8-16-1996

Pricing for Market Segmentation in Data Networks

Philipp Afeche
Stanford University

Haim Mendelson
Stanford University

Pricing for Market Segmentation in Data Networks

Philipp Afèche and Haim Mendelson
Graduate School of Business, Stanford University

The use of data networks has increased tremendously over the past few years. A prime example is the Internet, where the proliferation of commercial Internet Service Providers and the development of new bandwidth-intensive multimedia applications drives exponential growth rates, both in terms of the number of users and the traffic volume they generate. While technology advances are expanding the available bandwidth, it is unlikely that absent some structural changes, sufficient capacity can be economically supplied in the near future to keep up with traffic growth. As a result, the Internet is plagued by congestion problems which are likely to prevail and become more serious unless congestion is effectively controlled.

Fueled by the commercialization of the Internet and by the National Information Infrastructure (NII) initiative, this has raised many economic and public policy questions pertaining to the effects of network service pricing on usage patterns and the benefits of providers, users and society as a whole (e.g., MacKie-Mason and Varian, 1994a, 1994b, 1995).

Our objective is to provide insights into some of these effects as a basis for decision makers. Specifically, we focus on the commercial network provider who seeks to maximize profits by segmenting a heterogeneous user market with multiple message classes in the presence of network congestion. In this setting, we investigate the effects of two pricing schemes under best-effort and priority service on network usage levels, provider profits and consumer surplus in the various market segments and in aggregate. While our analysis is motivated by the Internet scene, it is applicable to other congestible data networks as well.

This article is a summary of our current findings on these issues. Specifically, section 1 describes our modeling approach and sections 2 - 4 summarize our results. A paper providing the substance of our analysis is under way.

1. Model

Our modeling approach for describing the user popula-tion, the data network and demand for network services is motivated by Mendelson (1985). The users are characterized by their message valuations, message sizes and delay sensitivities. We consider a continuum of atomistic users, each potentially generating on average one message per unit time. Users are heterogeneous with respect to their message valuations. We assume that users' ordered gross message values are given by constant elasticity marginal value curves.

The message sizes, expressed in bits, are independent random variables and also independent of users' message valuations. The sizes of messages within the same class are identically distributed.

The net value users derive from a message transmis-sion is sensitive to the quality of service delivered by the network. We focus here on traditional packet switched networks without resource reservation and assume that the packet switch buffers are of infinite capacity. Therefore, quality of service is measured by the message transmission delay, which accounts for both the queueing delays in the packet switch buffers and the actual transmission time. Delays are assumed to be costly to users, either due to the opportunity costs of time they incur while awaiting the system response or due to tight technological delay bounds for certain real-time voice and video applications. We capture this delay sensitivity by multiplying users' gross message value by a delay discount factor which is decreasing in the transmission delay. This specification of users' delay sensitivities is fairly general and encompasses a number of special delay discount functions of interest. The three cases which we focus on in parts of this paper are those of linear, exponential, and deadline delay discount functions. These are suitable for different kinds of applications such as E-mail, remote login, file transfer or real-time video.

We focus on the special case of a single transmission link, whose delay performance we model as a single-server M/G/1 queue (Kleinrock, 1976). We assume that the link bandwidth is fixed over the relevant decision horizon. Therefore, capacity costs do not enter the analysis. The times between transmission requests and the message transmission times are independent random variables, drawn from separate probability distributions. We first assume that all users are offered a single class of best-effort service, as is the case in the current Internet, and then extend the model to the case of multiple priority classes.

The demand rates (i.e., the rates of transmission requests per unit time) for the various message classes are determined by users' gross message values, expected delay costs, and network service charges. We assume that users are self-interested, i.e., in making their network usage decisions, they only seek to maximize their own net value regardless of the effect on the net value of other users. For example, in the case of a single class of best-effort service with a uniform message price, each user transmits a message if and only if its value net of delay cost equals or exceeds this price. The inverse demand relationships are derived by equating the marginal users' *net* message value to the price charged. The network usage decisions for the case of multiple prices and/or service classes are similarly determined.

In this setting, we study the pricing decisions of *(i)* a welfare maximizing and *(ii)* a profit maximizing network provider under best-effort service with uniform pricing and price discrimination, and under service and price differentiation.

2. Uniform Pricing

We consider as a benchmark the case in which all message sizes are identically distributed and users are offered a single class of best-effort service with a uniform message price. We find the remarkable result that the monopolist sets the *socially optimal* price leading to welfare maximizing network usage levels, while only capturing a fraction of the social net value. This result is fundamentally different from other cases in the microeconomics literature where a monopolist leads to socially efficient consumption levels. In those cases, the monopolist uses some form of price discrimination which allows her to extract all surplus from the consumers (e.g., Tirole, 1988; Varian, 1989). Since the monopolist then earns the entire social net value, the quantity maximizing monopoly profits is socially efficient even though the distributional effect is potentially questionable. Here, the social optimum is obtained without requiring that the monopolist capture the entire surplus.

3. Price Discrimination

While a nondiscriminating monopoly network provider sets the same price as a social planner, she only extracts part of the consumer surplus. In this section, we study what happens when the monopolist can use information about each user's message values to price discriminate. We assume that the monopolist first determines the *degree of discrimination* M, i.e., the number of different prices M to be charged. She then derives the profit maximizing levels of these prices, knowing that each message can be charged the highest price not exceeding its net value. This enables the monopolist to charge M different prices, one for each discernible user segment.

We examine this scheme's impact on the profit maximizing network usage levels and the resulting monopoly profit, social net value and segmentation characteristics as a function of the degree of discrimination M. We find that:

a) The profit maximizing aggregate network usage and the resulting social net value are invariant to the degree of discrimination and equal the socially efficient levels obtained in the case of no price discrimination.

b) The monopoly profit is increasing and concave in M. It equals the social net value in the limiting case of perfect discrimination with infinitely many prices. Here, the monopolist extracts all surplus from each individual user.

c) For a given degree of discrimination M, the segment size and revenue are increasing, and the segment consumer surplus is decreasing, as we move to the lower-valued segments.

Hence, while the effect of price discrimination on the net benefits to society is neutral, it results in a redistribution of the surplus away from the users and to the network provider.

4. Service and Price Differentiation

Pure price discrimination is very attractive for the monopolist in theory. However, it is difficult to implement in practice, since the required information on users' message valuations is only partially available. In this section, we study what happens when the provider, using service differentiation, can induce users to pay higher prices in return for a higher service priority. Specifically, we assume that the provider offers users a schedule of M prices and priority service classes. The schedule requires no information on individual users' message values and is compatible with their incentives. Users choose from this price-service schedule, which results in M market segments. In contrast to previous work on priority pricing (e.g., Bohn et al., 1993; Cocchi et al., 1992; Gupta et al., forthcoming; Mendelson and Whang, 1990), the number of prices and priority classes, i.e., the *degree of differentiation* M, is endogenous in our model.

We study the impact of this pricing scheme on network usage levels and the resulting provider profits, social net value and segmentation characteristics as a function of M and under the alternative objectives of profit or social welfare maximization. For simplicity, we focus on the M/M/1 case with a linear delay discount function. We expect similar results to hold for the cases of exponential and deadline delay discount functions. We find that:

a) The profit maximizing prices are also socially efficient, leading to welfare maximizing network usage levels. This extends the result obtained in the case of uniform pricing and best-effort service.

b) The optimal aggregate network utilization and monopoly profit are increasing and concave in M. Compared to the case of a single class of best-effort service, they can grow by over 10-20% in the limiting case of infinitely many priority classes. The profit gains resulting fromprioritization are larger the more heterogeneous the users' message characteristics.

c) The aggregate consumer surplus, and hence the aggregate social net value, increase in proportion with the monopoly profit.

d) For a given degree of differentiation M, the segments comprising the lowest and the highest value users generate the smallest revenue, those consisting of medium value users the largest. The segment consumer surplus is decreasing as we move to lower-valued segments.

Thus, while price and service differentiation is not as profitable for the monopolist as price discrimination, it is amenable to implementation and results in small gains compared to the nondiscrimination case. Moreover, we find that service differentiation is also beneficial to users.

5. Conclusions

In conclusion, the research findings support current efforts for implementing network protocols that offer multiple priority classes, such as the next generation of the Internet protocols. While the proposed pricing mechanisms were analyzed for the case of a single communication link, they can be implemented on a network topology with multiple links by application to each individual network node. In order to provide guidelines for network design and management, future research will explore the economic performance of different degrees of service differentiation in experimental network environments as a function of various network and user characteristics.

Acknowledgment

Partial financial support by the Stanford Computer Industry Project is gratefully acknowledged.

References

Bohn, R., Braun, H.-W., Claffy, K.C. and Wolff, S. "Mitigating the coming Internet crunch: multiple service levels via Precedence," *Journal of High Speed Networks* (3:4), 1994, pp. 335-349.

Cocchi, R., Estin, D., Shenker, S. and Zhang, L. "Pricing in computer networks: motivation, formulation, and example," *IEEE/ACM Transactions on Networking* (1:6), December 1993, pp. 614-627.

Gupta, A., Stahl, D.O. and Whinston, A.B. "An economic approach to networked computing with priority classes," *Journal of Organizational Computing*, forthcoming.

Kleinrock, L. *Queueing Systems*, Vols. I and II, Wiley-Interscience, New York, NY, 1976.

MacKie-Mason, J.K., and Varian, H.R. "The economics of the Internet," *Dr. Dobb's Information Highway Sourcebook* , Winter 1994a, pp. 6-16.

MacKie-Mason, J.K. and Varian, H.R. "Pricing the Internet," in *Public Access to the Internet*, B. Kahin and J. Keller (eds.), Prentice Hall, Englewood Cliffs, NJ, 1994b.

MacKie-Mason, J.K. and Varian, H.R. "Pricing congestible network resources," *IEEE Journal on Selected Areas in Communications* (13:7), September 1995, pp. 1141-1149.

Mendelson, H. "Pricing computer services: queueing effects," *Communications of the ACM* (28:3), March 1985, pp. 312 - 321.

Mendelson, H. and Whang, S. "Optimal incentive-compatible priority pricing for the M/M/1 queue," *Operations Research* (38:5), September-October 1990, pp. 870 - 883.

Tirole, J. *The Theory of Industrial Organization*, MIT Press, Cambridge, MA, 1988.

Varian, H.R. "Price Discrimination," in *Handbook of Industrial Organization*, Vol. I of *Handbooks in Economics*, R.Schmalensee and R.D. Willig (eds.), Elsevier Science Publishing, New York, NY, 1989, pp. 597-654.