers.

Finally, a Stackelberg game is introduced in which the TP is the leader and

ISPs form the group of followers. Moreover, the game played between the ISPs (followers) is a non-cooperative game.

3.3 GAME STRUCTURE

3.3.1 Definitions

Let $\lambda_j \equiv$ rate of arrival of packets from customer j

Let $L_j \equiv$ packet length of customer j

Let $E[L_j] \equiv$ mean packet length of customer j

Let $\mu_i + \Delta\mu_i \equiv$ service rate of ISP_i where $\Delta\mu_i$ is the amount of extra bandwidth that ISP_i can increase to improve its overall payoff.

Let $E_i[\sigma_j] \equiv$ mean service time for packet j in ISP_i and is given by:

$$E_i[\sigma_j] = \frac{E[L_j]}{\mu_i + \Delta\mu_i} \tag{3.1}$$

The second moment of the service time for packet j in ISP_i is given by:

$$E_i[\sigma_j^2] = \frac{E[L_j^2]}{(\mu_i + \Delta\mu_i)^2} \tag{3.2}$$

The aggregate rate of traffic arrival is given by:

$$\lambda = \sum_{j=1}^{m} \lambda_j \tag{3.3}$$

The mean service time for all packets within ISP_i is given by the sum of the mean service time for packet j weighted by the probability of arrival of packet j and it is given by:

$$E_i[\sigma] = \sum_{j=1}^{m} E_i[\sigma_j] \frac{\lambda_j}{\lambda} \tag{3.4}$$

Furthermore, the second moment of the service time for all packets within ISP_i is given by:

$$E_i[\sigma^2] = \sum_{j=1}^{m} E_i[\sigma_j^2] \frac{\lambda_j}{\lambda} \tag{3.5}$$

The utilization factor for ISP_i is given by:

$$\rho_i = \lambda E_i[\sigma] \tag{3.6}$$

The total time spent in ISP_i on average is given by the Pollaczek-Khinchine formula below:

$$E_i[T] = E_i[\sigma] + \frac{\lambda E_i[\sigma^2]}{2(1-\rho_i)} \tag{3.7}$$

The total time spent in all the ISPs along the route from source to destination is given by:

$$E[T] = \sum_{i=1}^{n} E_i[T] \tag{3.8}$$

where $n \equiv$ number of ISPs

The price paid by customer j depends on three factors. First, it is inversely proportional to the maximum tolerable delay specified by customer j and denoted by T_{max_j}. Second, it is linearly proportional to the unit flow price, which is usually taken to be an exponentially decreasing function of λ_j, and it is denoted by $C_v(\lambda_j)$. Finally, the price paid by customer j is linearly proportional to the coefficient of variation of the traffic of customer j. The formula for customer j's price is given by:

$$P_{C_j} = \frac{a\, C_v(\lambda_j)\, \lambda_j\, (1 + \beta\, c_j)}{T_{max_j}} \tag{3.9}$$

where a is a decision variable that allows customer j to pay only for the amount of traffic that meets the QoS agreed upon. It can have only two values, 1 and 0. It takes a value of 1 if the total average time spent in the system by a packet is less than the minimum of the maximum tolerable delays specified by each customer. It takes the value of 0 otherwise. The variable a is given by:

$$a = \begin{cases} 1 & \textit{if } E[T] \leqslant \min\left(T_{max_1}, T_{max_2}, ..., T_{max_m}\right); \qquad m \equiv \text{number of customers} \\ 0 & \textit{otherwise} \end{cases} \tag{3.10}$$

In Equation (3.9), β is a factor to adjust the priority level for accepting different traffic variations from users. Moreover, c_j is the coefficient of variation of customer j's traffic and it is defined as the ratio between the standard deviation and the mean and it is given by:

$$c = \frac{\sqrt{E[\sigma^2] - (E[\sigma])^2}}{E[\sigma]} \tag{3.11}$$

3.3.2 Utility Functions

Third Party (TP)

The utility function of the TP is simply defined as the summation of the prices of all the customers multiplied by a percentage factor of the total profit reserved for the TP. The utility function is given by:

$$TP_U = M \sum_{j=1}^{m} P_{C_j}; \qquad \lambda = \sum_{j=1}^{m} \lambda_j \leqslant \sum_{j=1}^{m} \lambda_{max_j} \tag{3.12}$$

where $M \in [0, 1]$ which is the fraction of total profit reserved for the TP

It is to be noted that λ is the only decision variable for the TP, it is chosen to maximize TP_U.

Internet Service Providers (ISPs)

As mentioned before in the introduction of this chapter, each ISP_i has a certain amount of fixed bandwidth μ_i always available for all of the customers' traffic. To guarantee stability of the system the following condition must hold:

$$\sum_{j} \lambda_{max_j} \leqslant \mu_i \quad \forall i \tag{3.13}$$

Moreover, each ISP_i has the option to increase its bandwidth by $\Delta\mu_i$ subject to the condition $\Delta\mu_i \leqslant \Delta\mu_{max_i}$; therefore the actual bandwidth of ISP_i is given by $\mu_i + \Delta\mu_i$ at the price of p_i per unit of bandwidth added.

We next define the sharing mechanism by which each ISP is rewarded according to its contribution towards QoS. ISP_i's share is proportional to the difference between the minimum of the maximum tolerable delays specified by each customer and the average amount of time a packet spends in ISP_i's network. ISP_i's share is given by:

$$S_i = \frac{\min(T_{max_1}, T_{max_2}, ..., T_{max_m}) - E_i[T]}{\min(T_{max_1}, T_{max_2}, ..., T_{max_m})} \tag{3.14}$$

From Equation (3.14) we notice that as $E_i[T]$ increases S_i decreases so as to penalize the ISP that offers poor QoS and reward those offering better QoS.

The utility function of ISP_i is simply defined as the remaining portion of the total profit, after the TP took its share, multiplied by ISP_i's share (S_i) multiplied by the total price paid by all the customers. Then we subtract the cost of adding extra bandwidth from that amount. The utility function of ISP_i is given by:

$$ISP_{U_i} = (1 - M)\, S_i \sum_{j=1}^{m} P_{C_j} - p_i\, \Delta\mu_i \tag{3.15}$$

where a is given by Equation (3.10) and p_i is the unit flow price for each unit of added bandwidth.

It is to be noted that $\Delta\mu_i$ is the decision strategy for ISP_i.

3.4 FORMULATION OF THE STACKELBERG GAME

The game takes the form of a Stackelberg game. A Stackelberg game is a game which has a leader and some followers. The game is solved by backward induction. The leader considers what the best response of the followers is, then the leader picks a strategy for its decision variable that maximizes its own profit. The followers actually observe this and in equilibrium choose the expected quantity as a response. Of course in this game the leader has the upper hand and is in a better position.

In our model the leader is the TP whereas the ISPs are the followers of the game. As we mentioned before in our assumptions in Section 3.2, the game is a perfect information game. This implies that all strategies are transparent and available to all the players of the game.

The game is played as follows:

1. Each customer j notifies the TP with its desired T_{max_j} along with the source and destination points for its packets.

2. The TP measures the coefficient of variation of the traffic of each customer.

3. The TP negotiates with the customers the unit flow price of traffic $C_v(\lambda_j)$.

4. The TP looks up in its database for the ISPs that it can approach, which lie along the path from the source to the destination. It then passes the information it gained from the customers to the ISPs and waits for a response whether they are willing to be involved in the deal or not.

5. If an ISP replies with an approval, it will notify the TP with the initial amount of fixed bandwidth μ_i and the maximum extra bandwidth $\Delta\mu_{max_i}$ it might add.

6. For all possible λ_j's, the calculations of Section 3.3.1 are performed, then a non-cooperative game is played between all the ISPs to determine a Nash equilibrium between them. In this game each ISP_i chooses a decision strategy $\Delta\mu_i$ of which would give it a better payoff.

7. Now the TP has the values of the utility functions of all ISPs under any choice of λ_j's, working its way backwards, it will choose values of λ_j's that maximize its own payoff. Hence its payoff is now calculated.

8. Now that the customer traffic level is determined, the payoffs of the ISPs are automatically determined as well.

9. Finally, the customer's price is calculated.

The outcome of the whole game is the traffic rate of each customer λ_j, the profit of each ISP, the profit of the TP, and the price paid by each customer.

We would like to refer the reader to appendix A for a pseudocode example involving two customers and two ISPs. Moreover, a brief explanation of the method used to calculate the Nash equilibrium among the followers of the game is found in appendix B.

In this chapter we completely defined our model. We started off with some of the assumptions made, followed by the framework of the game and the definitions of the different utility functions of the players of the game. We then showed the steps of playing the game to achieve our desired outcome. In our next chapter we give simulation results that backup the theory presented in this chapter. It is interesting to see that changing some of the initial parameters have a significant effect on the results of the game.

Chapter 4

Numerical Results

4.1 INTRODUCTION

In this chapter we present the numerical results for the simulation of our model that was formulated in Chapter 3. We simulated four combinations of scenarios. The first one is having two customers and two ISPs. The second is with two customers and three ISPs. The third one is for three customers and two ISPs. Finally, the last scenario is for three customers and three ISPs. Once more we note that the customers here are really a group of customer not individual users.

In each of these scenarios we present a number of plots and tables. We are interested in finding the Nash equilibrium for each case and hence the outcome of the game. The results that identify the Nash equilibrium of the game are the values of the extra added unit bandwidth $\Delta\mu$'s for the ISPs. The outcome of the game are the values of the rate of arrival of the customers' traffic λ's, the profit of each ISP, the profit of the TP, and the price paid by each customer.

Furthermore, for the scenarios having two customers we are able to have three dimensional plots representing all possible values for the different outcomes of the game versus the λ_1 on the x-axis and λ_2 on the y-axis.

In each scenario, the simulation is run twice. Once for a general case of having different allowable maximum delay T_{max}'s for customers along with having variation in their traffic that is taken from different distributions. For the second case, we assume that all the customers have the same allowable maximum delay T_{max}, whereas the traffic variation in their traffic is still taken from different distributions. The goal of this second case is to capture the effect of traffic variations without the interference of T_{max}.

4.2 SCENARIO 1: TWO CUSTOMERS AND TWO ISPs

In this section we simulate the case of having the TP choose two ISPs along the path from source to destination for two different customers. As indicated in the introduction of this chapter, we will study two cases of simulation. The first case we call it the General Case in which the T_{max}'s of customers and their traffic variations are arbitrarily chosen to differ between the customers. The second case we call it the Fixed T_{max} Case in which T_{max} is the same for both customers and the simulation is run multiple times for different values of T_{max}. Again as mentioned in the introduction, the reason for having the second case is to study the effect of traffic variation alone without having T_{max} in the equation.

4.2.1 General Case

In this case we assumed that the first customer has exponential packet lengths, whereas the second one has deterministic packet lengths. Moreover, all the values taken for the different variables in the simulation are shown in Table 4.1. We would also like to note that some of the initial values used for these values were taken from [14].

The results of the simulation are shown in Table 4.2. These results represent the Nash equilibrium of the game. It is apparent from these results that ISP_2's profit is greater than that of ISP_1. This is fair since ISP_2's contribution of bandwidth is larger and hence it results in less delay for the overall system (Notice how $E_2[T]_{Nash} < E_1[T]_{Nash}$). Furthermore, since customer 1's traffic has more variations in it than the traffic of customer 2, therefore it consumes more network resources. For that reason and also, because $T_{max_1} < T_{max_2}$, we see that the price paid by customer 1 is higher than that of customer 2 (Notice that $P_{C_1} > P_{C_2}$). However,

Table 4.1: Initial Values for Simulating the General Case of two Customers and two ISPs.

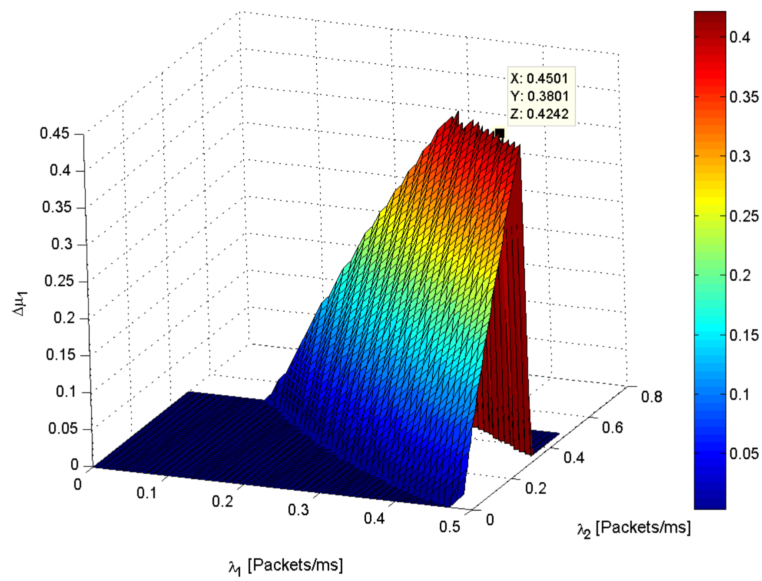
Variable	Value
β	$\frac{1}{3}$
λ_{max}	1 [packets/msec]
$\Delta\mu_{max_1}$	1 [packets/msec]
$\Delta\mu_{max_2}$	1 [packets/msec]
T_{max_1}	3 [msec]
T_{max_2}	4 [msec]
M	20%
μ_1	1.1 [packets/msec]
μ_2	1.2 [packets/msec]
p_1	0.075
p_2	0.055

the pricing model admits more of customer 1's traffic. This is, because in doing so, it gives better profit for the TP and for the ISPs. We will see later on, when we fix T_{max}, there is a point where admitting more of customer 1's traffic gives worse results for the TP and the ISPs.

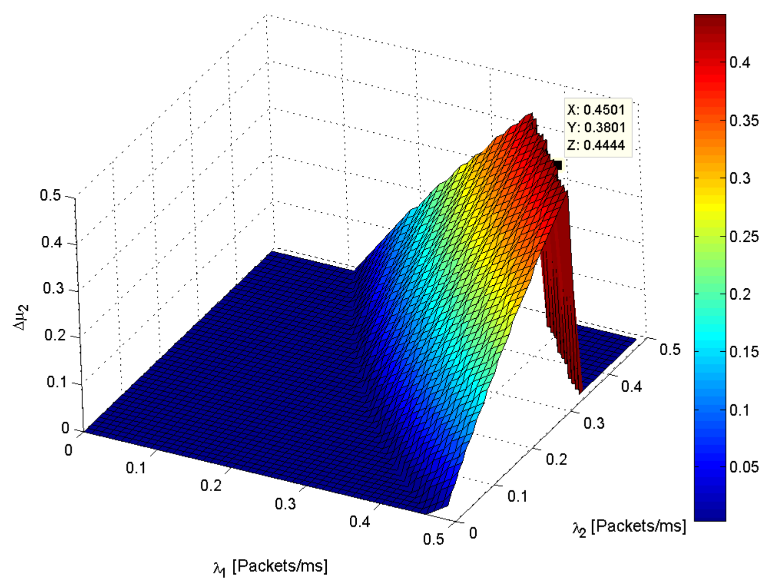
Table 4.2: Nash equilibrium for the game of two customers and two ISPs.

Variable	Value
λ_{Nash_1}	0.4501 [packets/msec]
λ_{Nash_2}	0.3801 [packets/msec]
λ_{Total}	0.8302 [packets/msec]
$\Delta\mu_{Nash_1}$	0.4242 [packets/msec]
$\Delta\mu_{Nash_2}$	0.4444 [packets/msec]
$E_1[T]_{Nash}$	1.2612 [msec]
$E_2[T]_{Nash}$	1.0862 [msec]
TP_{Nash}	0.0334
ISP_{Nash_1}	0.0457
ISP_{Nash_2}	0.0609
P_{C_1}	0.1098
P_{C_2}	0.0574

We also include a series of plots showing the outcome of the game. These plots are plotted against λ_1 and λ_2 and are shown in Figures 4.1, 4.3, and 4.2. It is apparent from these figures that they all tend to have a sharp fall at a certain value of λ_1 and λ_2. This is, because when increasing the rate of arrival of traffic to a certain point, QoS deteriorates and hence the utility functions of the TP, and the ISPs drop.

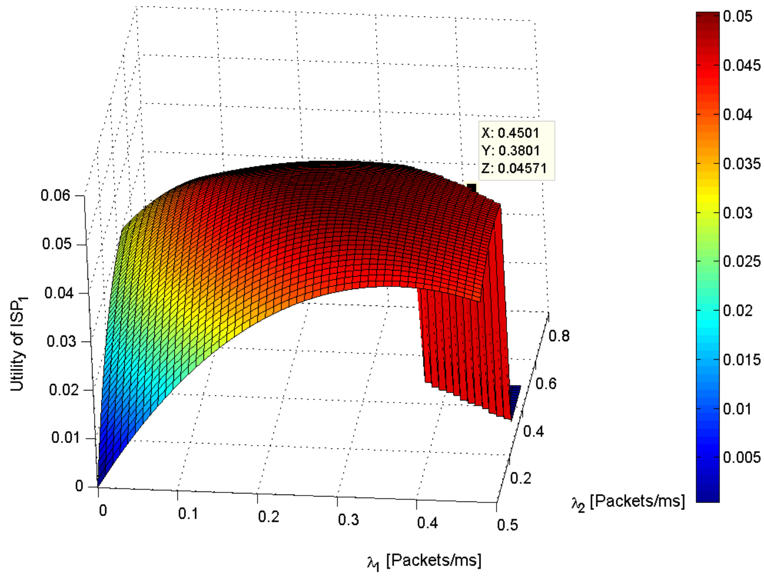


(a) $\Delta\mu_1$ vs. λ_1 and λ_2

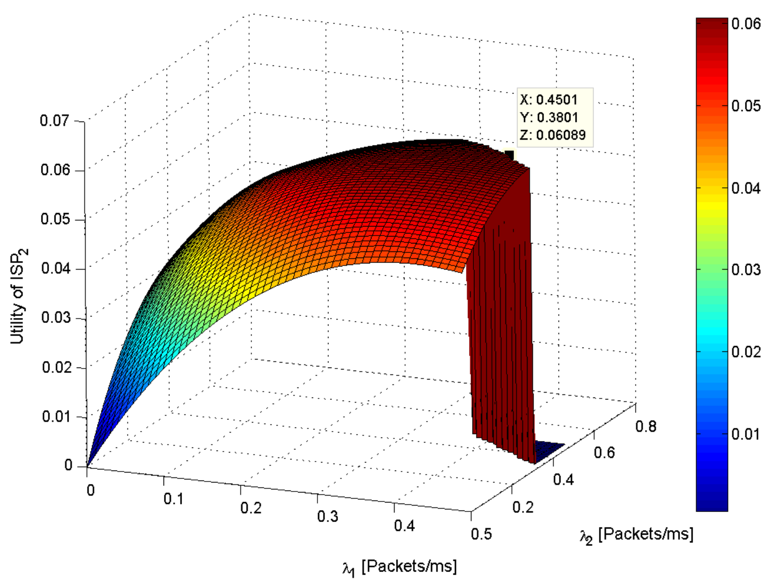


(b) $\Delta\mu_2$ vs. λ_1 and λ_2

Figure 4.1: Additional Bandwidth Acquired by ISP_1 and ISP_2 as the Outcome of the two Customers and two ISPs Game.



(a) *ISP$_1$* vs. λ_1 and λ_2



(b) *ISP$_2$* vs. λ_1 and λ_2

Figure 4.2: The Utility of *ISP$_1$* and *ISP$_2$* as the Outcome of the two Customers and two ISPs Game.

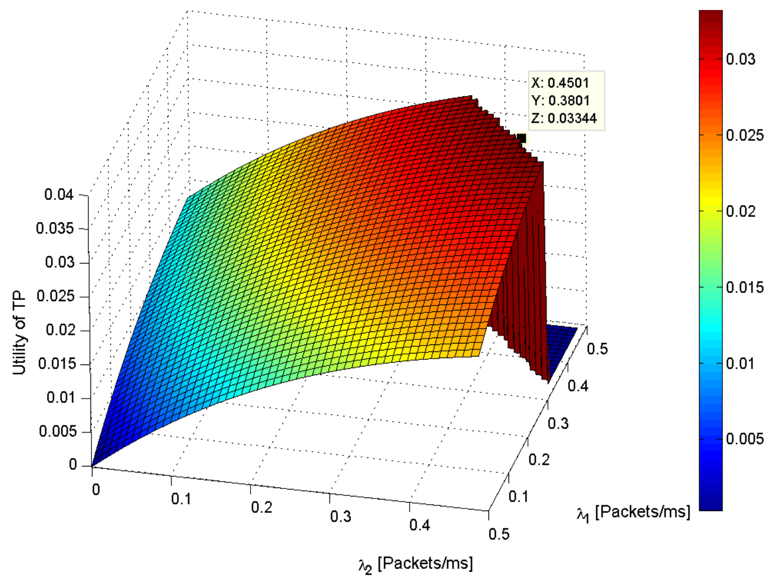


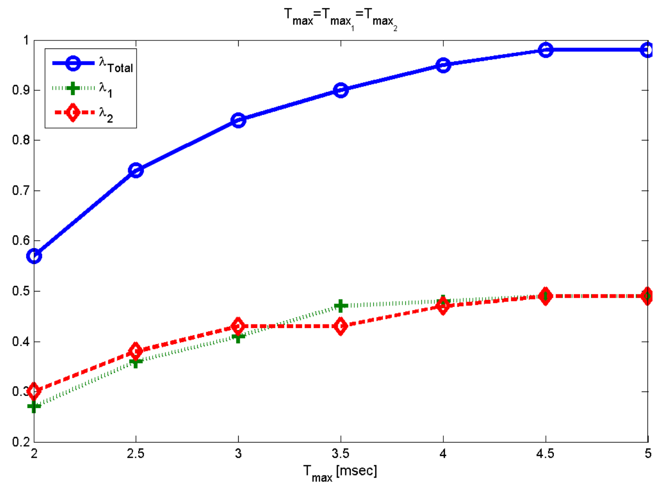
Figure 4.3: The Utility of the TP as the Outcome of the two Customers and two ISPs Game.

4.2.2 Fixed T_{max} Case

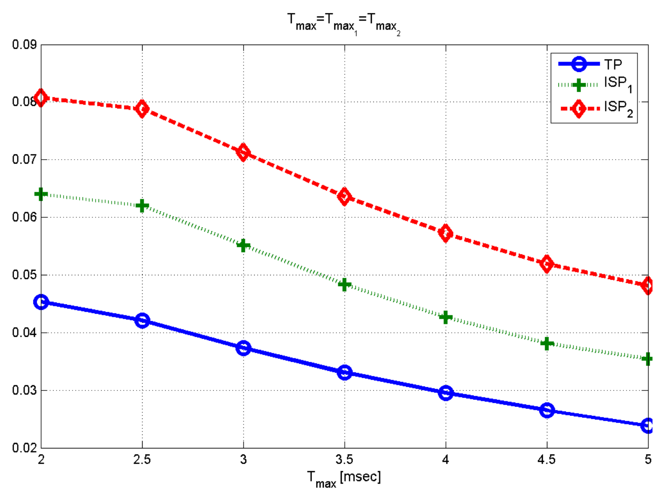
In this section we deliberately make $T_{max_1} = T_{max_2} = T_{max}$ so that the system chooses λ_1 and λ_2 based on the variation in the customers' traffic only. The initial conditions are the same as those given in Table 4.1. Once again the second customer has deterministic packet lengths, however, we try different variations for the first customer. We try first exponential packet lengths, then higher than exponential, and finally much higher than exponential. In each case we plot λ_1 and λ_2 vs. T_{max}, the utilities of the TP and the ISPs vs. T_{max}, and finally the price paid by each customer vs. T_{max}.

After observing Figures 4.4, 4.5, and 4.6, we would like to make some comments. In Figure 4.4a, we see that at low values of T_{max} the system admits more of customer 2's traffic. This is, because it has less variation in traffic. As the T_{max} requirement is relaxed (T_{max} is increased) the system admits more of customer 1's traffic, because that gives better payoff for the TP and the ISPs. In Figure 4.5a, we notice that all of the λ curves have decreased. This is due to the increase in the variation of customer 1's traffic. In Figure 4.6a, the situation becomes more dramatic and almost none of customer 1's traffic is admitted.

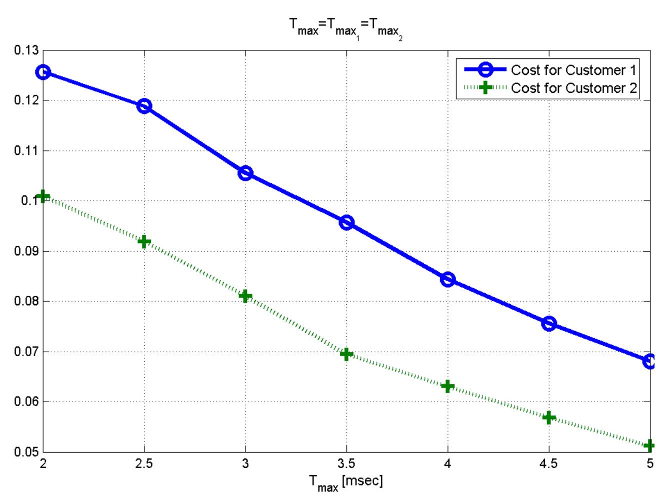
Moreover, we see from comparing Figures 4.4b, 4.5b, and 4.6b, that the profit



(a) λ_{Total}, λ_1, and λ_2 vs. T_{max}

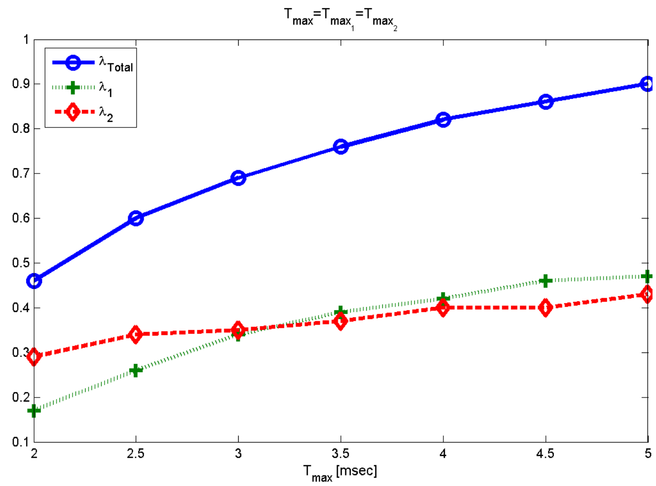


(b) *TP*, *ISP₁*, and *ISP₂* vs. T_{max}

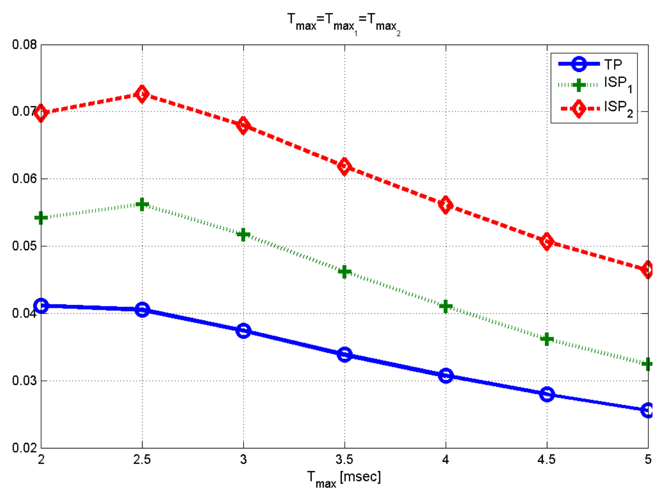


(c) Customers' Costs vs. T_{max}

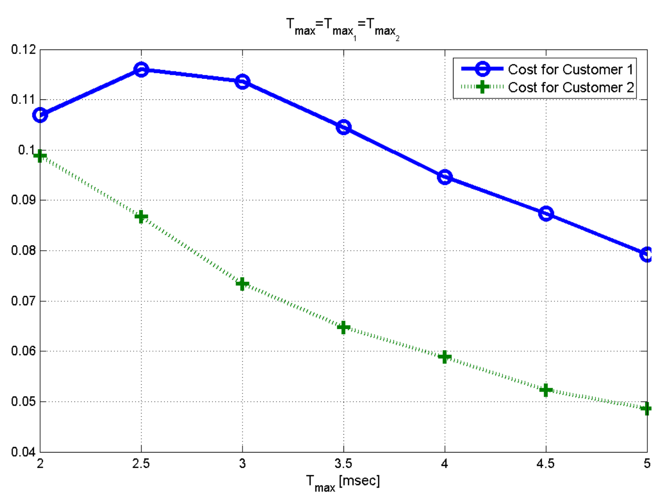
Figure 4.4: Outcome of the two Customers and two ISPs Game vs. T_{max} for Exponential Packet Lengths for Customer 1 and Deterministic Packet Lengths for Customer 2



(a) λ_{Total}, λ_1, and λ_2 vs. T_{max}

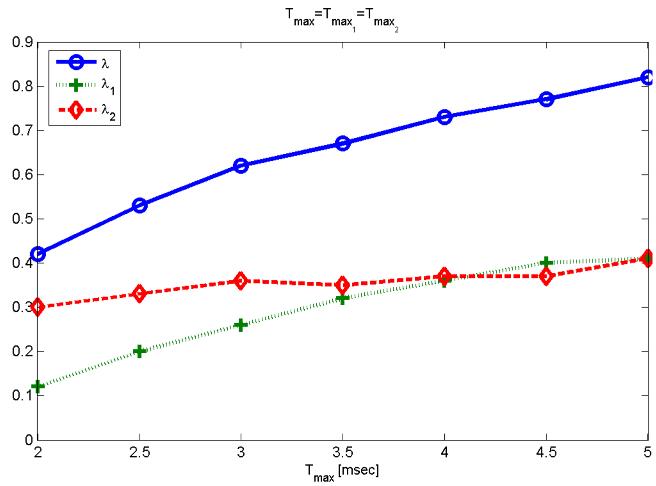


(b) *TP*, *ISP₁*, and *ISP₂* vs. T_{max}

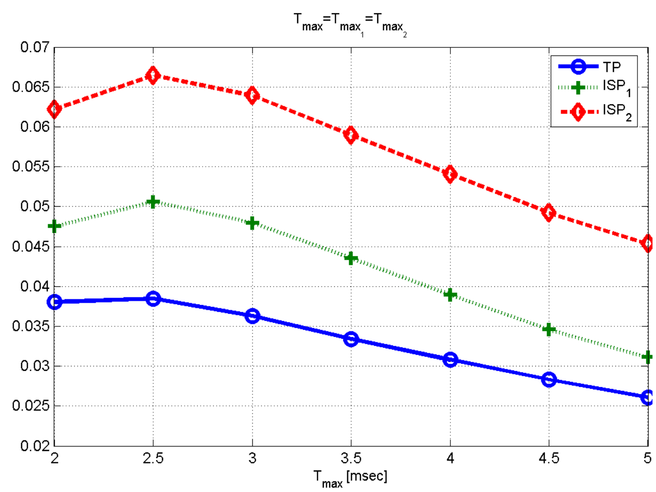


(c) Customers' Costs vs. T_{max}

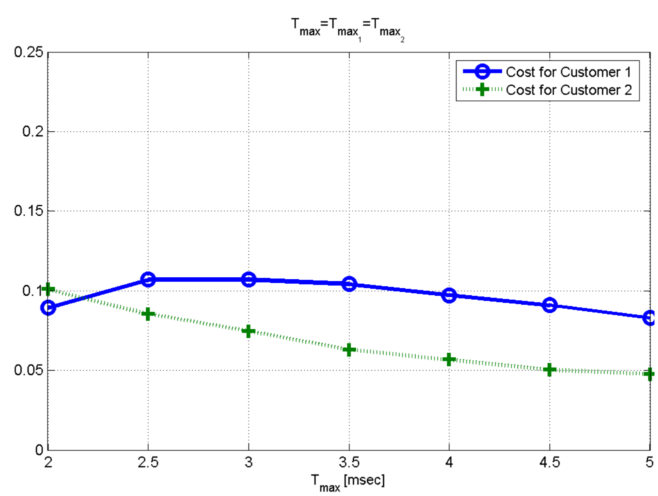
Figure 4.5: Outcome of the two Customers and two ISPs Game vs. T_{max} for Larger Variations than Exponential Packet Lengths for Customer 1 and Deterministic Packet Lengths for Customer 2



(a) λ_{Total}, λ_1, and λ_2 vs. T_{max}



(b) *TP*, *ISP*$_1$, and *ISP*$_2$ vs. T_{max}



(c) Customers' Costs vs. T_{max}

Figure 4.6: Outcome of the two Customers and two ISPs Game vs. T_{max} for Much Larger Variations than Exponential Packet Lengths for Customer 1 and Deterministic Packet Lengths for Customer 2

of the TP and the ISPs are degraded when the variation of customer 1's traffic increases. Finally, from observing Figures 4.4c, 4.5c, and 4.6c, we see how the customers only pay for the amount and quality of service they get. Hence they are treated fairly.

4.3 SCENARIO 2: TWO CUSTOMERS AND THREE ISPs

Now we assume the case of having the TP choose three ISPs along the path from source to destination. Again we assume two customers require their packets to be delivered along this path. Once more we study two cases, one is the General Case of arbitrarily choosing the traffic variation for each customer along with their maximum allowable delay T_{max}. The second case is the Fixed T_{max}, in which $T_{max_1} = T_{max_2} = T_{max}$ and we simulate for different values of T_{max}.

4.3.1 General Case

Here we assume that customer 1's traffic is of exponential packet lengths, whereas customer 2's traffic is of deterministic packet lengths. The initialization values for the different variables in our model are shown in Table 4.3. Once again, we would like to note that some of those values were taken from [14]. Moreover, the simulation results, which represent the Nash equilibrium of the game are given in Table 4.4.

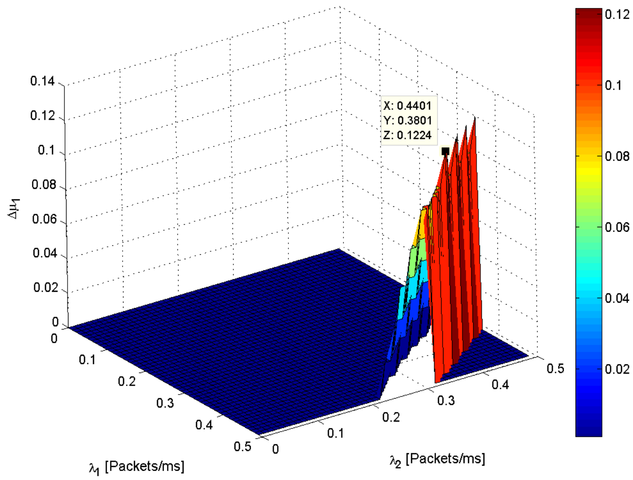
We then include a number of plots representing the outcome of the game. These plots can be found in Figures 4.7, 4.9, and 4.8.

Table 4.3: Initial Values for Simulating the General Case of two Customers and three ISPs.

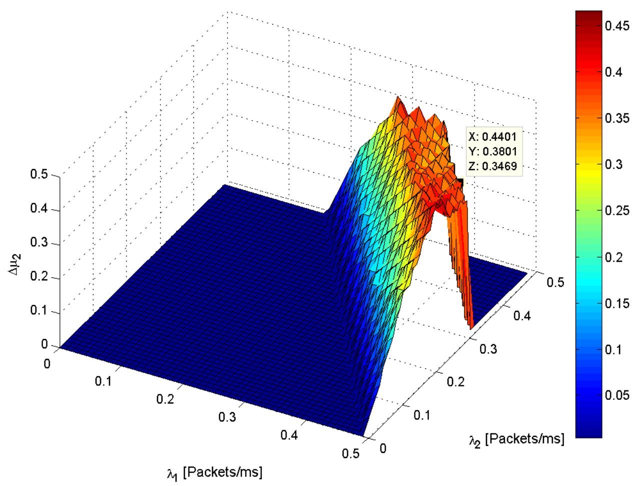
Variable	**Value**
β	$\frac{1}{3}$
λ_{max}	1 [packets/msec]
$\Delta\mu_{max_1}$	1 [packets/msec]
$\Delta\mu_{max_2}$	1 [packets/msec]
T_{max_1}	3 [msec]
T_{max_2}	4 [msec]
M	20%
μ_1	1.4 [packets/msec]
μ_2	1.5 [packets/msec]
μ_3	1.6 [packets/msec]
p_1	0.075
p_2	0.055
p_3	0.035

Table 4.4: Nash equilibrium for the game of two customers and three ISPs.

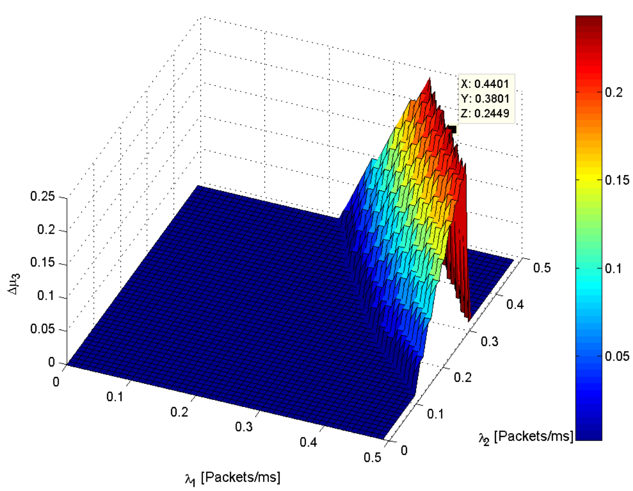
Variable	**Value**
λ_{Nash_1}	0.4401 [packets/msec]
λ_{Nash_2}	0.3801 [packets/msec]
λ_{Total}	0.8202 [packets/msec]
$\Delta\mu_{Nash_1}$	0.1224 [packets/msec]
$\Delta\mu_{Nash_2}$	0.3469 [packets/msec]
$\Delta\mu_{Nash_3}$	0.2449 [packets/msec]
$E_1[T]_{Nash}$	1.2463 [msec]
$E_2[T]_{Nash}$	0.8737 [msec]
$E_3[T]_{Nash}$	0.8754 [msec]
TP_{Nash}	0.0332
ISP_{Nash_1}	0.0685
ISP_{Nash_2}	0.0752
ISP_{Nash_3}	0.0856
P_{C_1}	0.1088
P_{C_2}	0.0574



(a) $\Delta\mu_1$ vs. λ_1 and λ_2

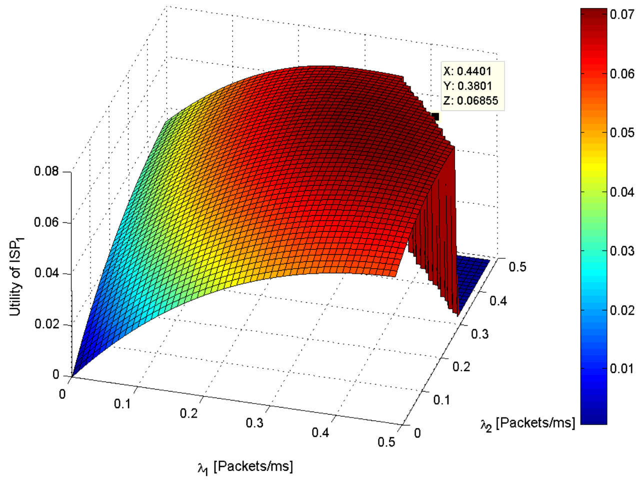


(b) $\Delta\mu_2$ vs. λ_1 and λ_2

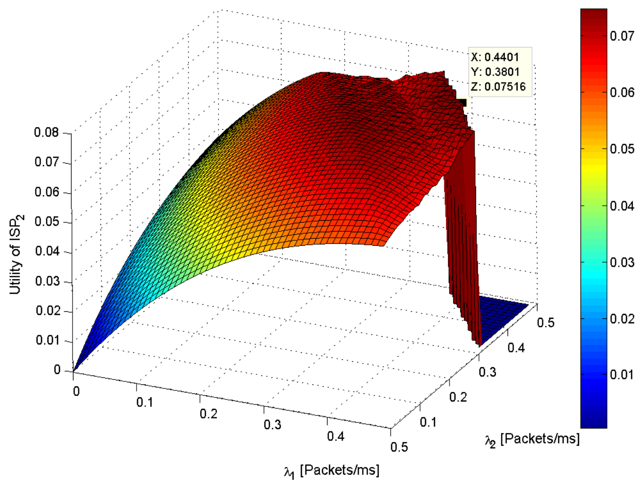


(c) $\Delta\mu_3$ vs. λ_1 and λ_2

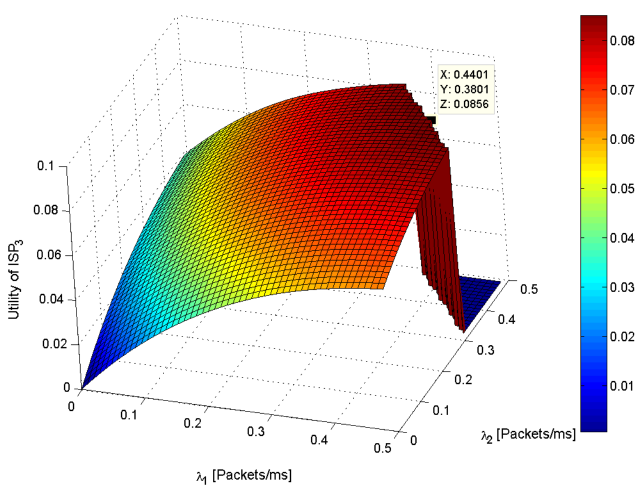
Figure 4.7: Additional Bandwidth Acquired by ISP_1, ISP_2, and ISP_3 as the Outcome of the two Customers and three ISPs Game.



(a) ISP_1 vs. λ_1 and λ_2



(b) ISP_2 vs. λ_1 and λ_2



(c) ISP_3 vs. λ_1 and λ_2

Figure 4.8: The Utility of ISP_1, ISP_2, and ISP_3 as the Outcome of the two Customers and three ISPs Game.

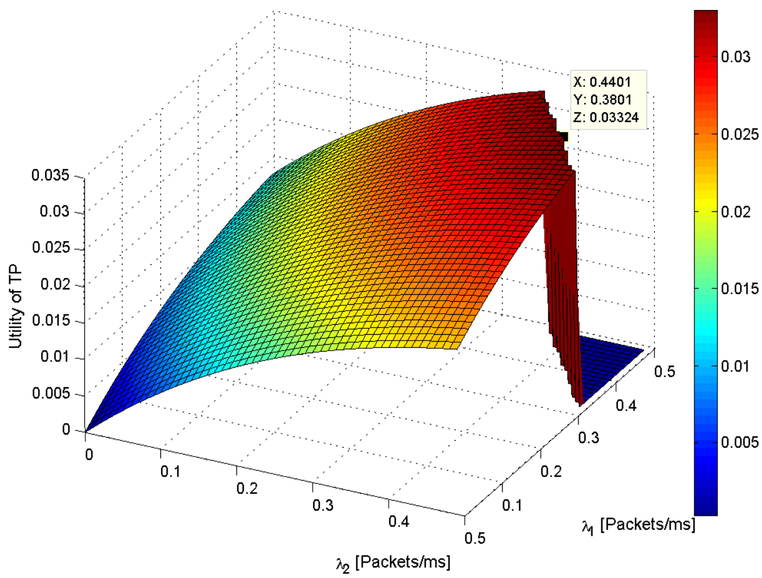
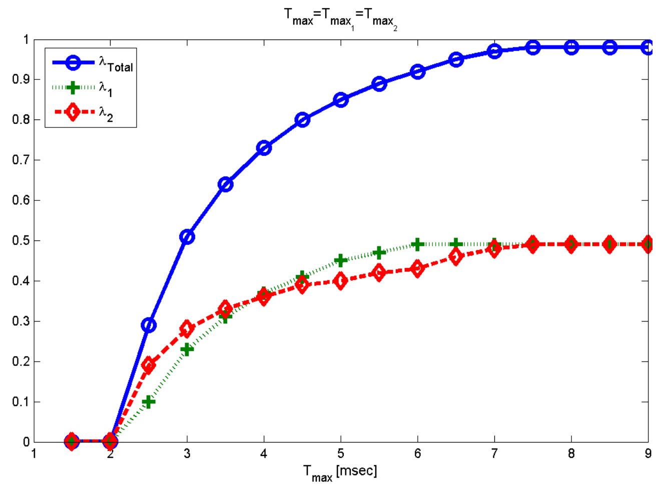


Figure 4.9: The Utility of the TP as the Outcome of the two Customers and three ISPs Game.

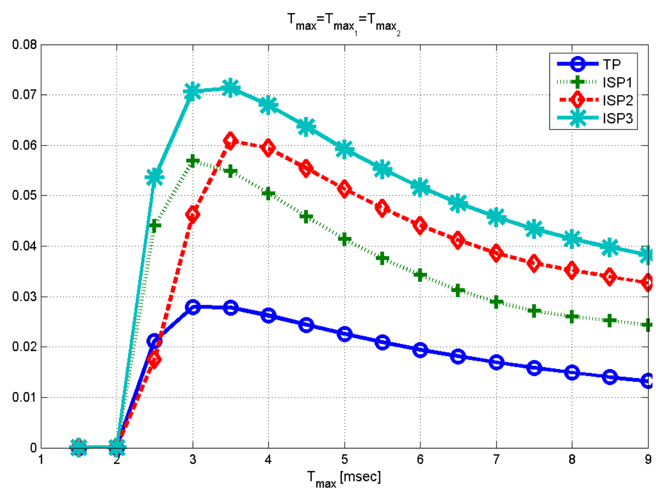
4.3.2 Fixed T_{max} Case

For this section we have $T_{max_1} = T_{max_2} = T_{max}$ and we study the effect of traffic variation on the model alone. We will only consider two cases, one with customer 1 having exponential packet lengths, whereas customer 2 has deterministic packet lengths. The other is when customer 1 has higher than exponential packet lengths, whereas customer 2 has deterministic packet lengths. The initial conditions for the simulation are given by Table 4.3. We then plot Figures 4.10, and 4.11. We can see in both Figures 4.10a and 4.11a that no traffic from either customer is admitted into the system if T_{max} is below 2 msec. This is, because in this case the ISPs can not provide this required QoS to the customers hence their profit would be negative i.e. it turns into loss. Once again we find that the differences between Figures 4.10, and 4.11 are only due to the traffic variation in customer 1's traffic. We observe how the utility functions degrade in the case of higher traffic variation in customer 1's traffic, which is shown in Figures 4.10b and 4.11b.

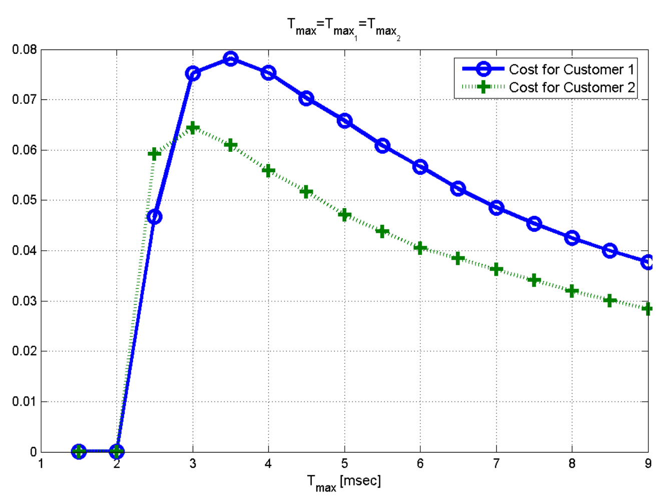
To conclude this section, we would like to note the similarities in the behavior of the model for the two scenarios presented so far. Namely, the scenario of having two customers and two ISPs and having two customers and three ISPs. The trend is the same which gives us a sense of comfort that the model can be expanded to



(a) λ_{Total}, λ_1, and λ_2 vs. T_{max}

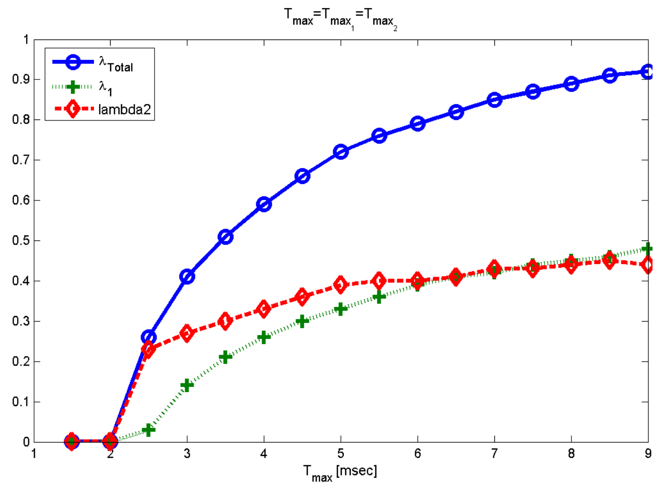


(b) *TP*, *ISP$_1$*, *ISP$_2$*, and *ISP$_3$* vs. T_{max}

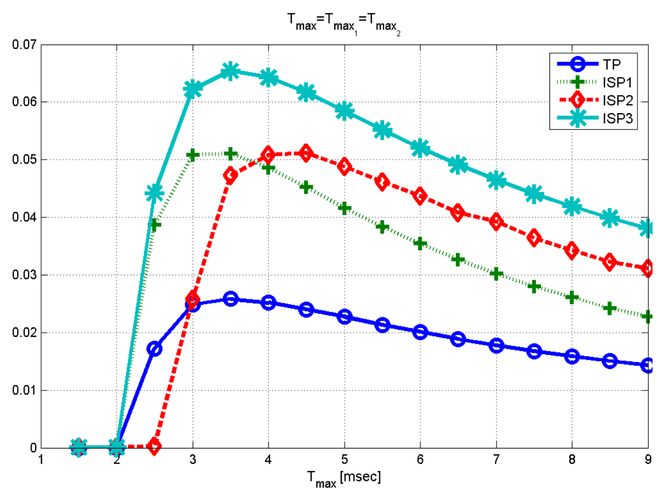


(c) Customers' Costs vs. T_{max}

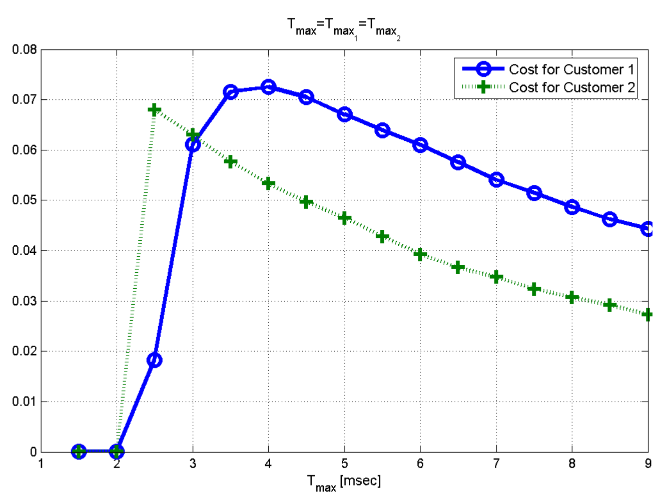
Figure 4.10: Outcome of the two Customers and three ISPs Game vs. T_{max} for Exponential Packet Lengths for Customer 1 and Deterministic Packet Lengths for Customer 2



(a) λ_{Total}, λ_1, and λ_2 vs. T_{max}



(b) TP, ISP_1, ISP_2, and ISP_3 vs. T_{max}



(c) Customers' Costs vs. T_{max}

Figure 4.11: Outcome of the two Customers and three ISPs Game vs. T_{max} for Larger Variations than Exponential Packet Lengths for Customer 1 and Deterministic Packet Lengths for Customer 2

any number of ISPs. In the final two sections of this chapter we continue to expand our basic model to include the scenarios of having three customers and two ISPs and finally three customers and three ISPs.

4.4 SCENARIO 3: THREE CUSTOMERS AND TWO ISPs

In this scenario, the TP chooses two ISPs along the path from the source to the destination. These two ISPs are serving three customers. Again we have two cases, the General Case and the Fixed T_{max} Case. Since we have three customers, hence it is not possible to give plots for the General Case. However, we include plots for the Fixed T_{max} Case.

4.4.1 General Case

Customer 1's traffic is of exponential packet lengths, whereas customer 3's traffic is of deterministic packet lengths. As for customer 2's traffic, its variation is in between that of customers 1 and 2. The initialization values for the different variables in our model are shown in Table 4.5. Some of those values were taken from [14]. Moreover, the simulation results, which represent the Nash equilibrium of the game are given in Table 4.6.

From these results we observe that customer 1 pays the most, because the system admits the most traffic for it. Moreover, it has the most variation in traffic and requires better QoS. We also see how ISP_2 makes more profit than ISP_1, because it promises more overall bandwidth and the overall time spent in its network is less than that of ISP_1.

4.4.2 Fixed T_{max} Case

In this section $T_{max_1} = T_{max_2} = T_{max_3} = T_{max}$ and we study the effect of traffic variation on the model alone. Two cases are considered, one with customer 1 having exponential packet lengths, customer 3 has deterministic packet lengths,

Table 4.5: Initial Values for Simulating the General Case of three Customers and two ISPs.

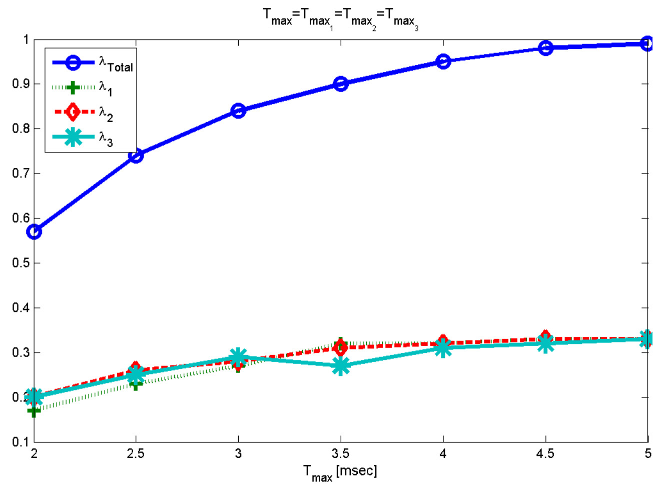
Variable	Value
β	$\frac{1}{3}$
λ_{max}	1 [packets/msec]
$\Delta\mu_{max_1}$	1 [packets/msec]
$\Delta\mu_{max_2}$	1 [packets/msec]
T_{max_1}	3 [msec]
T_{max_2}	3.5 [msec]
T_{max_3}	4 [msec]
M	20%
μ_1	1.1 [packets/msec]
μ_2	1.2 [packets/msec]
p_1	0.075
p_2	0.055

Table 4.6: Nash equilibrium for the game of three customers and two ISPs.

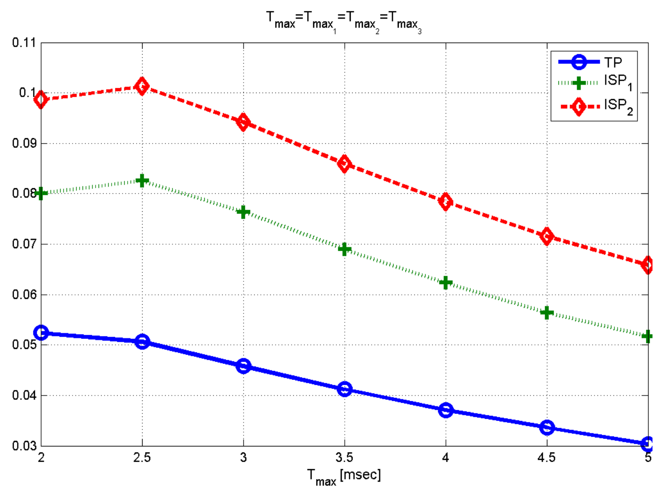
Variable	Value
λ_{Nash_1}	0.3101 [packets/msec]
λ_{Nash_2}	0.2901 [packets/msec]
λ_{Nash_3}	0.2301 [packets/msec]
λ_{Total}	0.8303 [packets/msec]
$\Delta\mu_{Nash_1}$	0.5051 [packets/msec]
$\Delta\mu_{Nash_2}$	0.5354 [packets/msec]
$E_1[T]_{Nash}$	1.1399 [msec]
$E_2[T]_{Nash}$	0.9855 [msec]
TP_{Nash}	0.0406
ISP_{Nash_1}	0.0629
ISP_{Nash_2}	0.0797
P_{C_1}	0.0911
P_{C_2}	0.0696
P_{C_3}	0.0425

and customer 2 having packet lengths of variation between the other two. The other case is when customer 1 has higher than exponential packet lengths, whereas customers 2 and 3 have the same packet length distributions as before. The initial conditions for the simulation are given by Table 4.5. We then plot Figures 4.12, and 4.13.

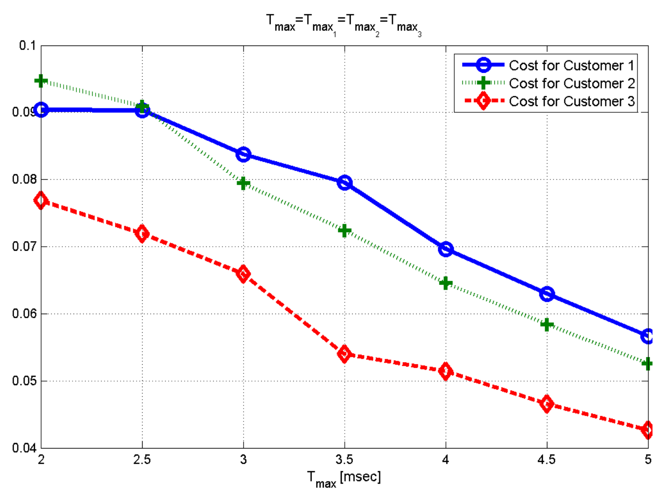
In Figure 4.12a, the curves for λ_1, λ_2, and λ_3 are very close, because the variations of traffic between the customers is not that great. However, in Figure 4.13a, we see a larger difference in the curves of λ_1, λ_2, and λ_3. Specifically, it is observed that customer 1's traffic gets the least rate of admittance into the system, because it has the largest variation in packet lengths. Furthermore, we observe a degradation of profit of ISPs and the TP between Figures 4.12b and 4.13b. It is also noted that in Figures 4.12c and 4.13c the cost paid by each customer depends on the rate of variation of its traffic, the amount of its traffic admitted to the system, and the QoS desired.



(a) λ_{Total}, λ_1, λ_2, and λ_3 vs. T_{max}

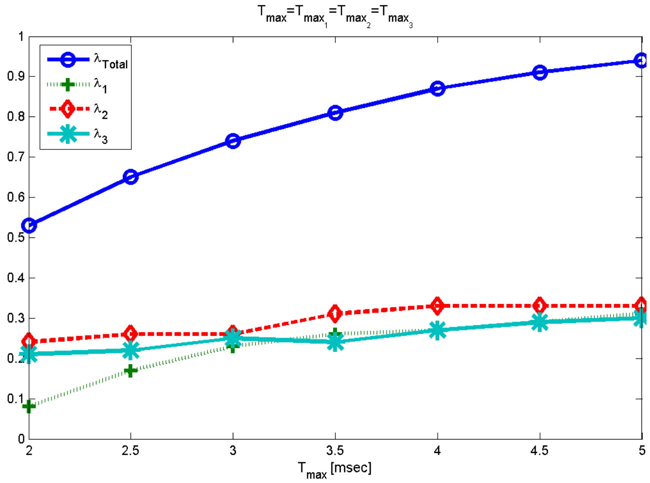


(b) *TP*, *ISP$_1$*, and *ISP$_2$* vs. T_{max}

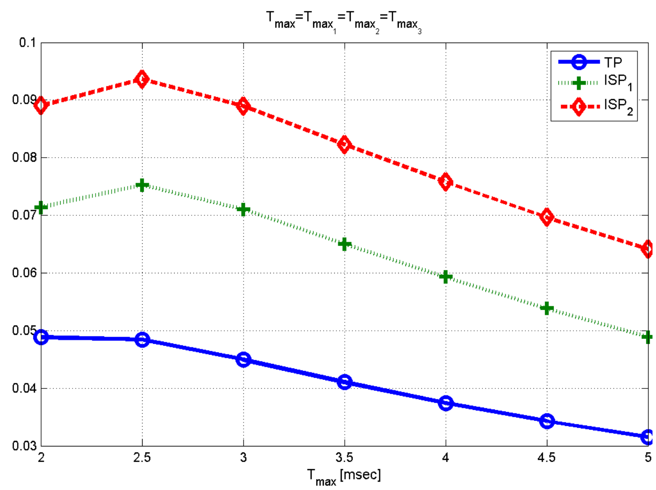


(c) Customers' Costs vs. T_{max}

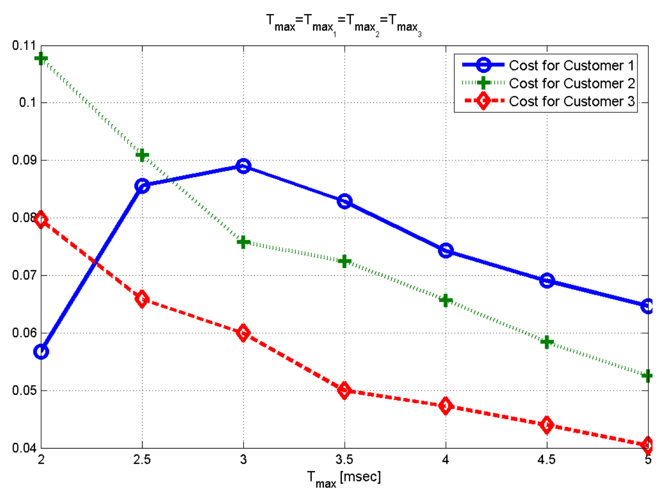
Figure 4.12: Outcome of the three Customers and two ISPs Game vs. T_{max} for Exponential Packet Lengths for Customer 1, Deterministic Packet Lengths for Customer 3, and in between for Customer 2



(a) λ_{Total}, λ_1, λ_2, and λ_3 vs. T_{max}



(b) *TP*, *ISP₁*, and *ISP₂* vs. T_{max}



(c) Customers' Costs vs. T_{max}

Figure 4.13: Outcome of the three Customers and two ISPs Game vs. T_{max} for Larger Variations than Exponential Packet Lengths for Customer 1, Deterministic Packet Lengths for Customer 3, and in between for Customer 2

4.5 SCENARIO 4: THREE CUSTOMERS AND THREE ISPs

We have reached the final scenario in which three ISPs serve three customers. We will also display the General Case and the Fixed T_{max} Case. No figures are shown for the General Case, because four dimensions are needed to draw. We give only the results. However, for the Fixed T_{max} Case, plots are presented.

4.5.1 General Case

We combine the assumptions made over the previous two scenarios here. Customer 1's traffic is of exponential packet lengths, whereas customer 3's traffic is of deterministic packet lengths. As for customer 2's traffic, its variation is in between that of customers 1 and 2. The initialization values for the different variables in our model are shown in Table 4.7. Some of those values were taken from [14]. The simulation results, which represent the Nash equilibrium of the game are given in Table 4.8.

Once again the results are consistent with the results of the other three scenarios. ISP_3's profit is the highest due to its highest contribution. Customer 1 pays the most monetary amount due to its high admittance rate, high variation in traffic, and high QoS demand.

4.5.2 Fixed T_{max} Case

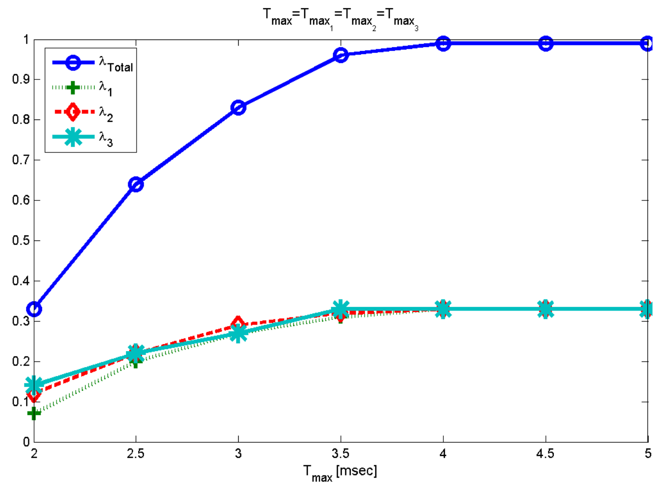
In this case $T_{max_1} = T_{max_2} = T_{max_3} = T_{max}$ and we only study the effect of traffic variation. There is a slight difference from scenario 3 in regards to the packet lengths. Two cases are considered, one with customer 1 having exponential packet lengths, customer 3 has deterministic packet lengths, and customer 2 having packet lengths of variation between the other two. The other case is when customer 1 has higher than exponential packet lengths, whereas customer 2 has exponential packet lengths, and customer 3 has deterministic packet lengths. The initial conditions for the simulation are given in Table 4.7. We then plot figures 4.14, and 4.15.

Table 4.7: Initial Values for Simulating the General Case of three Customers and three ISPs.

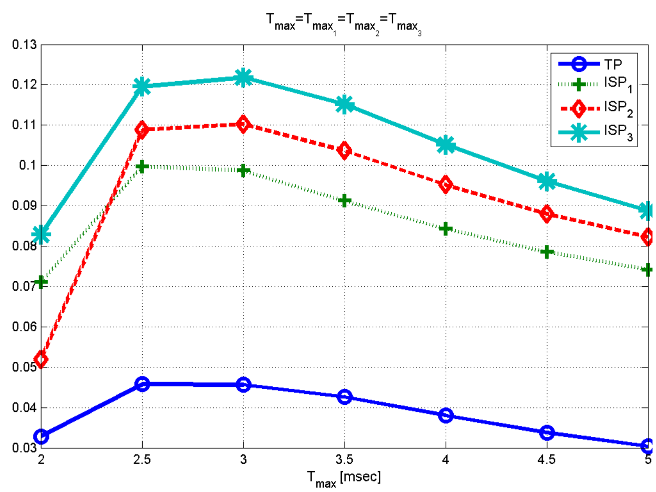
Variable	**Value**
β	$\frac{1}{3}$
λ_{max}	1 [packets/msec]
$\Delta\mu_{max_1}$	1 [packets/msec]
$\Delta\mu_{max_2}$	1 [packets/msec]
T_{max_1}	3 [msec]
T_{max_2}	3.5 [msec]
T_{max_3}	4 [msec]
M	20%
μ_1	1.4 [packets/msec]
μ_2	1.5 [packets/msec]
μ_3	1.6 [packets/msec]
p_1	0.075
p_2	0.055
p_3	0.035

Table 4.8: Nash equilibrium for the game of three customers and three ISPs.

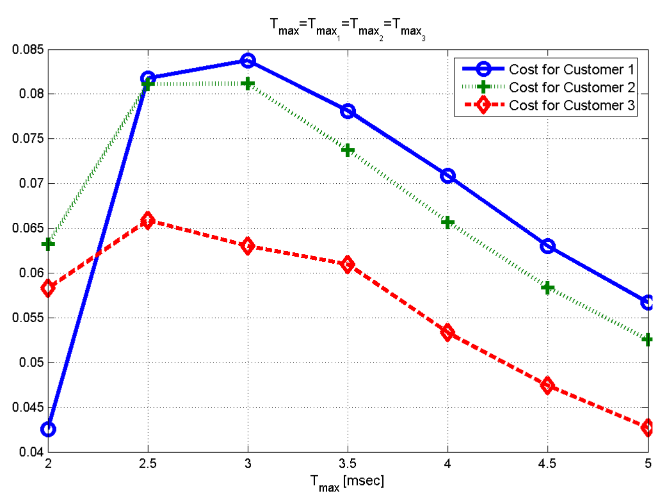
Variable	**Value**
λ_{Nash_1}	0.3001 [packets/msec]
λ_{Nash_2}	0.2801 [packets/msec]
λ_{Nash_3}	0.2401 [packets/msec]
λ_{Total}	0.8203 [packets/msec]
$\Delta\mu_{Nash_1}$	0.1724 [packets/msec]
$\Delta\mu_{Nash_2}$	0.2069 [packets/msec]
$\Delta\mu_{Nash_3}$	0.3448 [packets/msec]
$E_1[T]_{Nash}$	1.1689 [msec]
$E_2[T]_{Nash}$	1.0023 [msec]
$E_3[T]_{Nash}$	0.8023 [msec]
TP_{Nash}	0.0402
ISP_{Nash_1}	0.0853
ISP_{Nash_2}	0.0958
ISP_{Nash_3}	0.1058
P_{C_1}	0.1059
P_{C_2}	0.0681
P_{C_3}	0.0437



(a) λ_{Total}, λ_1, λ_2, and λ_3 vs. T_{max}

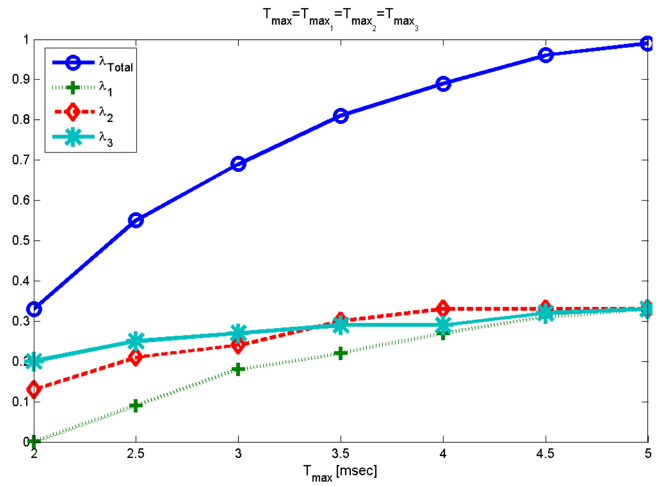


(b) TP, ISP_1, ISP_2, and ISP_3 vs. T_{max}

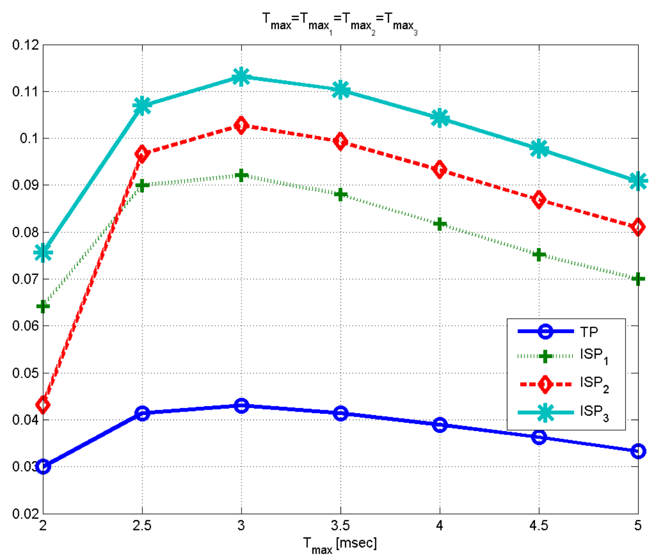


(c) Customers' Costs vs. T_{max}

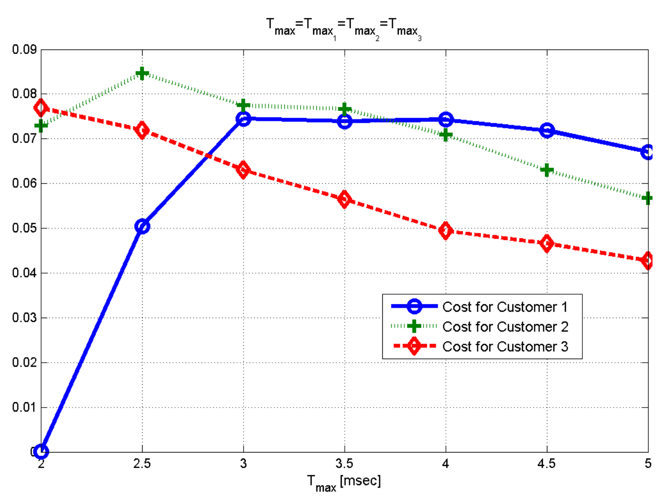
Figure 4.14: Outcome of the three Customers and three ISPs Game vs. T_{max} for Exponential Packet Lengths for Customer 1, Deterministic Packet Lengths for Customer 3, and Packet Length Distribution in between the two for Customer 2



(a) λ_{Total}, λ_1, λ_2, and λ_3 vs. T_{max}



(b) TP, ISP_1, ISP_2, and ISP_3 vs. T_{max}



(c) Customers' Costs vs. T_{max}

Figure 4.15: Outcome of the three Customers and three ISPs Game vs. T_{max} for Larger Variations than Exponential Packet Lengths for Customer 1, Exponential Packet Lengths for Customer 2, and Deterministic Packet Lengths for Customer 3

The results seen from the figures are self-explanatory. They are consistent with the previous scenarios. The conclusion we really want to make is that our model can be expanded to more customers and supported by more ISPs.

4.6 A NAIVE CASE

After demonstrating how our pricing model works for four different scenarios and how it can be extended to a larger number of customers and ISPs, a simple question may arise. Can we think of an easy naïve way of achieving the same results? Is there a simpler way of doing things in order to compare it with our more complex algorithm that uses game theory for optimization?

We will try to answer this question in this section by taking the first scenario of having two customers and two ISPs. Customer 1's traffic has higher coefficient of variation than that of exponential packet lengths, whereas customer 2 has deterministic packet lengths.

A naïve way would be to consider the TP to be non-sophisticated. The problem is that we do not know the actual rate of arrival of the customers' traffic, λ_1 and λ_2, which is needed to optimize the utility functions. Hence, the TP chooses λ_1 and λ_2 according to some pre-defined ratios. So we consider ratios of λ_1 and λ_2 such that $\lambda_{Total} = \lambda_1 + \lambda_2$, where $\lambda_{Total} \leqslant \lambda_{max_{Total}} = \sum_{j=1}^{m} \lambda_{max_j}$. We have $m = 2$ for two customers and $\lambda_{max_{Total}} = 1$. We chose the ratios for λ's as follows:

$$\lambda_1 = \frac{c_2 T_{max_1}}{c_1 T_{max_2} + c_2 T_{max_1}} \tag{4.1}$$

and

$$\lambda_2 = \frac{c_1 T_{max_2}}{c_1 T_{max_2} + c_2 T_{max_1}} \tag{4.2}$$

where c_1 and c_2 are the coefficients of variation of customer 1 and customer 2's traffic.

We first need to calculate the value of λ_{Total}. To do so, we assume that our system contains only one customer. Hence we apply our model for one customer and two ISPs. From that, we find the λ that is optimal, which we will take to be our λ_{Total}. In Figure 4.16, we see that $\lambda = 0.5011$ packets/ms is the value that maximizes the profit of both ISPs in a fair manner, which is taken to be the average of the two λ's maximizing each ISP's profit alone.

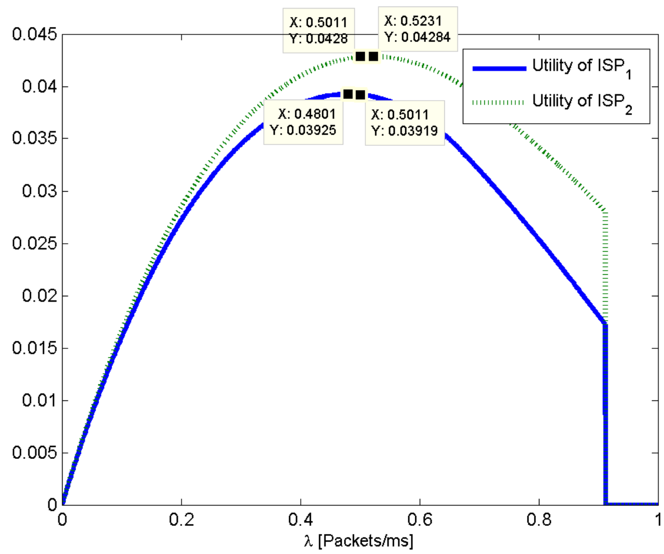


Figure 4.16: Finding Optimal λ to Maximize Utilities of ISPs, assuming one customer and two ISPs.

Then we run the naïve scenario in which λ_{Total} is split into λ_1 and λ_2 according to equations (4.1) and (4.2). The results are found in Table 4.9. *Ratio*, *Ratio$_1$*, and *Ratio$_2$* are defined as:

$$Ratio = \frac{\rho}{E_{Total}} = \frac{\lambda_{Total} E[\sigma]}{E_{Total}} \tag{4.3}$$

where E_{Total} is the total average delay in the whole system end to end.

$$Ratio_1 = \frac{\rho_1}{E_1} = \frac{\lambda_1 E[\sigma_1]}{E_1} \tag{4.4}$$

where E_1 is the total average delay for customer 1 alone throughout the whole system end to end.

$$Ratio_2 = \frac{\rho_2}{E_2} = \frac{\lambda_2 E[\sigma_2]}{E_2} \tag{4.5}$$

where E_2 is the total average delay for customer 2 alone throughout the whole system end to end.

Table 4.9: Comparison between our Model and the Naive Case.

	λ_{Total}	λ_1	λ_2	ISP_1	ISP_2	P_{C_1}	P_{C_2}	*Ratio*	$Ratio_1$	$Ratio_2$
Our Model	0.830	0.450	0.380	0.046	0.061	0.110	0.057	**0.224**	**0.161**	**0.164**
Naïve Case	0.501	0.137	0.364	0.046	0.051	0.051	0.056	**0.168**	**0.060**	**0.148**

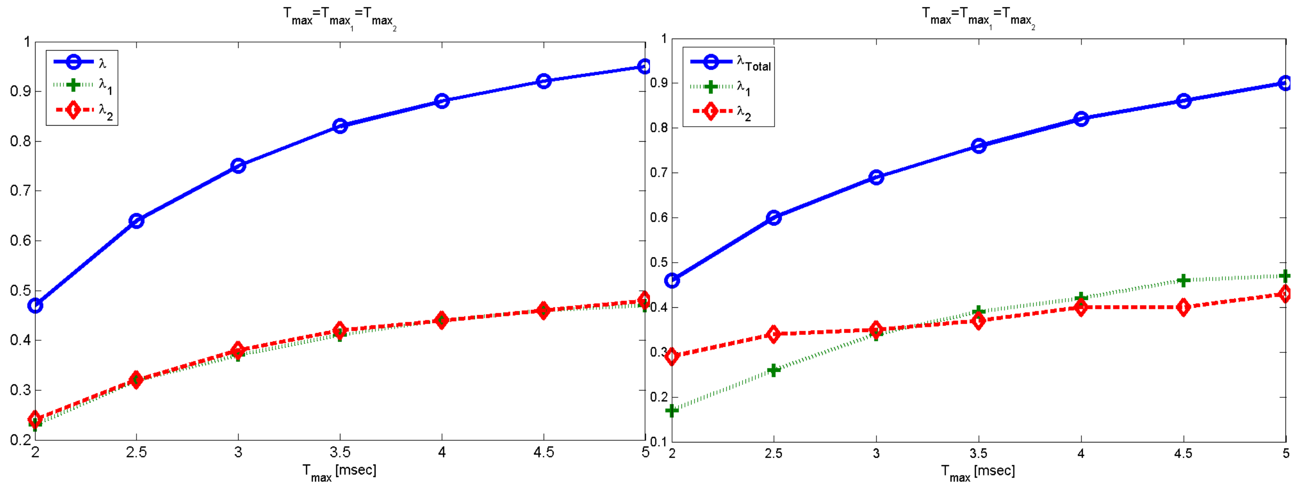
The problem here is that this naïve scenario always gives less λ to the customer of higher coefficient of variation of traffic, because of the ratios assigned. This does not necessarily give the optimal payoffs for the ISPs. Furthermore, it provides poorer service to the customers. As shown by Table 4.9, the values of *Ratio*, $Ratio_1$, and $Ratio_2$ are strictly higher for our model. These variables are just the utilization factor divided by the average delay. High values for these variables are indicators of better service for the customers. Moreover, in the naïve case, we still needed to use our model for the first step, which is to obtain the value for λ_{Total} which is then split into λ_1 and λ_2.

Therefore, there is no easy fix to this problem, and we need to apply game theory to optimize our outcome. So, not only do we need the TP, because it knows the Internet structure and routing topology, but it is also needed to choose a suitable λ_{Total} to maximize the profits of the ISPs and get some profit for itself.

4.7 FINAL REMARKS

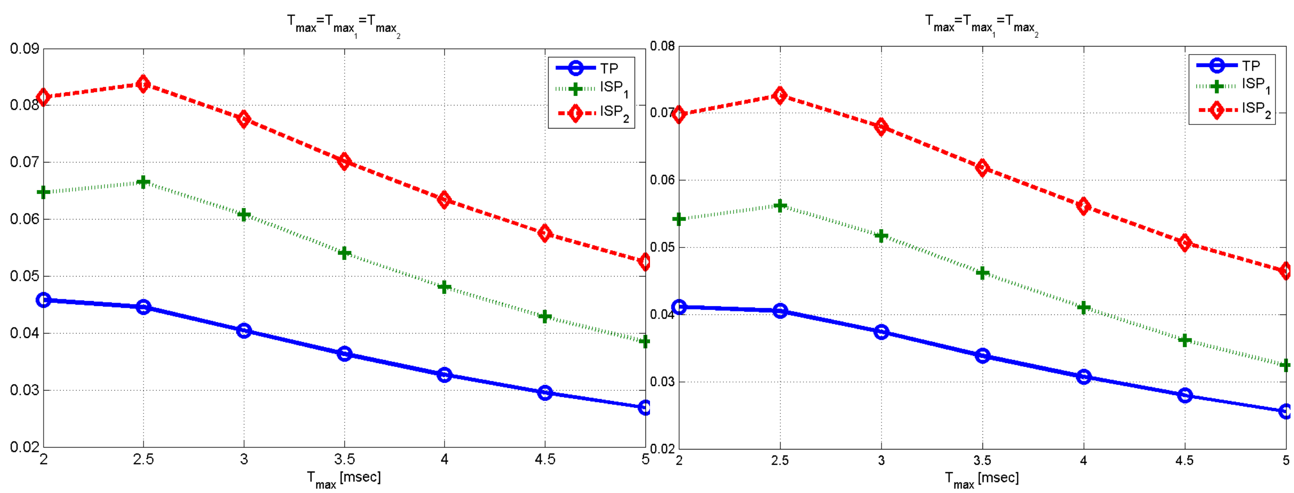
In this section we make a comparison between the results of our model and that of Malhamé et al in [14]. The comparison takes place for a two customers and a two ISPs scenario. Customer 1's traffic has higher coefficient of variation than that of exponential packet lengths, whereas customer 2 has deterministic packet lengths. Moreover, we have $T_{max_1} = T_{max_2} = T_{max}$ and we plot the outcome against T_{max}. We emphasize the necessity of taking the variation in traffic into account when placing a model for pricing in a multiple ISP domain.

As seen in Figure 4.17, if we take the M/M/1 model as in [14], it gives us a false indication that things look better. When adopting an M/G/1 model, we see how the variation in traffic affects the rate of traffic admitted and it provides fairness to customers of less variation in traffic. Moreover, from Figure 4.18, we also see



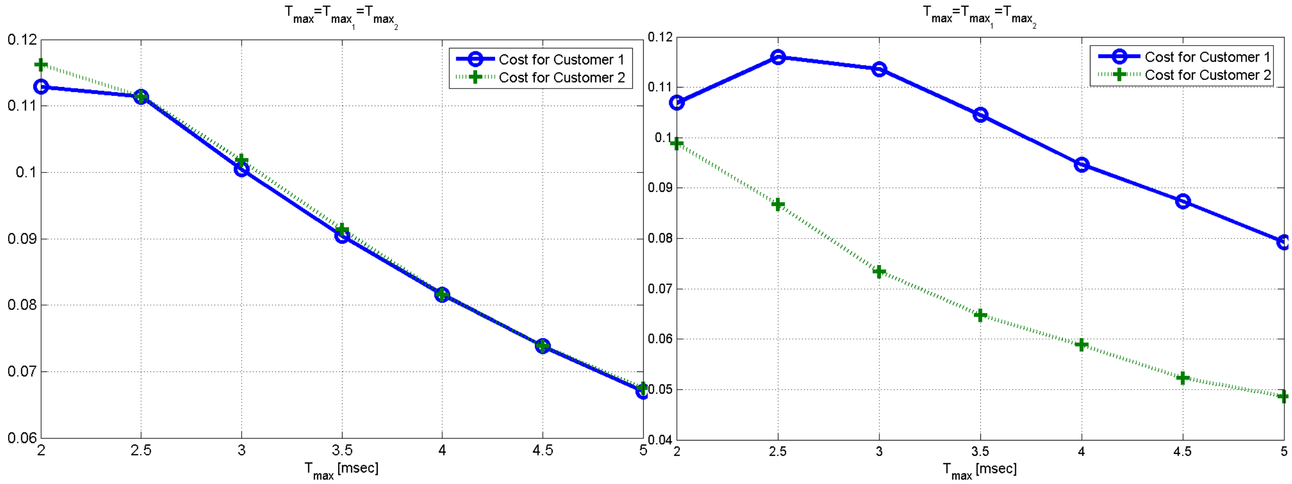
(a) λ_{Total}, λ_1, and λ_2 vs. T_{max} for Malhamé's Model (b) λ_{Total}, λ_1, and λ_2 vs. T_{max} for Our Model

Figure 4.17: Comparing λ's for Our Model and Malhamé's



(a) *TP*, *ISP₁*, and *ISP₂* vs. T_{max} for Malhamé's Model (b) *TP*, *ISP₁*, and *ISP₂* vs. T_{max} for Our Model

Figure 4.18: Comparing Utility functions of the TP and the ISPs for Our Model and Malhamé's



(a) Customers' Costs vs. T_{max} for Malhamé's Model (b) Customers' Costs vs. T_{max} for Our Model

Figure 4.19: Comparing Customers' Costs for Our Model and Malhamé's

how the M/M/1 model gives a false indication of higher profits, whereas our model gives a more realistic measure of profit. It is also shown in Figure 4.19 that our model gives fairness to customers. The customer with more variations in its traffic needs to pay more than the one with less variations.

We also show that our model provides better service for customers. In Figure 4.20, we have a chart which shows that the delays from our model are less than that of Malhamé's case. We note that although customers pay more, because of their variable delays, however, they get better service in terms of their overall delay if they adopt our model. So not only does our model provide more accurate payoffs for the ISPs, it also provides the customers with enhanced service as a return for their higher costs, while adding fairness between customers. Moreover, we would like to note that Figure 4.20 shows strictly less delay times for customer 2's traffic in contrast to customer 1's traffic. The reason behind this, is that customer's maximum tolerable delay (T_{max_1}) is initiated to be 3 msec, whereas that of customer 2, is initiated to be 4 msec.

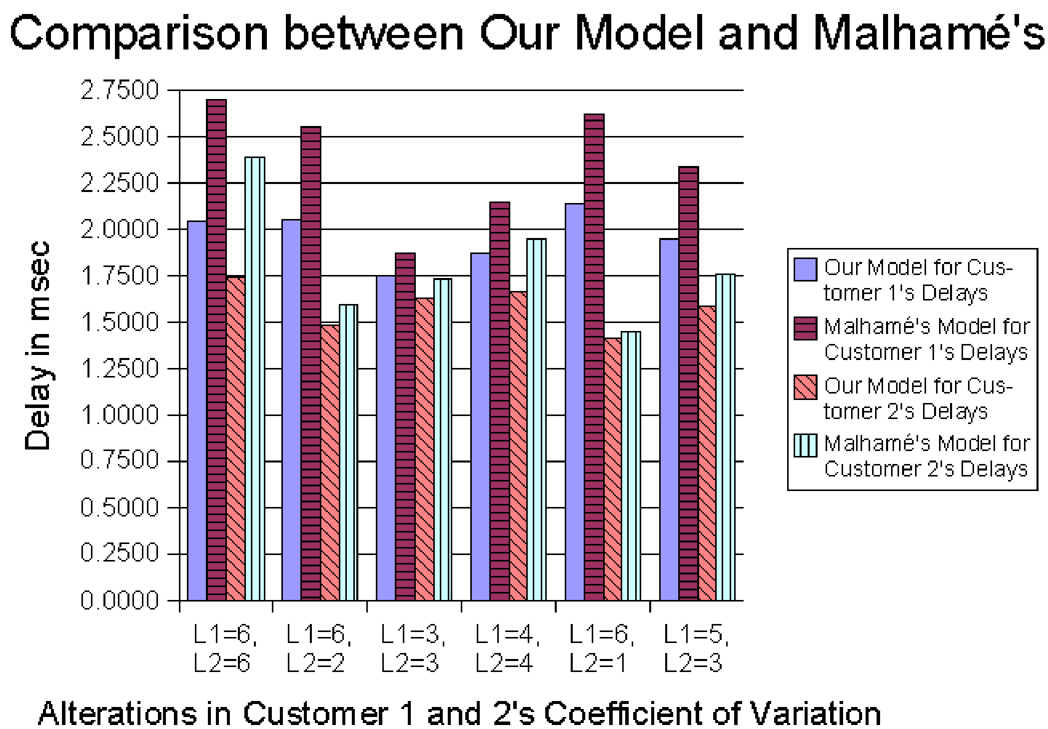


Figure 4.20: Comparison between Our Model and Malhamé's Model in terms of Customer Delays. Taking into consideration that $T_{max_1} = 3$ msec and $T_{max_2} = 4$ msec.

Chapter 5

Conclusion

In this thesis, we have studied many issues. We took a look at the structure of the current Internet. We saw how it has a hierarchical form. Moreover, we discussed how peering is becoming popular and the advantages and disadvantages of peering. We also looked at the future of the Internet structure and how it tends to flatten as the traffic volume increases gradually over the years.

Furthermore, we discussed the factors of a good pricing scheme which are:

- A customer's preferences
- Efficiency of the network
- The way an ISP monitors a customer's usage

We saw how balancing these factors would result in a good pricing scheme. A number of pricing schemes were presented. They were studied to see how well they comply with the three above mentioned factors. We saw how the cumulus pricing scheme was a good candidate, but needed to be applied in a multiple ISP environment.

Pricing in a multiple ISP domain was then discussed. We took a brief look at some papers written in this field. Most of these papers used Game Theory to solve this issue. Different models were presented along with the players of each game.

We then showed why we chose the model implemented by [14] as a base for our work. This was, because it fulfills the three main factors for a good pricing scheme along with having a simple structure that can be implemented in a multiple ISP domain. Some extensions to [14] were made to include more than one customer to the game. These customers have various traffic classes with varying packet lengths. As a result of varying packet lengths, each ISP is now represented by an M/G/1 queue instead of an M/M/1 queue.

Our model was then described in detail. We presented all the assumptions made, the game framework, and formulation. Then numerical results were given for four different scenarios of the game. Finally, we showed the effectiveness of adopting our model as opposed to the model by [14] through a comparison. This result really shows that traffic variation is an important factor to take into consideration when placing a pricing scheme. We also thought of an easy way to solve this problem by developing a naïve scheme. However, it turned out to be inadequate to solve our problem.

We conclude this thesis by noting that the TP could be assigned to choose ISPs along multiple paths, which is considered an open problem. Furthermore, we would like to mention that further research of pricing is needed in areas such as mobile networks, multi-hop radio networks, and ad-hoc wide-area 802.11b networks, which do not have any a priori infrastructure. This complicates accounting and resource allocation problems.

APPENDICES

Appendix A

Pseudocode for Two ISPs serving Two Customers

In this section, we present a pseudocode of the program developed in MATLAB for our pricing model. The scenario of two customers and two ISPs is shown here, however, it is straight forward to extend the code for various scenarios.

<u>**Initializations**</u>

$\beta = \frac{1}{3}$
$\lambda_1 = \lambda_2 \in [0, 0.5]$ (100 points)
$\lambda_{max} = 1$
$\Delta\mu_1 = \Delta\mu_2 \in [0, 1]$ (100 points)
$\Delta\mu_{max_1} = \Delta\mu_{max_2} = 1$
$T_{max_1} = 3$
$T_{max_2} = 4$
$L_1 = 2$ (This is for exponential packet lengths)
$L_2 = 1$ (This is for deterministic packet lengths)
$M = 0.2$
$\mu_1 = 1.1$
$\mu_2 = 1.2$
$p_1 = 0.075$
$p_2 = 0.055$

Procedure Begins

For $u = 1$ to $length(\lambda_2)$

For $i = 1$ to $length(\lambda_1)$

$C_v(\lambda_{1_i}) = \exp(-\frac{\lambda_{1_i}}{0.75})$

$C_v(\lambda_{2_u}) = \exp(-\frac{\lambda_{2_u}}{0.75})$

$\lambda = \lambda_{1_i} + \lambda_{2_u}$

For $j = 1$ to $length(\Delta\mu_1)$

For $k = 1$ to $length(\Delta\mu_2)$

$E_1^1[\sigma] = \frac{1}{\mu_1 + \Delta\mu_{1_j}}$, $E_2^1[\sigma] = \frac{1}{\mu_1 + \Delta\mu_{1_j}}$

$E_1^2[\sigma] = \frac{1}{\mu_2 + \Delta\mu_{2_k}}$, $E_2^2[\sigma] = \frac{1}{\mu_2 + \Delta\mu_{2_k}}$

$E_1^1[\sigma^2] = \frac{L_1}{(\mu_1 + \Delta\mu_{1_j})^2}$, $E_2^1[\sigma^2] = \frac{L_2}{(\mu_1 + \Delta\mu_{1_j})^2}$

$E_1^2[\sigma^2] = \frac{L_1}{(\mu_2 + \Delta\mu_{2_k})^2}$, $E_2^2[\sigma^2] = \frac{L_2}{(\mu_2 + \Delta\mu_{2_k})^2}$

$E^1[\sigma] = \frac{E_1^1[\sigma]\lambda_{1_i} + E_2^1[\sigma]\lambda_{2_u}}{\lambda}$

$E^2[\sigma] = \frac{E_1^2[\sigma]\lambda_{1_i} + E_2^2[\sigma]\lambda_{2_u}}{\lambda}$

$E = E_1 + E_2$

$T_{max} = \min(T_{max_1}, T_{max_2})$

If $E \leqslant T_{max}$

$a = 1$

else

$a = 0$

end if

$S_1 = \frac{T_{max} - E_1}{T_{max}}$

$S_2 = \frac{T_{max} - E_2}{T_{max}}$

$P_C = \frac{C_v(\lambda_{1_i})\,\lambda_{1_i}}{T_{max_1}} + \frac{C_v(\lambda_{2_u})\,\lambda_{2_u}}{T_{max_2}}$

$ISP_1 = (1 - M)\,a\,P_C\,S_1 - p_1\,\Delta\mu_1$

$ISP_2 = (1 - M)\, a\, P_C\, S_2 - p_2\, \Delta\mu_2$

end for

end for

Calculate $\Delta\mu_1^*$ and $\Delta\mu_2^*$ by solving for NASH in ISP_1 and ISP_2

$$E_1^* = \frac{1}{\mu_1 + \Delta\mu_1^*} + \frac{(c_1^2 + 1)\,\lambda}{2(\mu_1 + \Delta\mu_1^* - \lambda)(\mu_1 + \Delta\mu_1^*)}$$

$$E_2^* = \frac{1}{\mu_2 + \Delta\mu_2^*} + \frac{(c_2^2 + 1)\,\lambda}{2(\mu_2 + \Delta\mu_2^* - \lambda)(\mu_2 + \Delta\mu_2^*)}$$

$$E^* = E_1^* + E_2^*$$

If $E \leqslant T_{max}$

$a = 1$

else

$a = 0$

end if

$TP_i = M\, a\, P_C$

end for

end for

Find λ_1^* and λ_2^* that maximize TP

Hence $\lambda^* = \lambda_1^* + \lambda_2^*$

Find $\Delta\mu_1^{**}$ and $\Delta\mu_2^{**}$ that correspond to λ_1^* and λ_2^* (These are the NASH equilibrium values)

Calculate ISP_{Nash_1} and ISP_{Nash_2} using λ_1^*, λ_2^*, $\Delta\mu_1^{**}$, and $\Delta\mu_2^{**}$

Display the values of λ^*,λ_1^*, λ_2^*, TP_{Nash}, $\Delta\mu_1^{**}$, $\Delta\mu_2^{**}$, ISP_{Nash_1} , and ISP_{Nash_2}

Appendix B

Brief Explanation of Nash Equilibrium used in our Program

This section presents the method implemented in our program to solve for the Nash equilibrium in pure strategies in the game played between the ISPs (group of followers). It is based on the cell-by-cell inspection method for simultaneous games, which is shown in [4]. We only discuss the method for a game of two players, yet it is very easy to extend it to a game of n players.

The best way to demonstrate this method is by using an example. In Figure B.1a, we give a numerical example for two players with three strategies in a simultaneous game. In each cell, first the payoff of player 1 is displayed then followed by the payoff of player 2. The three strategies for player 1 are Up, Middle, and Down, whereas those of player 2 are High, Medium, and Low.

The cell-by-cell inspection method works as follows. First, we assume that player 1 chooses the strategy Up. In this case player 2 is rational and chooses the strategy that will maximize its payoff. This strategy is Medium, because it gives it a payoff of 70. Hence player 2 eliminates its High and Low strategies for the case of player 1 playing Up. The method continues in the same manner for the strategies Middle and Down of player 1. By now player 2 would have eliminated all the choices it would not pick. Figure B.1b reflects our discussion so far.

Now, the exact opposite is done. Player 2 keeps its strategies constant while player 1 eliminates its choices. After doing this exercise we reach the state shown in

Figure B.1c. The Nash equilibrium is found to be the cell which has no eliminations. In our example it is the shaded one corresponding to the Middle and Medium strategies of player 1 and player 2 respectively.

We used the concept of this method of cell-by-cell inspection and applied it to our program with some minor necessary modifications to be adequate to the MATLAB environment.

Player 2

Player 1

	High	Medium	Low
Up	60, 60	36, 70	36, 35
Middle	70, 36	50, 50	30, 35
Down	35, 36	35, 50	25, 25

(a) An example of simultaneous game with two players and three strategies.

Player 2

Player 1

	High	Medium	Low
Up	60, ~~60~~	36, 70	36, ~~35~~
Middle	70, ~~36~~	50, 50	30, ~~35~~
Down	35, ~~36~~	35, 50	25, ~~25~~

(b) Hold player 1's strategies constant while eliminating player 2's strategies.

Player 2

Player 1

	High	Medium	Low
Up	~~60~~, ~~60~~	~~36~~, 70	36, ~~35~~
Middle	70, ~~36~~	50, 50	~~30~~, ~~35~~
Down	~~35~~, ~~36~~	~~35~~, 50	~~25~~, ~~25~~

(c) Hold player 2's strategies constant while eliminating player 1's strategies.

Figure B.1: Example to illustrate how to solve for Nash equilibrium using cell-by-cell inspection.

References

[1] J. Altmann and K. Chu, "How to charge for network services– flat-rate or usage-based?" *Special Issue on Networks and Economics, Computer Networks Journal*, August 2001. 1, 2, 13, 14, 15, 16

[2] Bob Briscoe, "The direction of value flow in connectionless networks," in *Networked Group Communication*, 1999, pp. 244–269. [Online]. Available: citeseer.ist.psu.edu/briscoe99direction.html 18

[3] Xi-Ren Cao, Hong-Xia Shen, R. Milito, and P. Wirth, "Internet pricing with a game theoretical approach: concepts and examples," *Networking, IEEE/ACM Transactions on*, vol. 10, no. 2, pp. 208–216, Apr 2002. 21

[4] Avinash K. Dixit and Susan Skeath, *Games of Strategy, Second Edition.* W. W. Norton, March 2004. 71

[5] Drew Fudenberg and Jean Tirole, *Game Theory.* MIT Press, October 1991.

[6] Linhai He and J. Walrand, "Pricing and revenue sharing strategies for internet service providers," *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, vol. 1, pp. 205–216 vol. 1, 13-17 March 2005. 21, 22

[7] T. Henderson, J. Crowcroft, and S. Bhatti, "Congestion pricing. paying your way in communication networks," *Internet Computing, IEEE*, vol. 5, no. 5, pp. 85–89, Sep/Oct 2001. 14, 17

[8] J. MacKie-Mason and H. Varian, "Pricing the internet," in Public Access to the Internet, Brian Kahin and James Keller, eds. Cambridge, MA: MIT Press,1995: 269-314. 17

[9] J. MacKie-Mason and H. Varian, "Pricing congestible network resources," *IEEE Journal on Selected Areas in Communications*, 1995. 17

[10] W. B. Norton, "A business case for isp peering," *White Paper*, 2002. [Online]. Available: http://www.equinix.com 8, 9

[11] W. B. Norton, "The evolution of the u.s. internet peering ecosystem," *White Paper*, 2003. [Online]. Available: http://www.equinix.com 4, 5, 10, 11

[12] Andrew Odlyzko, "Internet pricing and the history of communications," *Computer Networks (Amsterdam, Netherlands: 1999)*, vol. 36, no. 5–6, pp. 493–517, 2001. [Online]. Available: http://citeseer.ist.psu.edu/odlyzko01internet.html 1

[13] P. Reichl, P. Flury, J. Gerke, and B. Stiller, "How to overcome the feasibility problem for tariffing internet services: The cumulus pricing scheme," 2001. [Online]. Available: citeseer.ist.psu.edu/reichl01how.html 11, 17, 18, 19

[14] Soheil Saberi, Roland P. Malhamé, and Lorne G. Mason, "Pricing for qos provisioning across multiple internet service provider domains." NetCoop07, Avignon, France, June 2007. 3, 22, 23, 24, 35, 43, 50, 55, 61, 65, 66

[15] Srinivas Shakkottai and R. Srikant, "Economics of network pricing with multiple isps," *IEEE/ACM Trans. Netw.*, vol. 14, no. 6, pp. 1233–1245, 2006. 21

[16] Burkhard Stiller, Jan Gerke, Peter Reichl, and Placi Flury, "Management of differentiated services usage by the cumulus pricing scheme and a generic internet charging system." [Online]. Available: citeseer.ist.psu.edu/487587.html

[17] Burkhard Stiller, Peter Reichl, Jan Gerke, and Placi Flury, "A generic and modular internet charging system for the cumulus pricing scheme," *J. Network Syst. Manage.*, vol. 9, no. 3, 2001. 2

[18] William Vickrey, "Counterspeculation, auctions, and competitive sealed tenders," *The Journal of Finance*, vol. 16, no. 1, pp. 8–37, 1961. [Online]. Available: http://www.jstor.org/stable/2977633 17