A Framework for Pricing Virtual Circuit and Virtual Path Services in ATM Networks

Nikolaos Anerousis[a] *and Aurel A. Lazar*[b]

a) AT&T Research, 600 Mountain Avenue, Murray Hill, NJ 07974-0636, E-mail: nikos@research.att.com. WWW: http://www.research.att.com/~nikos.

b) Department of Electrical Engineering and Center for Telecommunications Research, Columbia University, New York, NY 10027, e-mail: aurel@ctr.columbia.edu

Abstract

Future ATM networks are expected to provide two basic connection services: Switched Virtual Circuit (VC) and Virtual Path (VP). Pricing can be used as a control mechanism to influence the user demand for services, thereby preventing network overloads both at the cell and the call level; it can also be used by the network operator as a means of increasing its revenue. We study an environment where prices are manipulated by the network operator to a) satisfy certain QOS constraints and b) maximize its revenue when possible. Our analysis shows that the problem can be decomposed into two separate optimization problems: one executed by users to determine the optimal service allocation given the current prices, and one employed by the network to find the prices that achieve objectives (a) and (b). By iterating between these two optimizations, prices reach an equilibrium with the following properties: VC prices are higher than VP prices per unit of capacity (to reflect the additional signaling cost); network utility is maximized with respect to prices; and every user's utility is maximized with respect to its current service allocation. We extend our findings to formulate a network-wide pricing optimization problem, where capacity is allocated by the network operator to every source-destination pair based on a measure of "profitability".

1. INTRODUCTION

ATM networks will offer a wide variety of services, ranging from simple unicast connections to Virtual Networks. Broadband services int this environment can be constructed using two fundamental building blocks: Virtual Circuits (in point-to-point or point-to-multipoint mode) and Virtual Paths. In our previous work [ANE95, ANE96a] we presented an integrated management and control framework for these services. Its objective was to provide the necessary monitoring and control capabilities to the network manager in order to guarantee Quality of Service both at the cell and the call level.

At the call level, our control framework produces a VP distribution policy that bounds certain call-level performance measures such as the call blocking probability and the call setup time for the VC service. However, the termination of the algorithm depends on the network load (the VC service demand). For some ranges of network load the call blocking probability constraint cannot be satisfied. This motivated us to investigate control structures that can *influence* the service demand (rather than simply trying to adjust the network configuration to absorb it). The control objective here is to regulate the service request rate at the boundaries of the network. An uncontrolled network load at the call level can cause congestion within the signalling network and deterioration of the Quality of Service (QOS) observed by the users.

One way of achieving this is by blocking a portion of the traffic offered to the network. This solution, however, is not practical if the network wishes to guarantee a certain level of QOS at the call level, such as the blocking probability for requesting a particular type of service. Alternatively, many researchers have proposed an economic model that can directly control user behavior. According to this model, the network manager as the service provider sets prices for various services. Users (service consumers) observe the prices and formulate their demand, by taking into account their own costs, but ignoring the congestion that they might impose on others. Economists refer to this phenomenon as a “congestion externality”. In the economy, resources can be allocated in a way that is beneficial to the network operator and, at the same time, users with diverse requirements achieve maximal benefit (individual optimality). Economists argue that prices internalize the congestion externality by making users face the costs that they impose on other users. An additional argument for the service pricing approach is that prices do not only discourage service demand when congestion is present, but also generate revenue that can be used for capacity expansion [MAC94a].

Pricing mechanisms have been studied in both connectionless and connection oriented networks. An extensive survey can be found in [JOR95]. In most approaches, users are assigned a surplus function which represents their total benefit minus costs. The network sets prices for network resources such as bandwidth and buffer space in order to maximize a utility function. In a cooperative environment, the network uses the total network welfare, defined as the sum of all user surpluses minus costs, as the utility function. This approach is taken in [LOW93], [MUR94], [COC93], [JIA95] and [SAR95]. By contrast, the network operator can set the prices to (selfishly) maximize its surplus. This environment is studied in [PAR92] and [HON95].

This paper proposes an economic approach to controlling the demand for VC and VP services in ATM networks. We introduce the notion of a “pricing agent” which is installed at all network entry points and implements a pricing policy controlled by a centralized network management facility. The pricing agent observes the service demand and possesses “intelligence” that manipulates the prices to achieve a set of control objectives. In our case, these objectives are the service blocking probability and the maximization of revenue from the provided services.

This paper is organized as follows: Section 2 describes the network service model, the pricing agent architecture and presents the pricing optimization problem for a single Source-Destination pair. Section 3 provides a framework for optimizing prices in a network-wide setting. Finally, Section 4 presents the conclusions and the directions of further study.

2. PRICING VC AND VP SERVICES

2.1 Network Model

We assume a network that supports a small number of well defined service classes (e.g., voice, video, data, etc.). We model users as Poisson sources that produce connection requests for traffic class k with rate λ_k and average duration $1/\mu_k$. The ATM network offers two basic unicast connection services: A Virtual Circuit (VC) connection service, suitable for carrying the data of a single connection, and a Virtual Path (VP) service. VPs here are used as a user-to-user service (i.e., they are terminated within or just outside user premises).

Users are connected to the network through a network boundary node, referred to as the network *Point-of-Presence* (POP). Every user submits its request for service to the POP with a signaling message, and the network accepts or denies the request according to capacity availability and other constraints. VC requests are presented to the POP when a user wishes to establish a connection to a user connected to another POP using the VC service. Users with high traffic for a destination POP or a specific user may also find attractive the VP service (which is the equivalent of a leased line in a packet switched environment). In this case, a VP is requested between the user and the destination POP or between the two users. The size of the VP is determined from the user's traffic load and QOS requirements. When the user needs to establish a connection, the connection is routed through the VP to its destination without making a VC service request to the local POP. The benefit for the user is that the connection observes a reduced call setup time since the signaling system of the public network is not involved. This allows the network operator to offer the VP service at a lower cost per unit of capacity. The user however needs to maintain utilization information for the VP and make the appropriate admission control decisions. From an economic perspective, the difference between the VC and the VP service is that in the VC case, the user is paying only for the network resources used, whereas in the VP case for the entire capacity that has been reserved, regardless if it is used or not. Therefore the VP service is beneficial to the users that have enough traffic to keep the VP at a high utilization.

Every POP pair in the network for which service requests are submitted is referred to as a *Source-Destination (SD) pair.* The network manager assigns a pricing agent to every POP. The agent is responsible for advertising service prices to the users on a per SD pair basis and measuring the service demand. The agent charges for the VC service per unit of time and prices can vary between service classes. Similarly, he VP is priced as a function of its total capacity.

2.2 Characterizing user demand

Users can select any amount of VC or VP service based on the current prices and their traffic demands. First, we look at one SD pair only and assume the existence of one traffic class, i.e., every connection request requires the same amount of networking resources. The VP service is provided at price w_x per "circuit" (one circuit corresponds to the networking capacity necessary to handle one call) and unit of time, and the VC service at price w_y per circuit and unit of time. Let also λ be the arrival rate (Poisson) that the user wishes to offer to a particular SD pair in the network. The user demand for the VP and VC services is denoted by x and y respectively. Note that x represents the number of calls that can be accepted in the VP while y is the load (arrival rate) offered to the VC service. We further assume that the demand function is a decreasing function of the prices with the demand reaching 0 for some w_x and w_y. Since all users exhibit this behavior, there exists a set of prices (w_x^*, w_y^*) with which the pricing agent can achieve the desired service demand.

It is worthwhile to note here that the two services (VC and VP) are *not* orthogonal. In fact, users may select either to carry their traffic. The requested amount of each service is determined by solving an optimization problem given the current prices and the user service values (the amount that the user is willing to "pay" for the service).

Without loss of generality, let us assume that the average call holding time is exponentially distributed with mean one time unit. Every user initially requests a VP with capacity x where $x \geq 0$. Subsequent call requests are first offered to the VP. If the VP is full or x=0, the user chooses to either block the call with probability 1-a or attempt to establish it using the VC service with probability a. The procedure is shown in Figure 1.

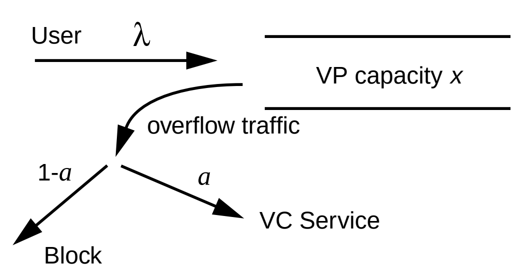


Figure 1: The user model for establishing calls

Given the current prices and rate λ, the user must calculate the x and a such that its utility is maximized. Assuming that the overflow traffic is approximately Poisson, the offered load to the VC service is given by $y = a\lambda E(\lambda, x)$, where $E()$ is the Erlang blocking formula. If the network provides the VC service for this particular SD pair with blocking probability P, the user throughput is given by $\gamma = \lambda[1 - E(\lambda, x)[1 - a(1 - P)]]$.

Now assume that the user values every call that goes through the system at w_u, and loses w_b for every call that is blocked. The user surplus function is given by:

$$U(x, y) = \gamma w_u - (\lambda - \gamma) w_b - x w_x - y(1 - P) w_y . \tag{EQ 1}$$

It can be verified that the surplus function is concave in x and a, and therefore exhibits a unique maximum.

After studying more carefully (EQ 1) we can see that there can exist 4 different types of users, depending on the values of w_x, w_y, w_u and w_b: 1) Users with $x>0$ and $a=1$ have enough traffic to allocate a VP and keep it into sufficient utilization levels, and in addition send their overflow traffic to the VC service; 2) Users with $x>0$ and $a=0$ allocate a VP but find the VC service too expensive; 3) Users with $x=0$ and $a=1$ do not have sufficient traffic to keep a VP of even one unit of capacity at reasonable utilization levels, but find the price of the VC service acceptable and use it exclusively for all their traffic; 4) Users with $x=0$ and $a=0$ find both services too expensive for their current service values and prefer to block all their traffic.

2.3 The Source-Destination Pair Optimization Problem

According to our model, the network advertises the following quantities to the users: 1) the price w_x of the VP service, 2) the price w_y of the VC service, and 3) the blocking probability P for the VC service (the latter can be based on measurements of the recent network history or an estimate based on the projected network load). The users compute their service demands by maximizing their utility functions (EQ 1) for each SD pair. Assuming that there are N users generating traffic for a given SD pair, let x_n and y_n denote the VP capacity allocation and offered load to the VC service for user n respectively. The total VP capacity demand and offered load to the VC service are given by:

$$x^{tot} = \sum_{n=1}^{N} x_n \qquad \text{and} \qquad y^{tot} = \sum_{n=1}^{N} y_n . \tag{EQ 2}$$

Also, let P^{max} denote the upper bound for the blocking probability of the VC service. This quantity is the QOS advertised by the network to its users. The network is said to guarantee QOS for the SD pair when P is below P^{max}. In addition, let us assume that the network manager imposes a quota x^{max} on the total capacity available for both the VP and VC services, for this SD pair. The SD pair is determined to be congested when $x^{tot} > x^{max}$ or $P > P^{max}$, where the blocking probability P is given by:

$$P = \begin{cases} E(y^{tot}, x^{max} - x^{tot}) , & \text{if} \quad x^{tot} \le x^{max} \\ 1 , & \text{otherwise} \end{cases} . \tag{EQ 3}$$

The role of x^{max} will become apparent in Section 3, where the network optimizes its revenue between all SD pairs. The network revenue for this SD pair is defined as follows:

$$R(w_x, w_y) = \begin{cases} x^{tot} w_x + y^{tot}(1 - P)(w_y - w_c) , & \text{if} \quad x^{tot} \le x^{max} \\ x_i^{max} w_x , & \text{otherwise} \end{cases} , \tag{EQ 4}$$

where w_c is the signaling cost incurred by every VC service request.

The network needs to find a price vector (w_x, w_y) that maximizes (EQ 4). The optimization problem for each SD pair can be stated as follows:

$$max\ R(w_x, w_y) \qquad \text{Subject to:} \qquad w_x \geq w_x^{min} , \quad w_y \geq w_y^{min} , \qquad \text{(EQ 5)}$$

where w_x^{min} and w_y^{min} are the minimum allowable prices that reflect the cost of the service. The optimal solution must also satisfy the QOS constraints $x^{tot} \leq x^{max}$ and $P \leq P^{max}$.

The revenue function R goes to 0 as w_x and w_y go to infinity. This can be easily verified since the user demands for the VP and VC services obtained by maximizing (EQ 1) are decreasing functions of the prices. Similarly, $R = 0$ when the prices are zero. Therefore, there exists a maximum revenue for some positive w_x and w_y.

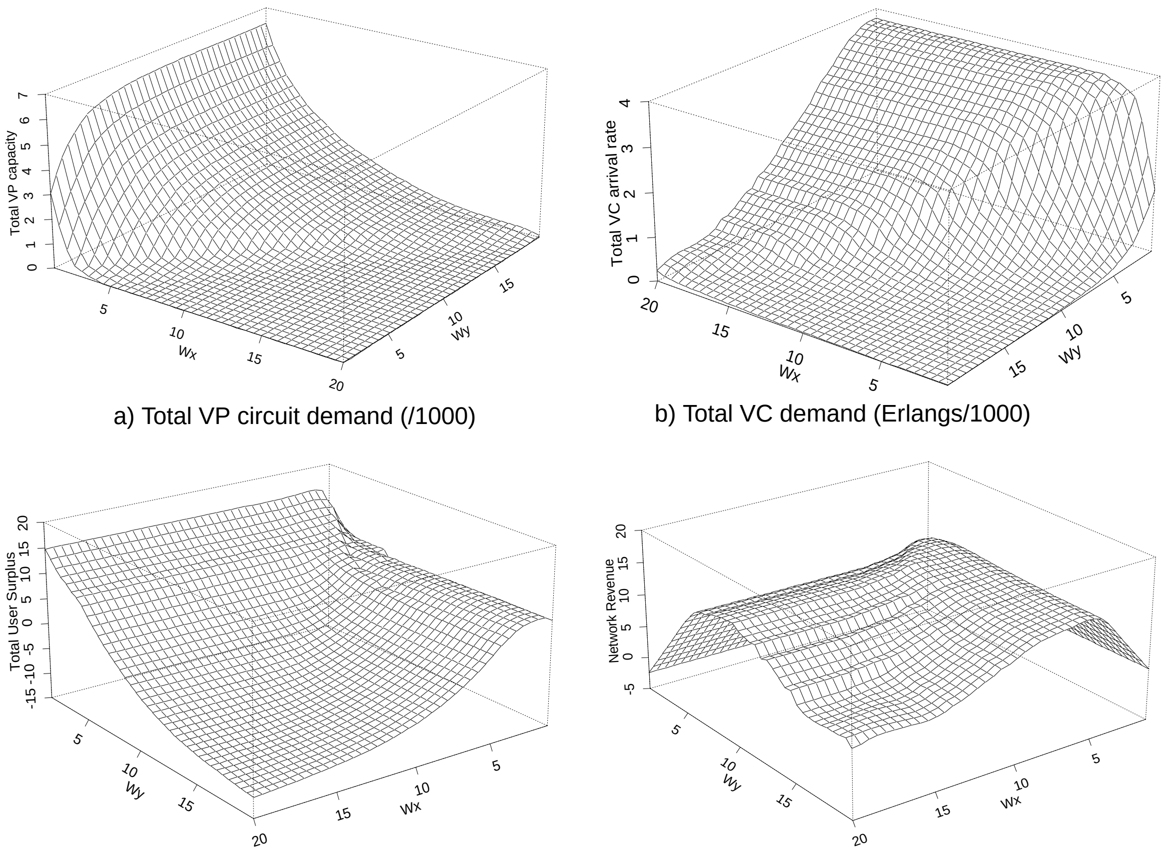
The above optimization problem is as a non-linear programming problem. The objective function is neither convex nor concave, and further, it cannot be transformed in a closed form as a function of the prices w_x and w_y because the user utility functions are not known to the network operator. The operator is only observing the service demand as a result of its pricing policy.

In order to better evaluate the problem, we have constructed an example with a population of 2000 users with diverse traffic requirements and service values. We model the arrival rate λ of every user in the population as a random variable. The first 1900 users have arrival rates exponentially distributed with mean $E\lambda = 1$ Erlang. The remaining 100 users have arrival rates also exponentially distributed with mean 20 Erlangs. In order to determine the w_u and w_b for each user, we have used the following arbitrary formulas:

$$w_u = w_y^{min} + q \cdot w_y^{min} \frac{\lambda}{E\lambda} , \qquad w_b = \frac{w_u}{2} . \qquad \text{(EQ 6)}$$

According to (EQ 6), high volume users are willing to pay more for the service. The constant q is 2 for the first population and 3.3 for the second. w_y^{min} was set at 1.2 and w_c to 2.0. In addition, let U^{tot} denote the total user surplus for the SD pair.

The plots for the total VP and VC demand, total surplus and revenue are shown in Figure 2. Notice that when the prices of the VP service are very low, user VP demand is higher than the total capacity available for the SD pair. For the minimum price of $w_x = 1.0$, the users require as much as 6313 circuits of VP service, which is more than twice the available capacity. Since we give priority to the VP service, the blocking probability for the VC service is 1 (Figure 3). As the VP service prices rise, the VP demand drops, and the VC blocking probability drops below 1 when the VP service demand reaches 3000 circuits. At the other extreme, when the VP prices are very high and VC prices very low, the demand for the VP service is zero and all user traffic requests the VC service (Figure 2b). The blocking probability in that case can be approximated by the Erlang formula for the total user load applied to a server system of 3000 servers. As VC prices increase, some users are discouraged from entering the system and as a result, the blocking probability is reduced. Figure 2c depicts the total user utility. Note that this is a strictly decreasing function of prices, except for a small region at the bottom right of the curve, where the VP service prices are low and the users allocate more VP service than the available capacity. The user utility is decreased in that case because users are allocated less than they have demanded. For any price vector that results in a total VP allocation less than the total capacity, the user utility function is strictly decreasing. Note that the higher user utility appears when prices are approximately at their minimum values. Therefore, a welfare maximizing policy would result in the lowest possible prices for the services such that network congestion would be avoided.



a) Total VP circuit demand (/1000)

b) Total VC demand (Erlangs/1000)

c) Total user surplus (/1000)

d) Network Revenue (/1000)

Figure 2: Plots of the total VP and VC demand, user utility and network revenue

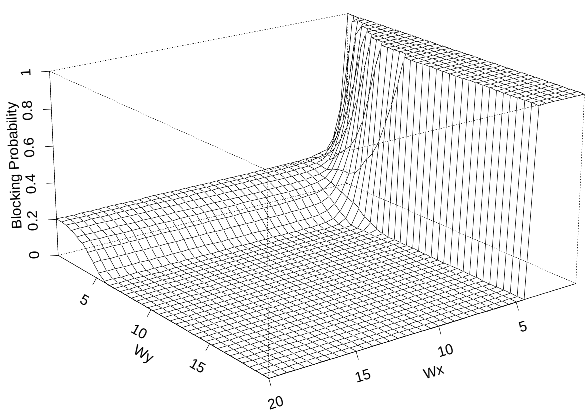


Figure 3: VC Blocking Probability

On the other hand, the network revenue plot appears to be strictly increasing for small prices w_x and w_y. The revenue is maximized for w_x = 6.0 and w_y = 8.2. It is worth noting that for these price values, blocking is nearly zero, i.e., the network is operating well below congestion levels. The total VP allocation is 1635 circuits and the total VC demand is 843 Erlangs. However, as prices increase further, the network revenue drops as more users are discouraged. Although the revenue function has a general concave shape, there exist many local maxima. The reason for this is that as the prices rise, a user may leave the system, which locally decreases the revenue. However, as the prices continue to rise, there is likely to be a price interval for which no users leave and as a result the revenue increases. When the prices are low, the increase in revenue appears to be almost linear, since all the users choose to allocate service.

2.4 Approaching the Optimal Prices

Since the user utility functions are not known to the network, we use an iterative procedure that converges to optimal prices. Intuitively, the shape of Figure 2d implies that by starting at the minimum, we can increase prices by a small amount and monitor the increase in network revenue. The rate at which revenue increases in either of the two directions can be used to provide the direction in which prices must be increased to approach the optimum. We have employed a

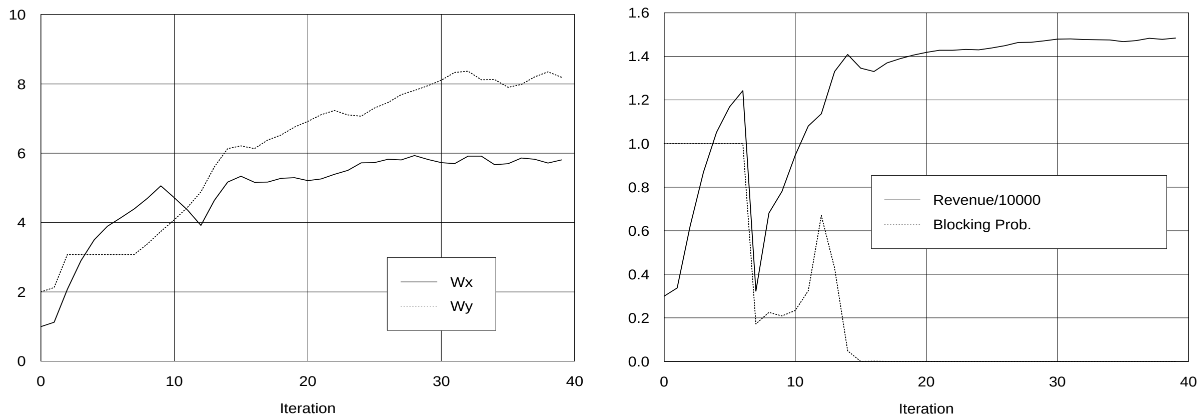


Figure 4: a) Prices vs. number of iterations b) Network revenue and VC blocking probability

price adjustment scheme based on the accelerated variation of the multidimensional exploratory search without derivatives [BAZ79]. The algorithm is described in detail in [ANE96b]. The price adjustment scheme is fundamentally a *Tatonnement* process [VAR78]. In our case, the excess demand function takes either the form of the excess VP demand or VC blocking probability. In addition, prices are further adjusted using a gradient method to maximize network revenue.

Figure 4 shows the behavior of the algorithm when applied to the user population of Section 2.3. In this figure, Plot a) shows the convergence of prices to the optimum values. Notice that convergence is attained in about 30 iterations. Plot b) shows the total network revenue and blocking probability. Notice also that the blocking probability is equal to 1 during the initialization of the algorithm, and remains at this level until the VP capacity demand has dropped below the total number of available circuits (which happens after iteration 7). The VC demand starts from 0 and increases as VP prices become prohibitively expensive. The user utility also reaches a maximum at iteration 7 when congestion ceases to exist, and drops subsequently as prices increase. During the first 7 iterations when the VP service demand cannot be entirely satisfied all users have portions of their VP demand rejected. Because the network does not distinguish based on each user's utility, the same demand percentage is rejected for each user. As a result, users with high values for the service observe the same blocking as others, and this causes the decreased utility value.

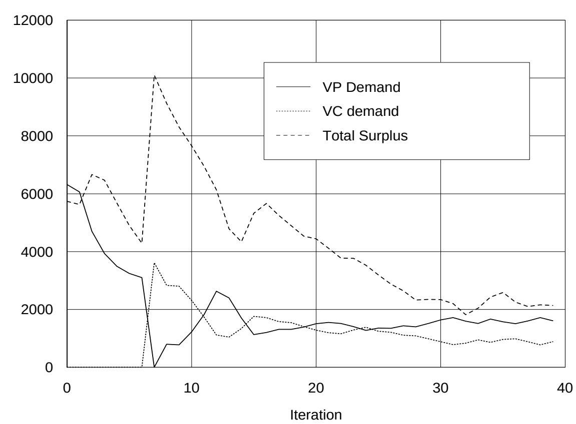


Figure 5: Total VP and VC demand and user surplus

2.5 The Effect of Load on Network Prices

We continue our analysis by focusing on an environment where the user traffic load changes over time. This behavior is typical in a commercial telecommunication system with both business and residential users. High volume business traffic is dominant during working hours, whereas traffic originating from residential users occupies the evening and night hours. The purpose of this section is to examine the behavior of the price equilibrium reached by the price adjustment scheme under load fluctuations. These fluctuations occur either because users change their behavior over time, or because new users enter or old users exit the system. We assume that the user value for services remains fixed and that the load changes occur on a much slower time scale than the one needed for the algorithm to converge for fixed user arrival rates. Our objective is to observe the behavior of prices vs. the user arrival rates.

Figure 6 shows the system parameters after the pricing algorithm has converged, parametrized by the mean arrival rate of the user population. In all cases, convergence was attained after approximately 30-40 iterations. Plot a) shows that the equilibrium prices increase as the user arrival rates increase, as one would expect. The price of the VC service is always higher than that

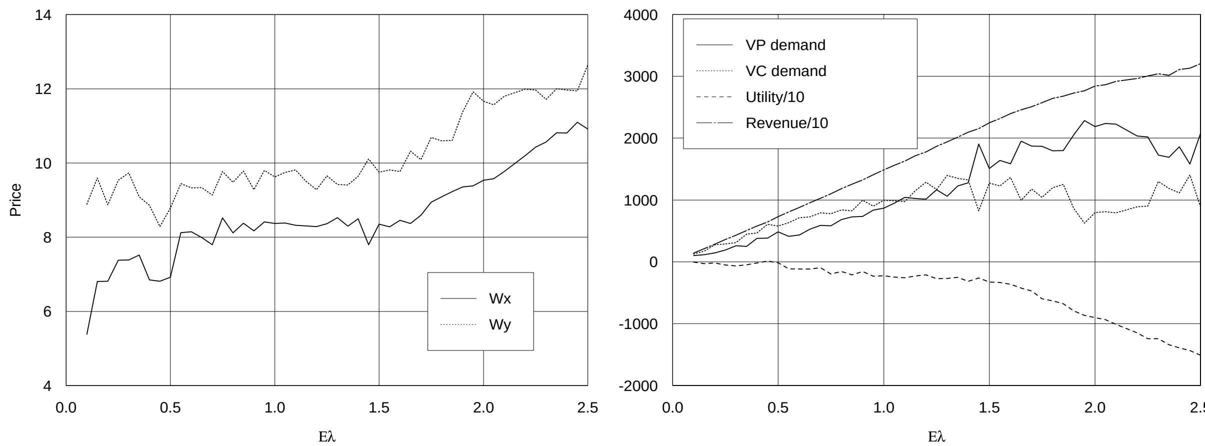


Figure 6: a) Equilibrium prices b) Demand, Utility and Revenue at equilibrium

of the VP service, and this reflects the additional signaling cost. The price curves are not strictly increasing because the algorithm stops after reaching one of the local revenue maxima. Since for every value of Eλ the algorithm is restarted, the prices are either slightly smaller or larger compared to the previous value of Eλ. This is an artifact introduced by the heuristic nature of the algorithm. By looking however at the general behavior, one can easily observe an increasing trend. The service demand is shown in Plot b). The plot reveals that as the load increases (and the prices increase as well) there is increased demand for the VP service. This is due to the fact that only the higher volume users (with higher service values) send traffic to the network. Since VP service prices are lower than the ones for the VC service, these users show preference for the VP service. On the other hand, as prices increase, low volume users that are the primary customers of the VC service are discouraged from entering the network, and as a result, the VC service demand drops. The network revenue appears to be a concave increasing function of the mean user traffic intensity. The total user utility however, increases slightly until $E\lambda = 1$ and then drops linearly, as prices increase and an increasing number of users are discouraged from the network. The general concave shape of both the prices and revenue curves shows that there is a maximum price beyond which no users enter the system. This price must be roughly equal to the highest w_u among the user population. At the limit, the total user arrival rate is so high, that the network wishes to discourage all users except the one with the highest service value. The arrival rate of this user is sufficient to use up all the available capacity.

2.6 The Effect of User Service Values

In this section, we focus on the effect of user service values on the price equilibrium. In this experiment, we hold constant the primary user arrival rate λ and multiply the initial user service value w_u by the same factor β for all users. Figure 7 demonstrates that the effect on service prices is almost linear. Decreased user values result also in decreased service prices in order to obtain the maximum revenue. As user values increase, service demand increases and the prices in-

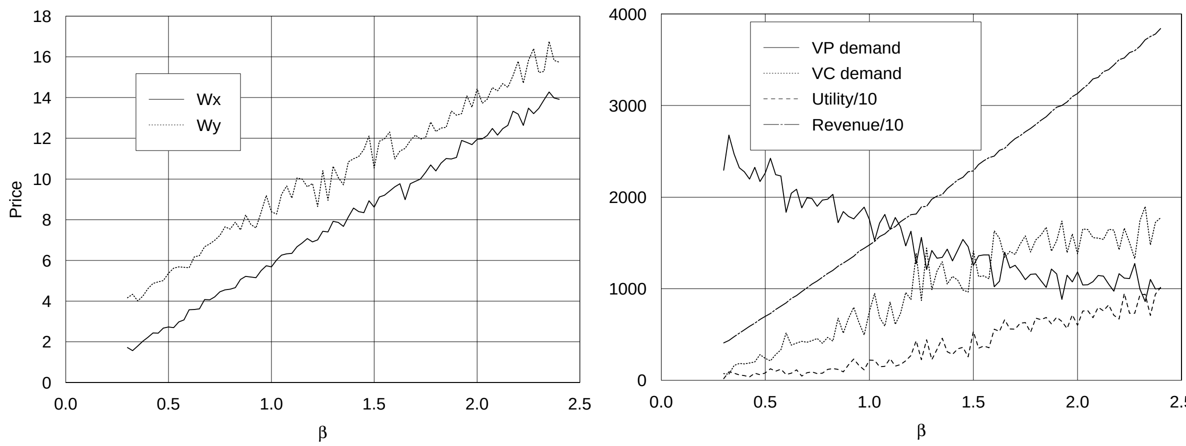


Figure 7: a) Equilibrium prices b) Demand, Utility and Revenue at equilibrium

crease as well in order to keep the network out of congestion. An interesting observation is that the VP service demand is decreasing with the user service values and the VC service demand is increasing. This is because when users have low service values, they will tend to allocate more of the lower priced service (the VP in this case). As their service value increases, the VC service becomes more economical to them. The network revenue appears to increase linearly with the user service value. The same holds for the total user utility.

2.7 Discussion

There are some practical considerations in the implementation of the algorithm. The most important is the number of iterations required for the algorithm to converge (around 30 from the initial prices in our experiments). One could imagine the network in an "auctioneer's" role: no trade is taking place until the equilibrium prices have been reached. This, however, is rather unrealistic for the commercial telecommunications market, where the user is not interested in a bidding process, but needs to know in advance the prices for the service in order to use it right away. There is, however, a partial answer to this problem: once convergence has been reached, minor deviations in the number of users, $E\lambda$ or w_u can be compensated in a few iterations. Assuming that the user behavior is changing slowly over time (and further, user behavior is approximately the same for the same time period from one day to another), the network can build a database for setting the initial prices according to the previously observed behavior, and subsequently make small adjustments to compensate for load or other fluctuations in the user population. However, although in this way faster convergence can be achieved, there still remains the question of determining the length of the time interval at which the network advertises its new prices and the users reply with their service demands. Clearly this interval must depend on the rate at which user behavior is changing.

Finally, a question can be raised here whether this iterative scheme is realistic for use in a future broadband network. A human user would be frustrated if it had to negotiate a price before being able to use the service. It is possible however that the user assigns an agent to participate in the price adjustment process. The user specifies its traffic demand, service values, and perhaps the maximum price it is willing to pay, and lets the agent negotiate the price on its behalf. The agent then decides if and how user traffic should be offered to the network. Assuming that communication overheads are of the order of milliseconds, convergence can be achieved in a few seconds, since all computation in this problem can be distributed.

3. NETWORK-WIDE SERVICE PRICING

The previous section provided a detailed analysis of the effect of network prices on user demand for one SD pair only. We now expand this analysis to a network-wide setting. In this setting, the network continues to price services by looking at the demand of individual SD pairs. The only interdependency that exists between the individual SD price optimization problems is that the VP capacity allocation or the demand for the VC service from one SD pair may affect the total available capacity for the VP service and the VC blocking probability of other pairs. In this section, we introduce another level of optimization (network level), where the network based on the current prices for every SD pair, decides on the maximum available capacity x^{max} available to every pair.

We further assume that the entire capacity x^{max} allocated to every SD pair follows the same path from the source to the destination POP. This means that all VPs and VCs between the two POPs will share the same physical route through the network. In reality, this assumption is not necessary and was omitted in the analysis of Section 2. However, it makes the network-wide pricing optimization problem more tractable.

3.1 Network-wide Pricing Optimization

After posting the new service prices, the network receives the user service demands for each SD pair and computes the total service demands x^{tot} and y^{tot}. Since this section looks at many SD pairs, let us denote the above quantities x_i^{tot} and y_i^{tot} respectively. According to our notation in this section, the index “i” refers to SD pair *i*. Assuming that the total available capacity to SD pair *i is* x_i^{max}, the SD pair revenue is given by (EQ 4) which is rewritten as follows:

$$R_i = \begin{cases} x_i^{tot} w_{x,i} + y_i^{tot}(1-P_i) w_{y,i}, & \text{if} \quad x_i^{tot} \le x_i^{max} \\ x_i^{max} w_{x,i}, & \text{otherwise} \end{cases} . \qquad \text{(EQ 7)}$$

The above equation reflects that the capacity which is available for the VC service is what remains from the total SD pair capacity after it has been distributed between the VP requests. That is, VP capacity allocation is given priority over the VC service, and in the case that $x_i^{tot} \ge x_i^{max}$, the blocking probability P_i will be 1.

We further assume that the capacity of every link is given by C_l. The objective of the network is to find the vector of capacity allocations $\boldsymbol{x^{max}}$ such that the total network revenue is maximized. The optimization problem can be written as:

$$max\ R = \sum_i R_i(x_i^{max}) \qquad \text{subject to:} \qquad \sum_i x_i^{max} \delta_{il} \le C_l, \qquad \forall l, \qquad \text{(EQ 8)}$$

where δ_{il} is 1 if SD pair *i* is routed through link *l* and 0 otherwise. Since this is a non-linear programming problem with a non-differentiable objective function, we can use the optimization methodology of [ANE96a]. When the final solution has been reached, SD pairs will have been sized based on their “profitability”.

The framework for the network pricing optimization problem is shown in Figure 8. Optimization is done at three levels: the user level, where users optimize their demands with respect to the current advertised prices, the SD pair level, where the network decides on a new price vector for the SD pair in order to satisfy the QOS constraints and maximize the local revenue, and the network level, where capacity is allocated between SD pairs according to their profitability.

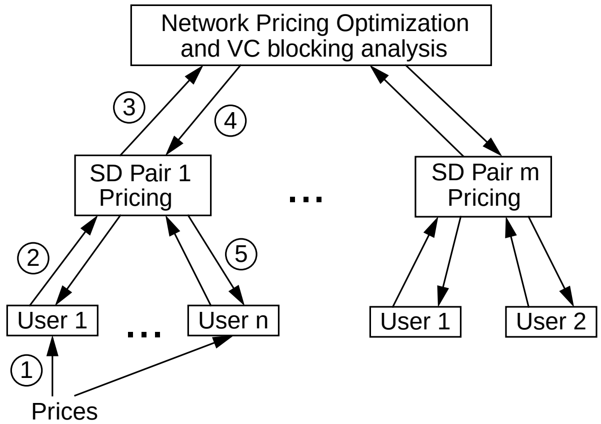


Figure 8: The Network Pricing Optimization

The network-wide pricing optimization proceeds as follows: 1) Users read the current service prices and optimize their service demands for each SD pair. 2) Users transmit their demands on a per SD pair basis to their POP. The POP maintains a computing facility that runs a number of SD pair optimization problems, one for each SD pair for which services are requested. 3) Each SD pair pricing module computes the total service demand x^{tot} and y^{tot}. This demand is transmitted to the central network analysis facility together with the current prices for the SD pair. The latter receives the total service demands for all SD pairs in the network, and decides on the maximum capacity x^{max} that can be allocated for each SD pair. 4) x^{max} is transmitted back to the SD pair pricing modules. The latter can now decide on the new service prices using the analysis of Section 2.4. 5) The new prices are transmitted back to the users, and the cycle starts from the beginning.

The network-wide pricing optimization requires a combination of distributed and centralized processing. Each user optimizes his service demand for each SD pair. Every POP calculates the new prices for all originating SD pairs. Finally, a centralized management facility decides on the optimal capacity allocation between the SD pairs.

3.2 Many Traffic Classes

The analysis of the previous section can be easily extended to more than one traffic classes. The user can have different values of the w_u and w_b for each class. Similarly, the network provides a different blocking estimate for each traffic class, and also charges a different rate. In addition, a different value of x^{max} is allocated from every SD pair to each traffic class.

4. SUMMARY

We examined how a pricing framework can be used to control the user demand for services in an ATM network. Knowing that the user demand is a decreasing function of the prices, the network manager controls the prices at regular time intervals to guarantee the advertised call-level QOS. Our environment considers the availability of two connection services (VC and VP) with a retail/wholesale relationship; the VP service can priced less per unit of capacity since every VC connection has a signaling cost. Users select any combination of the two services based on their offered traffic and utility function. We investigated how pricing of the two services influences user demand and developed a price adjustment scheme that can be used by the network manager for every Source-Destination pair in the network in order to satisfy two objectives: 1) guarantee the advertised call blocking constraints and 2) maximize the network revenue. Compared with similar schemes, our approach does not maximize a user welfare function but rather tries to drive the user demand to a point that is mostly beneficial to the service provider. This point has the characteristics of a Nash equilibrium: users do not wish to alter their demand as this would result in a drop of their utility; the network would also loose revenue by changing the prices. We studied the effect on the equilibrium when the user load or user value for the services change. A third level of optimization was introduced, where the network optimizes the total revenue by distributing the available capacity between SD pairs according to their profitability.

REFERENCES

[ANE96a] N. Anerousis and A.A. Lazar, "Virtual Path Control for ATM Networks with Call-Level Quality of Service Guarantees", *Proceedings of the 1996 INFOCOM*, San Francisco, CA, March 1996.

[ANE96b] N. Anerousis, "Managing Virtual Circuits and Virtual Paths on ATM Networks with Quality of Service Guarantees", Ph.D. dissertation, Columbia University, New York, NY, 1996.

[ANE95] N. Anerousis, and A.A. Lazar, "Managing Virtual Paths on Xunet III: Architecture, Experimental Platform and Performance", *Proceedings of the 1995 IFIP/IEEE International Symposium on Integrated Network Management,* Santa Barbara, CA, May 1-5, 1995.

[BAZ79] Bazaraa, M., and Shetty, C.M., "Nonlinear Programming, Theory and Algorithms", John Wiley and Sons, Inc., 1979.

[COC93] Cocchi, R., Shenker, S., Estrin, D., and Zhang, L., "Pricing in Computer Networks: Motivation, Formulation and Example", in *IEEE/ACM Transactions on Networking,* Vol. 1, No. 6, December 1993.

[HON95] Honig, M.L., Steiglitz, S., "Usage Based Pricing of Packet Data Generated by a Heterogeneous User Population", in *Proceedings of the 1995 INFOCOM*, Boston, MA, April 1995.

[JIA95] H. Jiang and S. Jordan, "The Role of Price in the Connection Establishment Process", *European Transactions on Telecommunications*, vol. 6, no. 4, July-August 1995, pp. 421-429.

[JOR95] S. Jordan & H. Jiang, "Connection Establishment in High Speed Networks", *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 7, September 1995, pp. 1150-1161.

[KEL94] F.P. Kelly, "Tariffs and Effective Bandwidths in Multiservice Networks", in *Proceedings of the 14th International Teletraffic Congress*, edited by J. Labetoulle and J.W. Roberts, Elsevier Science, 1994.

[LOW93] Low, S. and Varaiya, P.P., "A New Approach to Service provisioning in ATM Networks", in *IEEE/ACM Transactions on Networking*, Vol. 1 No. 5, pp. 547-553, October 1993.

[MAC94a] MacKie-Mason, J.K., and Varian, H.R., "Pricing Congestible Network Resources", Technical Report, University of Michigan, Ann Arbor, MI 48109.

[MAC94b] J.K. Mackie-Mason and H.R. Varian, "Pricing the Internet", *Proceedings of the Second International Conference on Telecommunication Systems Modeling and Analysis*, pp. 378-393, Nashville, Tennessee, March 24-27, 1994.

[MUR94] Murphy, J. Murphy, L. and Posner, E.C., "Distributed Pricing for Embedded ATM Networks", in *Proceedings of the 14th International Teletraffic Congress*, edited by J. Labetoulle and J.W. Roberts, Elsevier Science, 1994.

[PAR92] C. Parris, S. Keshav and D. Ferrari, "A Framework for the Study of Pricing in Integrated Networks", ICSI Technical Report TR-92-016 and AT&T Bell Labs Technical Memorandum TM-920105-03, January 1992.

[SAR95] J. Saraimesh, D.F. Ferguson and Y. Yemini, "An Approach to Pricing, Optimal Allocation and Quality of Service Provisioning in High-Speed Packet Networks", in *Proceedings of the 1995 INFOCOM*, Boston, MA, April 1995.

[VAR78] Varian, H.R., "Microeconomic Analysis", W.W. Norton & Company Inc., 1978.