Equilibrium Pricing in Multi-service Priority-based Networks *

Luiz A. DaSilva †, *David W. Petr* † *and Nail Akar* ‡

† *Information and Telecommunication Technology Center*
Department of Electrical Engineering & Computer Science
The University of Kansas
Lawrence, KS 66045

‡ *Sprint Corporation*
Overland Park, KS 66212

Abstract

The pricing of network services not only determines the economic viability of commercial networks but also plays an important role in traffic management through its influence on user behavior. This paper discusses the pricing of services in a priority-based network by employing game-theoretic concepts. Given any price difference between services and an estimate of users' utility functions, we describe a method for predicting what users' service choices will be. In this fashion, service providers can determine what price ranges better encourage users to exhibit behavior that is mutually beneficial to users and providers. Under certain strict assumptions for the traffic characteristics and utility functions, necessary and sufficient conditions for the achievement of an optimal equilibrium are presented.

1: Introduction

While a great deal of research has been carried out in the past few years on the subject of traffic management and traffic models for wide-area communication networks, the issue of pricing remains relatively unexplored.

The pricing structure not only affects revenue, but directly influences the load presented to the network and hence its performance (for instance, customers may choose to postpone offered traffic to times when lower tariffs are in effect). Therefore, the pricing policy plays an important role in the dimensioning of a commercial network. Several issues in traffic management, such as congestion control and Connection Admission Control (CAC), may also be affected by pricing.

This paper focuses on the study of desirable pricing ranges for priority classes, taking into account users' decision-making process and economic efficiency requirements (maximizing network provider and customer benefits). Customers' decisions consist of determining, at any given time, which of the services available optimizes their individual cost/benefit relationship. Network providers, on the other hand, must take into account the satisfaction of all users, as well as fairness, network performance and revenue when deciding what pricing structure to implement.

While at times users and providers may have conflicting objectives, we believe that pricing can be used as a way to encourage users to exhibit behavior that is beneficial to the network as a whole.

The goal of multi-service networks is to simultaneously meet the diverse quality of service (QoS) requirements of a variety of applications with acceptable user costs. Game theoretic concepts are utilized here in order to model the interaction among users and predict users' choices among service classes. This prediction enables us to determine whether the network is accomplishing the stated goal.

This paper is structured as follows: in Section 2, we propose a general pricing structure; the model used to describe the traffic contract and users' decisions is presented in Section 3; Sections 4 and 5 list some general results and specific examples, respectively, regarding pricing ranges for priority-based networks; finally, Section 6 summarizes our main conclusions. The Appendix contains the proofs of the propositions in Section 4.

2: Pricing Approach

The pricing of services in circuit-switched networks, and in particular telephone rate structures, is well understood [9]. However, pricing packet-switched networks presents additional challenges. Since these networks rely on statistical multiplexing, the reservation of resources is ideally kept to a minimum, and measuring usage is more difficult and costly.

The current debates on how to charge for Internet services [1, 8, 10] and ATM services [11, 12, 15] have sparked renewed interest in the field of network pricing.

We contend that a simple but relatively general pricing policy would assign a price P to service type j according to three main factors: resource allocation, usage and fixed connection costs. This relationship can be expressed as:

$$P_j = c_j + \int_{t_o}^{t_f} (f_j(t, R(t)) + g_j(t, U(t)))dt \qquad (1)$$

In the expression above, $R(t)$ refers to the amount of resources allocated to the call and $U(t)$ refers to usage of these

*This research is supported by Sprint Corporation.

resources (*e.g.*, number of packets delivered), both as functions of time. The integral is taken over the interval $[t_o, t_f]$, the duration of the call. The constant c_j refers to the fixed connection charge for service j. The functions f_j and g_j will determine the exact pricing structure for service type j; for instance, in service types with a strong resource allocation component (*e.g.* Constant Bit Rate service in ATM networks) one can expect f_j to dominate over g_j. The explicit dependency of f_j, g_j on t enables the implementation of time-of-day pricing, where rates are higher when the network is expected to be more heavily used.

The only restriction we place on f_j and g_j is that $f_j(t, 0) = g_j(t, 0) = 0 \; \forall \; t$. However, we do not expect f_j to be a linear function; this is best explained through an example. Suppose a user intends to transmit a 1 MB file over the network, having the choice of reserving 1 Mb/s or 500 kb/s of bandwidth. If $f_j(t, R(t)) \equiv aR(t)$, the price charged for the file transfer will be the same for either case, and the user will choose the first option. However, network providers generally prefer to spread resource allocation over a longer period of time; besides, the user will likely be willing to pay more for the first option, since it affords better quality of service. The combination of these two factors leads us to believe that f_j should be a non-linear function of R(t).

In the remainder of this paper, we discuss priority-based networks with no resource allocation ($R(t) \equiv 0$), so $f_j \equiv 0$.

Our study concentrates on relative prices, since they are the primary factor in determining customers' choices among service alternatives. Furthermore, the cost-recovery aspect of pricing, which is directly related to absolute prices, is not the subject of this work.

3: System Model

This work employs game theory to model customers' decisions when utilizing commercial networks. Game theory has been used for years as a tool of economic analysis, to understand and predict what will happen in economic contexts [7]. Its applications to networking problems include not only pricing but congestion control and CAC as well [3, 4, 14].

The use of game theory in the study of network service pricing allows us to model how a user's choice of service is impacted by beliefs about other customers' decisions. In the present work, we extend the pricing policy model described in [3], applied in the context of multi-service priority-based networks. In this scenario, each customer must choose the most appropriate service for her application taking into account the performance requirements of the application, prices of services, and interactions with other users. This approach is very similar to the one presented in [2, 3]; however, these papers concentrate on simulation of priority networks, whereas we build an analytical framework for the problem.

Let S_i be the set of choices available to user i when requesting service from the network. The joint strategy space, denoted by S, is the Cartesian product of the individual strategy sets; for N users, we have:

$$S = S_1 \times S_2 \times \ldots \times S_N = \{\mathbf{s} = (s_1, s_2, \ldots, s_N) : s_i \in S_i\} \quad (2)$$

We will denote the traffic statistics for customer i by $\mathbf{t_i}$; this may include information such as average transmission rate and statistics of message size. When the service request from user i, $(s_i, \mathbf{t_i})$, is accepted, the user receives some level of service from the network, characterized by $\mathbf{q_i}$, where

$$\mathbf{q_i} = \Psi(\mathbf{s}, \mathbf{t_1}, \ldots, \mathbf{t_N}) \quad (3)$$

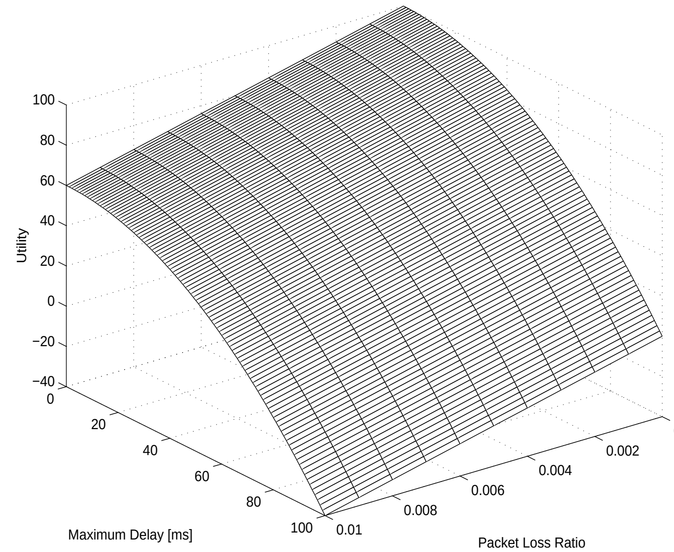

Figure 1. Example of utility function for a voice application.

The precise characterization of the function Ψ depends on the network topology and service disciplines; specific examples are provided in Sections 4 and 5.

3.1: Utility Functions

Utility functions serve to quantify the tradeoffs that customers are willing to make between the quality of a service received and its price. We consider that individual users associate a value to each service level; this value, referred to as the user's *utility function* $U_i(\mathbf{q_i})$, can be interpreted as the amount the user is willing to pay for a given QoS.

Utility functions are widely used in pricing theory; some common assumptions about U_i include concavity and strict monotonicity. It is seldom possible in practice to determine users' exact utility functions; rather, we postulate a specific function based on the known characteristics of the application. For instance, voice applications are typically very sensitive to maximum delay and delay variation, and yet can be made relatively insensitive to packet losses. It is reasonable, then, to represent the value a customer associates to a voice application as a function of QoS parameters such as maximum delay and packet loss ratio as shown in Figure 1.

3.2: Evaluating Pricing Policies

The *consumer's surplus* $C_i^{(\mathbf{s})}$ is then defined as the difference between the utility obtained with a given service choice and the price paid for the service. Users will decide on a service request (*i.e.*, a pure strategy) that maximizes their surplus given the other players' strategies. The effectiveness of a pricing policy can be evaluated based on the existence of an equilibrium joint strategy. The idea is that we should be able to predict users' choices in order to ensure some level of satisfaction to users and revenue to providers; the equilibrium strategy combination, when one exists, is considered a consistent prediction [5].

A Nash equilibrium is a strategy combination where no user can unilaterally increase her utility by changing her strategy [3, 7]. More precisely, we offer the following definition:

Definition 1 (Nash Equilibrium) *Strategy combination* $\mathbf{s}$ *is a Nash equilibrium if* $C_i^{(\mathbf{s})} \geq C_i^{(s_i^*, \mathbf{s_{-i}})} \; \forall \; s_i^* \in S_i, i \in$

$\{1, 2, ..., N\}$. [1]

If an equilibrium exists, we would like to determine whether it is efficient; for this purpose we use the concept of Pareto optimality. A strategy combination is Pareto optimal if there is no other strategy combination which at least one user would prefer and to which all others would be indifferent, or, more formally [14]:

Definition 2 (Pareto Optimality) *A strategy combination* $\mathbf{s}$ *is Pareto optimal if there does not exist* $\mathbf{s}' \in S$ *such that:*

1. $C_i^{(\mathbf{s}')} \geq C_i^{(\mathbf{s})}$ $\forall$ *i; and*
2. $C_i^{(\mathbf{s}')} > C_i^{(\mathbf{s})}$ *for at least one i.*

4: Non-preemptive Priority System

In order to analyze the effect of pricing on the equilibrium achieved, we consider the case of a single FIFO queue with non-preemptive priorities ("high" and "low" priority levels). The study of a single queue is applicable to local area networks (LANs) and metropolitan area networks (MANs), which are sometimes modeled as a single server with a queue that is distributed among all stations [13]. Besides, this relatively simple system is analytically tractable, whereas for the analysis of a general network one must rely on simulation and/or experiments.

Let N be the number of customers utilizing the queue at a given time. Each user can choose to tag a percentage s_i of her traffic as high priority, paying a price p_H for the bandwidth utilized; the remainder of the traffic is transmitted as low-priority at a price p_L. This in effect yields a joint strategy space to the game $S = [0, 1]^N$.

If surplus functions C_i are differentiable over S and strictly monotonic functions of the QoS parameter, we can state that:

Proposition 1 1. *A pure strategy Nash equilibrium exists in the interior of the joint strategy set* ($S^o \equiv (0, 1)^N$) *if and only if it is a solution for the system of equations:*

$$\frac{\partial C_i}{\partial s_i} = 0 \tag{4}$$

2. *If a pure strategy Nash equilibrium* $\hat{s}$ *exists in* $S \backslash S^o$*, then* $\hat{s} \in \{0, 1\}^N$ *(i.e.,* $\hat{s}$ *is an extreme point).*

Furthermore, if $\frac{\partial C_i}{\partial s_i} > 0$ over S, then $\hat{s_i} = 1$; conversely, $\frac{\partial C_i}{\partial s_i} < 0$ would imply $\hat{s_i} = 0$.

Therefore, the procedure for determining the Nash equilibrium reduces (in the worst case) to solving a system of N equations plus a search of the 2^N extreme points of S.

An important quantity in determining the equilibrium (and its optimality) is $\Delta p \equiv p_H - p_L$. Intuitively, one can expect that if Δp is made large enough, all users will tend to choose low priority service (or simply exit the system). On the other hand, if Δp is sufficiently low, all users will tend to choose high priority service, a clearly poor result from the standpoint of Pareto optimality.

For the sake of concreteness, quality of service in this system will be measured by average waiting time in the queue, denoted by W_i. Utility functions will be approximated by:

$$C_i^{\mathbf{s}} = \underbrace{A_i - B_i\ (W_i)^{d_i}}_{U_i} - [p_H s_i \lambda_i - p_L(1 - s_i)\lambda_i] \tag{5}$$

Differences in sensitivity to QoS changes between users can be modeled by varying the parameters A_i, B_i and d_i. Furthermore, A_i plays no role in the determination of an equilibrium [2]. Parameters A_i, B_i and d_i are not arbitrary; in the determination of a pricing policy, service providers must estimate (usually based on empirical evidence) how sensitive customers are to changes in prices and QoS.

Due to the scarcity of closed-form results for delay in G/G/1 priority queueing systems, we assume Poisson arrivals to the queue, with $\mathbf{t_i} = (\lambda_\mathbf{i}, \overline{\mathbf{x_i}}, \overline{\mathbf{x_i^2}})$, where λ_i is the average arrival rate, and $\overline{x_i}$ and $\overline{x_i^2}$ are the first two moments of message length for customer i.

Let us for a moment restrict the strategy space to the extreme points of S. In the simple case of 2 users with independent identically distributed arrivals to the queue, exponentially-distributed message lengths of mean μ and constant marginal utility B_i ($d_i = 1$), we can determine sufficient and necessary conditions on Δp for the existence of an optimal equilibrium, in closed-form:

Proposition 2 *Under the assumptions listed above, a two-user system achieves a unique Nash equilibrium that is Pareto optimal and maximizes revenue if and only if:*

$$(\min_i B_i)\frac{2\lambda/\mu}{(\mu - 2\lambda)(\mu - \lambda)} < \Delta p < (\max_i B_i)\frac{2\lambda/\mu}{(\mu - 2\lambda)(\mu - \lambda)} \tag{6}$$

In more complex cases, for any given Δp, the Nash equilibrium can be calculated (and its Pareto optimality evaluated) using the procedure delineated above. We provide two examples in the next section.

5: Examples of Nash Equilibria in Priority Systems

Utilizing the general form for the utility function described in equation 5 and well-know queueing theory results [6], and assuming $\overline{x_i} = \overline{x}$, $\overline{x_i^2} = \overline{x^2}$ $\forall i$, we get:

$$\begin{aligned} \frac{\partial C_i}{\partial s_i} &= B_i K^{d_i} d_i \overline{x} \frac{(1-\overline{x}\lambda_T s_i)^{d_i - 1}}{(1-\overline{x}\sum_{j=1}^{N} s_j \lambda_j)^{d_i+1}} [\textstyle\sum_{j \neq i} \lambda_j \\ &\quad - \overline{x}\lambda_T \textstyle\sum_{j \neq i} s_j \lambda_j] - \Delta p \lambda_i \end{aligned} \tag{7}$$

where $\lambda_T = \sum_{j=1}^{N} \lambda_i$ and

$$K = \frac{\overline{x^2}\lambda_T}{2(1 - \overline{x}\lambda_T)} \tag{8}$$

To illustrate the search for an equilibrium in the extreme points of S, we investigate a two-user case with non-linear utilities (represented in Figure 2); the results confirm the intuitive interpretations of Proposition 2. Typical consumers' surplus combinations as we vary Δp are shown in Table 1; the equilibrium $\hat{s}$ is found to be:

[1] We use the following convention: $\mathbf{s_{-i}}$ denotes all components of $\mathbf{s}$ except its i^{th} component.

[2] In competitive pricing, it can be assumed that customers are no longer willing to pay for the service if the utility falls below a certain threshold, and that is where A_i plays an important part.

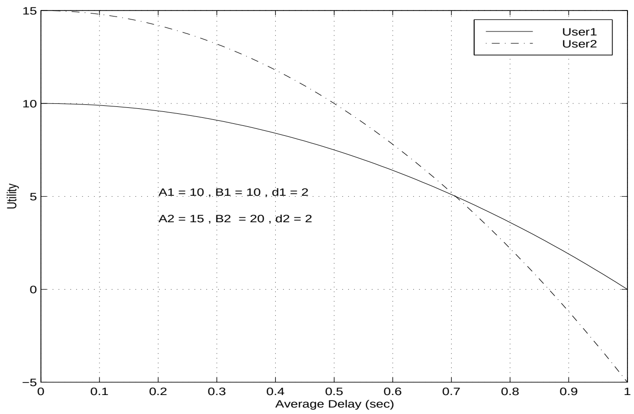

Figure 2. User characteristics for example 1.

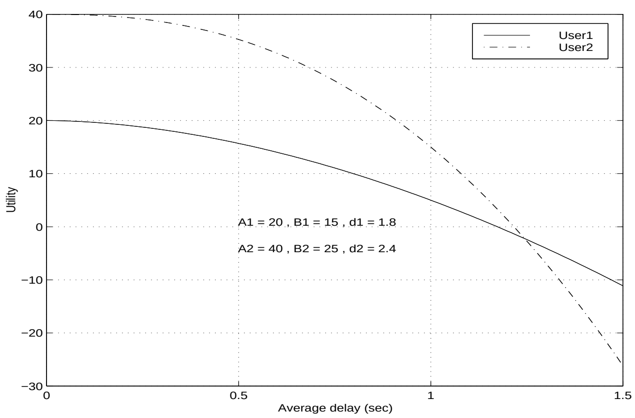

Figure 3. User characteristics for example 2.

$$\hat{s} = \begin{cases} (1,1) & \text{if } \Delta p \leq 0.9 \\ (0,1) & \text{if } 0.9 < \Delta p \leq 1.3 \\ (0,0) & \text{if } \Delta p > 1.3 \end{cases} \tag{9}$$

These results suggest the existence of an optimum price range (namely, $0.9 < \Delta p \leq 1.3$) that takes advantage of the structure of the network to benefit all users according to their sensitivity to QoS. In this case, no equilibrium exists in S^o.

In order to illustrate the existence of an equilibrium in S^o, we present an example consisting of two users whose applications are characterized in Figure 3. The partial derivatives of C_i with respect to s_i are plotted in Figure 4; the intersection of the two contours is a strictly monotonic curve, yielding a unique equilibrium in S^o (where this curve crosses the zero plane).

6: Conclusions and Future Work

We have outlined a method for finding the equilibrium in a priority system given any price difference between services and an estimate of users' utility functions. In this fashion, a network provider can determine the pricing range that will yield a desirable equilibrium. The results presented here confirm some of the conclusions of [2], namely that by providing monetary incentives one can tailor the services provided to customers' needs.

Integrated services networks in the near future are expected to offer service classes based on a mix of priority and resource allocation. For instance, in ATM networks,

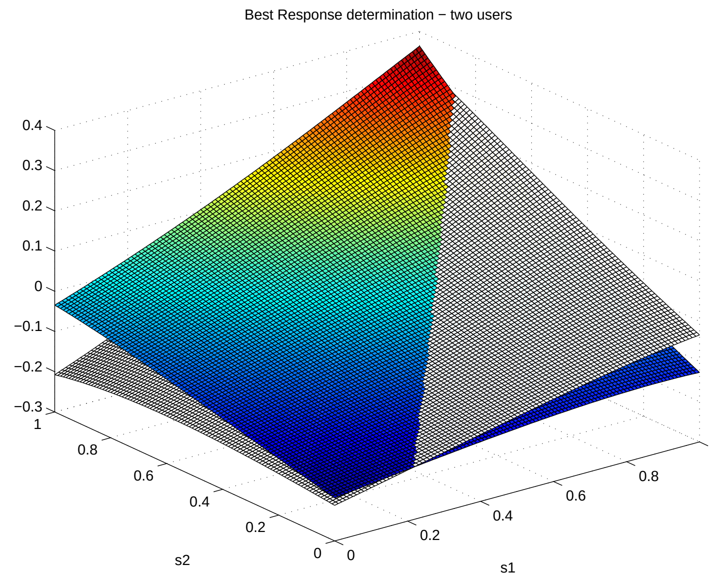

Figure 4. Determination of an equilibrium in S^o.

some service classes differ primarily in the amount of bandwidth reserved, and others, where virtually no reservation is necessary, differ primarily in the assignment of priorities. In order to extend the results of this paper to such networks, the authors are currently conducting a simulation-based investigation into the problem of equilibrium pricing for networks with allocation of resources.

A Proofs

In this appendix, we provide the proofs of the propositions in Section 4.

Proof 1 *1. A Nash equilibrium $\hat{s} = (\hat{s_1}, \hat{s_2})$ implies that $\hat{s_1}$ is the best response to $\hat{s_2}$ and vice-versa [5]. Therefore, if $\hat{s}$ belongs to an open set it must maximize C_i with respect to s_{-i}. The concavity of the utility functions makes the first order necessary conditions also sufficient.*

2. The assumption of strictly monotonic utility functions forces the equilibria (if they exist) to reside in one of the extreme points.

Proof 2 *In this system, each player has two alternatives: high (H) and low (L) priority services. Let us denote by $C_k^{(i,j)}$ the payoff (consumer's surplus) derived by user k when user 1 chooses type-i service and user 2 chooses type-j. The ordered pair (i,j) is the pure strategy combination $\mathbf{s}$.*

The average waiting time for an M/M/1 queue with non-preemptive priority is a well-known result in queueing theory (see, for instance, [6]). For each pure strategy combination $\mathbf{s} \in \{(i,j) : i,j \in \{H,L\}\}$ we can find the expected payoff for each user $C_k^{(\mathbf{s})}, k \in \{1,2\}$. The payoffs are shown in Tables 2 and 3.

Let us now consider the following cases:

1. *Let $\Delta p < \min(B_i) \frac{2\lambda/\mu}{(\mu-\lambda)(\mu-2\lambda)}$.*

   In this case, $C_1^{(H,H)} > C_1^{(L,H)}$ and $C_2^{(H,H)} > C_2^{(H,L)}$, and therefore, (H,H) is a Nash equilibrium. This is the unique Nash equilibrium, since (H,L) is not an equilibrium ($C_2^{(H,L)} < C_2^{(H,H)}$), and neither are (L,H)

		$\Delta p = 0.7$		$\Delta p = 1.1$		$\Delta p = 1.4$	
		s_2		s_2		s_2	
		0	1	0	1	0	1
s_1	0	8.13 , 12.50	7.64 , 12.89	8.13 , 12.50	7.64, 12.64	8.12, 12.50	7.64, 12.49
	1	8.17 , 11.53	7.83 , 12.20	7.92 , 11.53	7.58 , 11.95	7.77 , 11.53	7.43 , 11.80

Table 1. Variation in consumers' surplus as Δp changes. Surplus is presented as ordered pairs C_1^s, C_2^s.

		User 2	
		H	*L*
User 1	*H*	$A_1 - \frac{2B_1\lambda/\mu}{\mu-2\lambda} - p_H\lambda$	$A_1 - \frac{2B_1\lambda/\mu}{\mu-\lambda} - p_H\lambda$
	L	$A_1 - \frac{2B_1\lambda}{(\mu-\lambda)(\mu-2\lambda)} - p_L\lambda$	$A_1 - \frac{2B_1\lambda/\mu}{\mu-2\lambda} - p_L\lambda$

Table 2. Expected payoff function $C_1^{(s)}$ for user 1.

		User 2	
		H	*L*
User 1	*H*	$A_2 - \frac{2B_2\lambda/\mu}{\mu-2\lambda} - p_H\lambda$	$A_2 - \frac{2B_2\lambda}{(\mu-\lambda)(\mu-2\lambda)} - p_L\lambda$
	L	$A_2 - \frac{2B_2\lambda/\mu}{\mu-\lambda} - p_H\lambda$	$A_2 - \frac{2B_2\lambda/\mu}{\mu-2\lambda} - p_L\lambda$

Table 3. Expected payoff function $C_2^{(s)}$ for user 2.

($C_1^{(L,H)} < C_1^{(H,H)}$) or (L,L) ($C_1^{(L,L)} < C_1^{(H,L)}$). However, this unique equilibrium is clearly not Pareto optimal, since $C_i^{(H,H)} < C_i^{(L,L)}$, $i = 1, 2$.

2. *Let $\Delta p > \max(B_i)\frac{2\lambda/\mu}{(\mu-\lambda)(\mu-2\lambda)}$.*

   Now, $C_1^{(L,L)} > C_1^{(H,L)}$ and $C_2^{(L,L)} > C_2^{(L,H)}$, making (L,L) a Nash equilibrium. This is the unique equilibrium point, since (H,H) is not an equilibrium ($C_1^{(H,H)} < C_1^{(L,H)}$) and neither are (H,L) ($C_1^{(H,L)} < C_1^{(L,L)}$) or (L,H) ($C_2^{(L,H)} < C_2^{(L,L)}$).

3. *Let $\min(B_i)\frac{2\lambda/\mu}{(\mu-2\lambda)(\mu-\lambda)} < \Delta p < \max(B_i)\frac{2\lambda/\mu}{(\mu-2\lambda)(\mu-\lambda)}$.*

   Suppose $B_1 > B_2$. In this case, $C_1^{(H,L)} > C_1^{(L,L)}$ and $C_2^{(H,L)} > C_2^{(H,H)}$; therefore, (H,L) is a Nash equilibrium. Again, it is unique; by the observation above, neither (L,L) nor (H,H) is an equilibrium, and one can easily verify that $C_1^{(L,H)} < C_1^{(H,H)}$. By symmetry, analogous results are obtained when $B_1 < B_2$.

4. *For the sake of completeness, we must consider the cases of equality: $\Delta p = B_i\frac{2\lambda/\mu}{(\mu-\lambda)(\mu-2\lambda)}, i = 1, 2$.*

   In either case, there is no longer a unique equilibrium, since more than one strategy will be equivalent in the eyes of one of the users.

The equilibria obtained in the cases 2 and 3 are both Pareto optimal. It is easy to see that revenue is maximized in case 3 (since in case 2, both users choose low priority service). •

References

[1] Herb Brody. Internet@crossroads.$$$. *Technology Review*, May/June 1995.

[2] Ron Cocchi, Deborah Estrin, Scott Shenker, and Lixia Zhang. A Study of Priority Pricing in Multiple Service Class Networks. In *SIGCOMM Symposium on Communications Architectures and Protocols*, pages 123–130, Zurich, Switzerland, September 1991.

[3] Ron Cocchi, Scott Shenker, Deborah Estrin, and Lixia Zhang. Pricing in Computer Networks: Motivation, Formulation, and Example. *IEEE/ACM Transactions on Networking*, 1(6):614–627, December 1993.

[4] Zbigniew Dziong and Lorne G. Mason. Fair-Efficient Call Admission Control Policies for Broadband Networks - A Game-Theoretic Framework. *IEEE/ACM Transactions on Networking*, 4(1):123–136, February 1996.

[5] Drew Fudenberg and Jean Tirole. *Game Theory.* MIT Press, 1993.

[6] Leonard Kleinrock. *Queueing Systems, volume II: Computer Applications.* John Wiley and Sons, 1976.

[7] David M. Kreps. *Game Theory and Economic Modeling.* Clarendon Press, 1990.

[8] Jeffrey K. MacKie-Mason and Hal R. Varian. Pricing the Internet. In B. Kahin and J. Keller, editors, *Public Access to the Internet.* Prentice-Hall, 1994.

[9] Bridger M. Mitchell and Ingo Vogelsang. *Telecommunications Pricing – Theory and Practice.* Cambridge University Press, 1991.

[10] Tracie Monk and K. Claffy. Cooperation in Internet Data Acquisition and Analysis. In *Coordination and Administration of the Internet*, Cambridge, MA, September 1996.

[11] John Murphy and Liam Murphy. Bandwidth Allocation by Pricing in ATM Networks. Technical report, Dublin City University, Ireland, July 1994.

[12] Liam Murphy, John Murphy, and Jeffrey K. MacKie-Mason. Feedback and Efficiency in ATM Networks. In *Proc. of the 1996 Intl Conf. on Communications (ICC'96)*, pages 1045–1049, Dallas, TX, June 1996.

[13] Jon M. Peha and Fouad A. Tobagi. Cost-Based Scheduling and Dropping Algorithms to Support Integrated Services. *IEEE Transactions on Communications*, 44(2):192–201, February 1996.

[14] Scott J. Shenker. Making Greed Work in Networks: A Game-Theoretic Analysis of Switch Service Disciplines. *IEEE/ACM Transactions on Networking*, 3(6):819–831, December 1995.

[15] Qiong Wang, Jon M. Peha, and Marvin A. Sirbu. The Design of an Optimal Pricing Scheme for ATM Integrated Services Networks. *Journal of Electronic Publishing*, 1995. Special Issue on Internet Economics.