Proceedings of the International Conference on ATM, pages 78-84, 1998

ABR Rate Control for Multimedia Traffic Using Microeconomics*

Errin W. Fulp and Douglas S. Reeves
Departments of Electrical and Computer Engineering, and Computer Science
North Carolina State University
Box 7534, Raleigh, North Carolina 27695 USA
email: ewfulp|reeves@eos.ncsu.edu

Abstract

Multimedia applications are expected to play a more prevalent role in integrated service networks. One method of efficiently transmitting such traffic uses the ABR service class. However, rate control for this class becomes more difficult due to the bursty and somewhat unpredictable behavior of multimedia traffic. This paper presents a microeconomic-based ABR rate control technique that models the network as competitive markets. Prices are affixed to ABR bandwidth based upon supply and demand, and users purchase bandwidth to maximize their individual QoS. This yields a state-less rate control method that provides Pareto-optimal and QoS-fair bandwidth distributions, as well as high utilization. Simulation results using actual MPEG-compressed video traffic show utilization over 95% and better QoS control than max-min or demand-based weighted max-min.

Keywords: ABR rate control, multimedia applications, QoS perception, microeconomics.

*This work was supported by AFOSR grants F49620-96-1-0061 and F49620-97-1-0351. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the AFOSR or the U.S. Government.

1 Introduction

It has been demonstrated that transmitting multimedia traffic with the ABR service class is beneficial [1, 2]. For example in [2], Roberts demonstrated how the ABR service class can be used to transmit MPEG-compressed video resulting in utilization over 95% as compared to 30-60% using the Variable Bit Rate (VBR) service class. However, due to the dynamic and unpredictable nature of multimedia traffic, proper ABR rate control becomes more difficult.

ABR explicit rate control relies on network feedback provided by Resource Management (RM) cells that are circulated for each connection [3]. A RM-cell traveling from the source to the destination will be referred to as moving *upstream*, while a RM-cell traveling from the destination to the source will be referred to as moving *downstream*. The RM-cell consists of several fields, one of which is the Expected Rate (ER). This field indicates the maximum rate the network can support for this user. As the RM-cell travels along the path, a switch and/or destination may alter its contents. Exactly how this is done depends on the strategy. Once the cell reaches the destination it is returned to the source, which must alter transmission based on the RM-cell information.

Methods that perform rate allocation can be generally classified on whether they maintain per-connection state information [4]. Methods

that maintain state information that is directly used in the calculation of the allowable rate of a user will be referred to as *state-maintaining*. Alternatively, if per-connection state information is not required for the calculation of the allowable rate, it will be referred to as *state-less*. Of these two categories, a state-less method is preferred. Such a method does not require the overhead (storage and computational) of connection tables when computing allowable rates. Also, state-less implementations are scalable to larger networks since additional data structures are not required.

When a switch becomes congested, many of these ABR rate control strategies attempt to allocate bandwidth in a fair (weighted max-min) manner [3]. Examples of ABR rate control methods that achieve weighted max-min fairness include [1, 5, 6] and can be differentiated based on how weights are assigned. Using the Minimum Cell Rate (MCR), that is declared in the RM cell, as the weight was done by [5] and [6]. A Source with a larger MCR would receive a larger portion of the available bandwidth; yet, no evidence was provided indicating weights assigned in this manner are appropriate (especially for multimedia) or provide better results than max-min. Lakshman, et al. introduced another state-maintaining ABR rate control method for transmitting compressed video, where weights were based on the desired bandwidth of a user (demand-based weighted max-min) [1]. Simulation results demonstrated that this method can perform better than max-min when transmitting compressed video; yet, it assumes traffic that requires the most bandwidth is more sensitive to bandwidth reductions, which we will demonstrate is not necessarily true. Furthermore, these implementations were unfortunately state-maintaining.

An alternative method of bandwidth allocation incorporates microeconomic theory [7, 8, 9, 10, 11]. Microeconomic methods have been proven to achieve optimal distributions, such as *Pareto-optimal* and *equitable* allocations [12] (optimality and fairness definitions are provided in section 2). While these methods can achieve optimal allocations, they are not appropriate for ABR rate control, since network conditions are not permitted to change dynamically.

Microeconomic-based techniques designed specifically for ABR rate control include [13, 14]. In [13], switches allocate ABR bandwidth in a proportionally fair manner based on the "willingness-to-pay" provided by each user. When conditions change, users determine a new willingness-to-pay via a curve fitting process which relies on a history of previously optimal decisions. In the ABR rate control method of [14], users bid for some amount of effective bandwidth. While effective bandwidth allocates over a longer time scale, these techniques are difficult to apply to sources with little or no a priori information (for example, live and interactive video) and can be considered too conservative [14].

In this paper we introduce a state-less ABR rate control technique based on the "dynamic competitive market" model [15]. This model adapts to changing networks conditions (such as users entering/exiting the network and multimedia traffic) and is therefore appropriate for ABR rate control. In our method, ABR bandwidth is priced and users pay for their usage; however, it is important to note that this is done for **rate control** only, not revenue generation or cost recovery. Pricing ABR services for revenue, *not* rate control, is described in [16]. Advantages of our ABR rate control strategy include:

- State-less implementation
- High bandwidth utilization
- Equitable allocations (QoS-fair)
- Control of individual QoS
- Simple computations for switches and users

The remainder of this paper is structured as follows. Section 2 describes the competitive

market model and optimal allocations. Section 3 describes our microeconomic-based rate control method in detail. Section 4 describes the simulation results including comparisons to max-min and Lakshman's ABR rate control method, using actual MPEG-compressed traces. Finally, section 5 summarizes the results and discusses some open questions.

2 Competitive Market Model

The competitive market model consists of scarce resources and two types of agents, consumers and producers. Consumers require resources to satisfy wants. Producers create or own the resources sought by consumers. These agents come together at a market, where they buy or sell resources. Usually these exchanges are intermediated with money and the exchange rate of a resource is called its price. One method for setting the price in a competitive market is the tâtonnement process. First proposed by Lêon Walras, the tâtonnement process iteratively determines the price of a resource based excess demand [17]. The excess demand is a function of the total (aggregate) demand and supply of the resource. The price increases if the demand is greater than the supply and decreases when the demand is less than the supply. The iterative process repeats until a price is reached such that supply equals demand; at this point the market and price are in equilibrium. We will refer to the prices calculated before the equilibrium price is reached as *intermediate prices*. Buying and selling normally do not occur with the intermediate prices [18]; however, this constraint will not apply to our rate control method. This allows demands to change dynamically (hence the name "dynamic competitive market") and is achieved using a modified tâtonnement process described in the next section. When the market is in equilibrium the resulting allocation is weighted max-min fair, proportionally fair per unit charge and Pareto-optimal [12]. Pareto optimality is the allocation of finite resources such that no sub-set of agents can improve on their allocation without lowering the utility (satisfaction) of another. Many different Pareto-optimal allocations may exist in a competitive market in equilibrium. For this reason, we employ a social welfare criterion, the *equitable* criterion, to compare and rank Pareto-optimal allocations. In economics, the *equitable* criterion states that each agent should enjoy approximately the same level of utility. When applied to a computer network, an equitable allocation is one in which all users have the same perception of QoS (also referred to as QoS-fair). This is the measure of fairness used in this paper.

This model was chosen for our ABR rate allocation technique because of its ability to achieve certain desirable goals, such as Pareto-optimal distributions and price stability. The competitive market also has a simple structure and a well founded mathematical basis for analysis. We again emphasize that our goal is ABR traffic management with QoS.

3 A Proposed ABR Rate Control Strategy

This proposed ABR rate control strategy is based on the dynamic competitive market model [15], where pricing is done to promote high utilization as well as Pareto-optimal and equitable allocations. The resource priced is ABR link bandwidth[1] and will be considered a non-storable resource (similar to residential electricity). For this reason, users cannot purchase bandwidth with the intent to use it at a later date.

3.1 Switch

The switch owns the ABR link bandwidth that is sought by users. The network consists of several switches interconnected with links. For a unidirectional link between two switches, we consider the sending switch as owner of the bandwidth of that link. Each switch prices its

[1]Since we are pricing only ABR link bandwidth, all references to bandwidth will refer to ABR bandwidth.

ABR bandwidth based on local supply and demand. Therefore a single switch, having multiple output ports, will have one price associated with each output port, where i represents the ith link of the economy.

As defined by the ABR service class, a switch will periodically receive RM-cells. The RM-cell provides the user feedback about the links (or destination) in their route. We propose using the link price as feedback, since users must scale bandwidth consumption due to budget constraints. The price in the RM-cell is initialized to 0 by the source node. At the switch, the *current* price for link i is inserted into the RM-cell traversing link i if it is greater than the price already stored in the RM-cell. The switch can update a RM-cell traveling upstream or downstream. We assume that the price is stored in the ER field of the RM-cell; therefore, no additional field is required and no other information is placed/altered in the RM-cell.

The price for link i is calculated at the switch, at discrete intervals. At the end of the nth interval, the switch updates the price of link i using a modified tâtonnement process. A limitation of the tâtonnement process, in its original form, is the inability to dynamically adapt to changing demands (which are prevalent in networks). To handle such dynamics the following *modified* tâtonnement process [15] is used,

$$p^i_{n+1} = p^i_n \cdot \frac{d^i_n}{\alpha \cdot S^i} \qquad (1)$$

where p^i_{n+1} is the new price, p^i_n is the current price, S^i is the link capacity and d^i_n is the aggregate demand for bandwidth. In a dynamic competitive market, the modified tâtonnement process adjusts the price at regular intervals, based on the demand (received traffic) and the supply. The bandwidth supply is the total bandwidth times a constant α, where $0 < \alpha \leq 1$. This modification causes the price to increase after some percentage (α) of the total bandwidth has been sold. An *equilibrium price* p^i_* is reached at link i when the supply equals the demand. At equilibrium, the resulting allocation

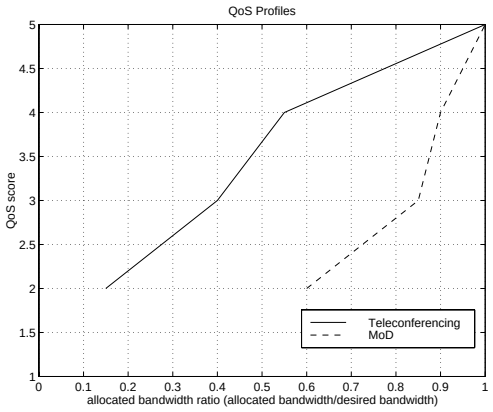


Figure 1: QoS profiles for teleconferencing and Multimedia on Demand (MoD) applications.

is Pareto-optimal and equitable [12]. A single equilibrium price does not exist for all time; however, the process will move towards the new equilibrium price when demand changes (proof provided in [12]). Since the price is calculated using only the aggregate (not individual) demand, supply and current price, this rate control method is state-less.

3.2 User

User j, executing a network application, desires a maximum amount of bandwidth for transmission, b^j_m. This amount of bandwidth maximizes the utility of the user and is expected to change over time (for example compressed video). In order to determine the allowable transmission rate a^j_r, the user must know: the route price, their wealth and their utility function.

As defined by the ABR service class, user j periodically generates RM-cells that circulate through the route to obtain feedback about network conditions. When the RM-cell reaches the destination it is returned (via the same route) to the user. The returned RM-cell contains the *route price*, which corresponds to the bottlenecked link in the route. The user is charged continuously for the duration of the session (analogous to a meter). To pay for the expenses, we will assume the user provides an equal amount of money over regular periods of

time. We will refer to this as the budget rate of user j, w^j (\$/sec). A method of wealth distribution that achieves an equitable allocation is provided in [12].

Based on prices and wealth, user j can afford a range of bandwidth, less than or equal to b_m^j. Preferences in the amount of bandwidth to use is provided with a utility function (individually defined for each user). We will use *QoS profiles* for the utility functions. The QoS profile is a function relating satisfaction to resource allocation, and is determined from psycho-visual experiments. The profile can be approximated by a piece-wise linear curve with three different slopes (examples are shown in figure 1). The slope of each linear segment represents the rate at which the performance of the application degrades when the network allocates a percentage of the maximum desired bandwidth (b_m^j). A steeper slope indicates the inability of the application to easily scale bandwidth (for example, high quality video), while a flatter slope signifies the application can more readily scale bandwidth requirements (for example, teleconferencing or data transmission). The horizontal axis measures the bandwidth ratio of allocated bandwidth to maximum desired bandwidth (b_m^j). The vertical axis measures the satisfaction and is referred to as a QoS score. Our QoS scores range from 1 to 5, with 5 representing an excellent perceived quality and 1 representing very poor quality. We will refer to an *acceptable* QoS score as any value greater than or equal to 3. As seen in the figure, if the allocated bandwidth is equal to the maximum desired bandwidth (b_m^j), the ratio is 1 and the corresponding QoS score is 5 (excellent quality). As this ratio becomes smaller the QoS score reduces as well. Profiles can be created for a variety of applications and redefined as users gain more experience. New and updated profiles can be easily incorporated within the economy as they become available. More information about QoS profiles and the relationship between bit-rate and quality can be found in [19, 20].

Finally, a new allowable transmission rate, a_{r+1}^j, is determined in response to a new price, or a change in application demand,

$$a_{r+1}^j = \begin{cases} \min\{\frac{w^j}{p_n}, b_m^j\} & \text{if } \check{b}^j \leq \frac{w^j}{p_n} \\ \emptyset & \text{otherwise, } \check{b}^j \text{ was not affordable} \end{cases} \quad (2)$$

where p_n is the most recent route price, $\frac{w^j}{p_n}$ is the maximum amount of bandwidth affordable and $\check{b}^j$ is the minimum bandwidth acceptable (determined from the QoS profile). As noted in the equation it is possible that the minimum is not affordable, due to the QoS constraint, prices and budgets. Properly managing such a situation is an area for future work.

4 Experimental Results

In this section the performance of the price-based rate control method is investigated via simulation. Experiments performed will consist of a realistic network configuration, allow users to enter/exit the network, have different application types and use actual MPEG-compressed traffic. A comparison is made with two other ABR rate allocation methods, max-min and weighted max-min. The max-min fairness criterion was chosen since it is sought by many current ABR rate control methods [3]. The max-min implementation was centralized and no communication overhead was included; therefore the max-min results presented here should be considered better than what is possible in practice. The weighted max-min rate control algorithm by Lakshman, et al. [1] was selected because it is described as an ABR rate allocation method for transmitting compressed video. Weights are equal to the desired bandwidth of each application; therefore this method will be referred to as "demand-based weighted max-min." This method requires frame prediction to allocate bandwidth before it is required; however a look-ahead buffer was used instead. For this reason, the performance of this method should be considered best possible[2]. Exper-

[2]A correction was made to the algorithm presented in [1] and was confirmed by the author.

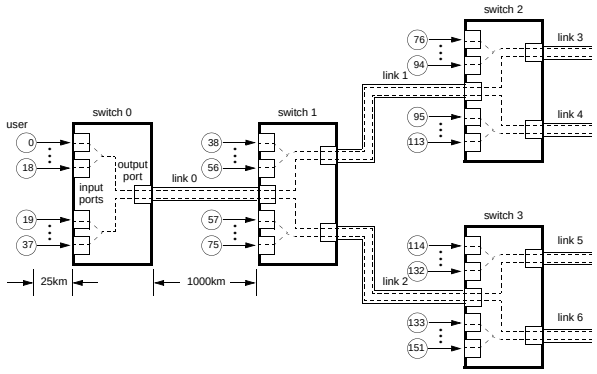


Figure 2: Network configuration used in simulations.

imental results will show that the proposed price-based rate control technique achieves high link utilization and equitable (QoS-fair) allocations, as well as better QoS control than max-min or demand-based weighted max-min.

The network simulated consisted of 152 users, four switches and seven 55 Mbps links, as seen in figure 2. The network can be described as a"parking lot" configuration, where multiple sources use a primary path. This configuration was agreed upon by members of the ATM Forum [21] as a suitable benchmark for allocation methods; it models substantial competition between users with differing routes and widely-varying propagation delays.

Since there is a variety of applications that transmit compressed video, user applications were considered one of two types: Multimedia on Demand (MoD) or teleconferencing. We are interested in determining if the rate allocation methods are able to provide equivalent QoS scores (utility) regardless of application type (equitable allocation). MoD applications require the transmission of high quality voice and video. These applications can scale bandwidth requirements only within a limited range, since bandwidth control is achieved through quantizer control. Teleconferencing applications, in contrast, can transmit lower voice and video quality. This is primarily due to quantizer control as well as the ability to transmit below the standard 24 or 30 frames per second. The difference is apparent from QoS profile associated with each application type, as seen in figure 1. Regardless of the type of application, the source for each user was one of 15 MPEG-compressed traces obtained from Oliver Rose at the University of Würzburg, Germany [22][3]. Users entered the network at random times uniformly distributed between 0 and 120 seconds.

The pricing strategy had the following initial values. MoD users had a budget rate[4], w, of 3×10^8/sec, while teleconferencing users had a budget rate of 1.5×10^8/sec. Teleconferencing users are given a lower budget because they are able to scale bandwidth requirements more readily. This was done to achieve a more equitable allocation. Although a method for determining the wealth of each user is presented in [12], for this simulation wealth was assigned based on the bandwidth ratio required to achieve a QoS score of 3. While the wealth assignment method is less complex, it will result in an allocation slightly less equitable than possible. Switches initialized their prices to 50 and α (the target utilization) to 95%. Switches updated their link prices at 10 msec intervals, a compromise between the desire for responsiveness, and the need for stability.

[3]Traces can be obtained from the ftp site `ftp-info3.informatik.uni-wuerzburg.de` in the directory `/pub/MPEG`

[4]The denomination is based on bps; if based on Mbps, the budget would be 300/sec.

For comparisons, we are interested in the link bandwidth utilization and the QoS provided to each user. Allocation graphs are provided to measure the utilization of link bandwidth. To measure the QoS observed, average QoS graphs, percent Good or Better (GoB) measurements and average QoS scores are provided. Average QoS graphs measure the average QoS score observed over time and are based on all users or on individual type. The percent Good or Better (GoB) measurement is the average percentage of time a user had a quality score of at least 3.

Results from the simulation are given in figure 3 and in table 1. As seen in figure 3(a), the allocation provided by the price method for link 0 indicates the total allocation stayed in the vicinity of 95% (α, the target utilization), yet never crossed 100%. Therefore, pricing was able to properly manage bandwidth demand (allocation results for the other links are very similar). For all users, the max-min and demand-based weighted max-min methods yielded lower average QoS and percent GoB values. This indicates, on average, users experienced lower QoS scores and enjoyed an acceptable QoS for shorter durations than the pricing method. More importantly, the pricing method provided both application types similar QoS scores and percent GoB values. This represents a more *equitable* (QoS-fair) allocation by the price method than max-min or demand-based weighted max-min. This is due to the inability of max-min or demand-based weighted max-min to differentiate between different classes of users.

5 Conclusions

This paper introduced an ABR rate control method based on microeconomics. Switches own the ABR bandwidth sought by users, and price their bandwidth based on local supply and demand. A user requires bandwidth to maximize their individual QoS. This competitive market structure encourages high utilization, with Pareto-optimal and equitable (QoS-fair) allocations. This results in a state-less rate control method that requires only simple computations. Simulation results demonstrate the ability of the economy to successfully allocate bandwidth of a network to a large number of diverse users, each transmitting an actual MPEG-compressed video trace. The economy also provided substantially better control of QoS than max-min or demand-based weighted max-min [1]. Finally, we believe the implementation cost will be very reasonable and the method can be incorporated into the existing ABR service class (no additional fields in the RM-cell are required and connection tables are not needed when determining allowable rates).

Acknowledgements The authors wish to thank Maximilian Ott and Daniel Reininger of C & C Research Laboratories, NEC USA for their significant contributions to this research.

References

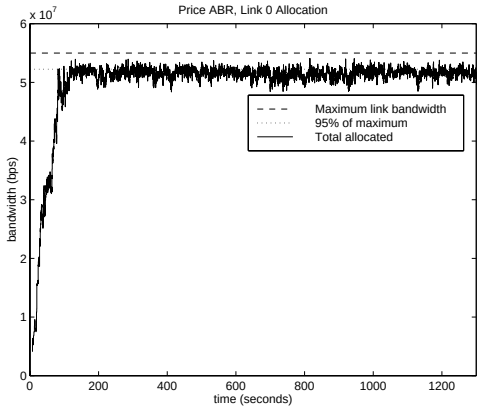
[1] T. V. Lakshman, P. P. Mishra, and K. K. Ramakrishnan, "Transporting Compressed Video over ATM Networks with Explicit Rate Feedback Control," in *Proceedings of the IEEE INFOCOM*, pp. 38 – 47, 1997.

[2] L. G. Roberts, "Can ABR Service Replace VBR Service in ATM Networks," in *Proceedings of the IEEE COMPCON*, pp. 346 – 348, 1995.

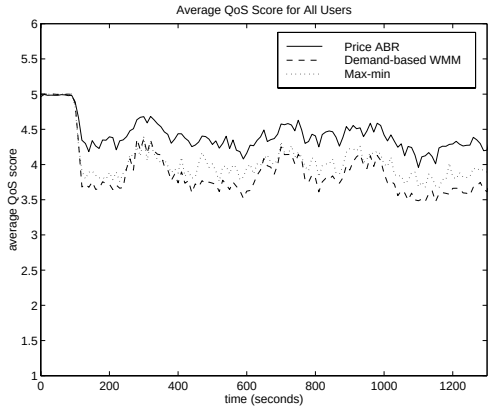
[3] ATM Forum Technical Committee, "Traffic Management Specification." Available through `ftp://ftp.atmforum.com/pub/approved-specs/af-tm-0056.000.ps`, 1996.

[4] L. Kalampoukas, *Congestion Management in High Speed Networks*. PhD thesis, University of California Santa Cruz, 1997.

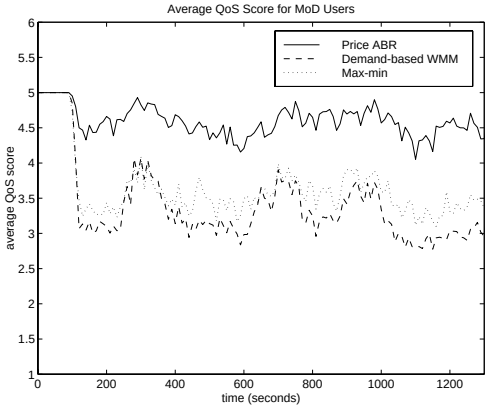
[5] Y. T. Hou, H. H.-Y. Tzeng, and S. S. Panwar, "A Weighted Max-Min Fair Rate Al-



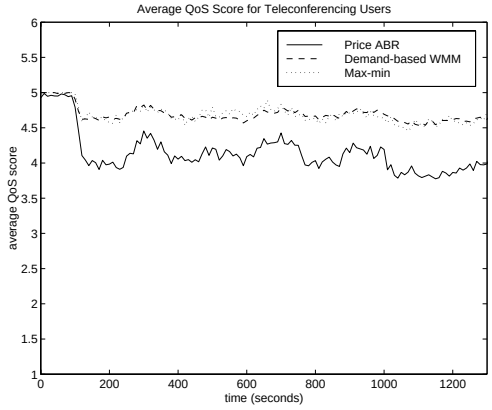
(a) Price ABR link 0 allocation.



(b) Average QoS score for all users.



(c) Average QoS score for MoD users.



(d) Average QoS score for teleconferencing users.

Figure 3: Allocation and average QoS scores.

	%GoB			Average QoS Score		
	All	MoD	Teleconf.	All	MoD	Teleconf.
Price ABR	90	90	90	4.43	4.63	4.14
Demand-based WMM	72	54	99	3.88	3.36	4.68
Max-min	80	66	99	4.25	3.92	4.76

Table 1: Percent GoB and average QoS scores.

location for Available Bit Rate Service," in *Proceedings of the IEEE GLOBECOM*, pp. 492 – 497, 1997.

[6] C. S. C. Lee, K. F. Cheung, and D. H. K. Tsang, "Generalized Weighted Fairness Criterion: Formulation and Application on Prioritized ABR Service," in *Proceedings of the IEEE Symposium on Computers and Communications*, pp. 512 – 516, 1997.

[7] N. Anerousis and A. A. Lazar, "A Framework for Pricing Virtual Circuit and Virtual Path Services in ATM Networks," *ITC-15*, pp. 791 – 802, 1997.

[8] D. F. Ferguson, C. Nikolaou, J. Sairamesh, and Y. Yemini, "Economic Models for Allocating Resources in Computer Systems," in *Market Based Control of Distributed Systems* (S. Clearwater, ed.), World Scientific Press, 1996.

[9] F. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate Control for Communication Networks: Shadow Prices, Proportional Fairness and Stability," *Journal of the Operational Research Society*, vol. 49, pp. 237 – 252, 1998.

[10] J. Murphy, L. Murphy, and E. C. Posner, "Distributed Pricing for ATM Networks," *ITC-14*, pp. 1053 – 1063, 1994.

[11] J. Sairamesh, 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 IEEE INFOCOM*, pp. 1111 – 1119, 1995.

[12] E. W. Fulp, *Allocation Methods for QoS Management in Computer Networks*. PhD thesis, North Carolina State University, 1999.

[13] C. Courcoubetis, C. Manolakis, and G. D. Stamoulis, "An Intelligent Agent for Negotiating QoS in Priced ABR Connections," in *Proceeding of the International Conference on Telecommunications*, 1998.

[14] C. Courcoubetis and V. A. Siris, "An Approach to Pricing and Resource Sharing for Available Bit Rate (ABR) Services," in *Proceedings of the IEEE GLOBECOM*, 1998.

[15] E. W. Fulp, M. Ott, D. Reininger, and D. S. Reeves, "Paying for QoS: An Optimal Distributed Algorithm for Pricing Network Resources," in *Proceedings of the IEEE Sixth International Workshop on Quality of Service*, pp. 75 – 84, 1998.

[16] C. Courcoubetis, V. A. Siris, and G. D. Stamoulis, "Integration of Pricing and Flow Control for Available Bit Rate Services in ATM Networks," in *Proceedings of the IEEE GLOBECOM*, pp. 644 – 648, 1996.

[17] L. Walras, *Elements of Pure Economics*. Richard D. Irwin, 1954. trans. W. Jaffé.

[18] A. Takayama, *Mathematical Economics*. Cambridge University Press, 1985.

[19] E. Nakasu, K. Aoi, R. Yajima, K. Kanatsugu, and K. Kubota, "Statistical Analysis of MPEG-2 Picture Quality for Television Broadcasting," in *Proceedings of the 7th International Workshop on Packet Video*, vol. 11, pp. 702 – 711, Nov. 1996.

[20] D. Reininger and R. Izmailov, "Soft Quality-of-Service for VBR+ Video," in *Proceedings of the International Workshop on Audio-Visual Services over Packet Networks, AVSPN'97*, Sept. 1997.

[21] A. Kolarov and G. Ramamurthy, "Comparison of Congestion Control Schemes for ABR Service in ATM Local Area Networks," in *Proceedings of the IEEE GLOBECOM*, pp. 913 – 918, 1994.

[22] O. Rose, "Statistical Properties of MPEG Video Traffic and Their Impact on Traffic Modeling in ATM Systems," Tech. Rep. 101, University of Würzburg Institute of Computer Science, Feb. 1995.