December 2001

Pricing and Capacity Planning for Internet Dial-Up Lines

Hsing Cheng
University of Florida

Kutsal Dogan
University of Florida

Richard Elnicki
University of Florida

PRICING AND CAPACITY PLANNING FOR INTERNET DIAL-UP LINES

Hsing Kenneth Cheng
University of Florida
hkcheng@ufl.edu

Kutsal Dogan
University of Florida
kutsal@ufl.edu

Richard A. Elnicki
University of Florida
dicke@ufl.edu

Abstract

Recent events experienced by Internet Service Providers have proved that unlimited e-mail and Internet access provided at flat fee causes congestion and is not the best method for pricing dial-up services. It is suggested that the use of congestion pricing instead of unlimited access at a flat rate would soften the congestion problem. The congestion and pricing situation is rather different for nonprofit Internet Service Providers. The mandate for these nonprofit entities is to recover the costs of providing their services while maintaining a guaranteed QoS level. This paper proposes a model that effectively solves this challenging pricing and capacity-planning problem.

Keywords: Internet Service Provider, pricing, capacity planning, e-commerce

Introduction

Internet access provided at a flat fee is not necessarily the best method for pricing dial-up services. Unlimited access results in excessive usage, and users have to live with low Quality of Service (QoS) levels and experience frustratingly long waiting times. With a flat fee pricing mechanism, the users have no incentive to plan or limit their use of the network resources. It is suggested that the use of congestion pricing instead of unlimited access at a flat rate would soften the congestion problem. MacKie-Mason and Varian (1995) argue that the use of connect fee and usage pricing is much more socially beneficial than a simple flat fee mechanism. Gupta, et al. (1996) examine the use of a different mechanism that imposes a "congestion toll" on the users according to additional delay they cause to other users. Their model also provides increased total net benefit compared to free access.

There is clear evidence in literature that suggests the use of congestion pricing as best way to price dial-up network services. This might be true for for-profit Internet Service Providers (ISPs); however, the congestion and pricing situation is rather different for nonprofit ISPs such as the one analyzed by this paper. The mandate for these nonprofit entities is to recover the costs of providing their services while maintaining a guaranteed QoS level. Their pricing and capacity-planning problem has three important aspects.

First, pricing based on cost recovery has inherent drawbacks. The administrators of these nonprofit entities have to take into consideration the trade off between capacity level and QoS level. In the context of Internet dial-up services, capacity is defined as the number of available dial-up lines and QoS level is the percentage of customers who get a dial-up line within one minute. This can also be interpreted as the probability that a customer will get a line within one minute. For the same demand level, the capacity required to provide 99% QoS level is larger than the capacity required for 95% QoS level. Similarly, the larger the capacity or the better the QoS level, the higher the average costs. But there is more to this simple relationship. If the desired QoS level is too high, this might result in high average cost that might drive away the demand and cost recovery might become impossible. On the other hand, at sufficiently low QoS level the congestion might result in extreme customer dissatisfaction, driving away the demand and making cost recovery impossible again.

Second, we have to tackle the variation of growth in demand, which might call for a continuous capacity expansion. The demand growth is very unpredictable. The market for ISPs becomes more competitive as new services and providers enter the market. The competition may drive away the demand if certain QoS level is not provided at an average cost that is competitive. Some customers might even select a high price-high quality service even though low quality service is provided free. Customers' utility definitely deteriorates exponentially with increased waiting times.

Third, capacity expansion comes only in bulks and has a considerable lead-time. The number of available dial-up lines cannot be increased continuously. Nonprofit ISPs should consider economies of scale in capacity expansion. Timing of expansion is also important. Capacity expansion decision should consider the seasonal fluctuations in demand. During the low demand periods, the capacity expansion may be delayed to reduce the cost and the risk of failure.

To solve this threefold problem, nonprofit ISPs need a mechanism to determine the optimum capacity level by taking into account the desired QoS and future demand growth. This paper proposes a model that will effectively resolve aforementioned issues.

The Model

The model proposed here is tested using data from a nonprofit ISP: the State of Florida's Northeast Regional Data Center (NERDC) at the University of Florida. Background of NERDC and system configuration is reported in an earlier work of one of the authors, see Elnicki (1997). The services provided by the NERDC include access to graphic – or character – based services and access to the Internet. Anyone affiliated with the University of Florida can acquire a dial-up line with a valid user ID and a password either from their homes or from their offices. There are currently 1,330 lines available to connect, each of which provides connection speed up to 56,000 bps (V.90). The current cost of access time is $0.008 per minute via local lines and $0.20 per minute via long distance lines.

The measurement of demand has significant effect in modeling. Our model uses the busiest minute demand that is analogous to peak-load demand used in earlier studies described in Elnicki (1997). The busiest minute for a given period of time is found using charge records that provides connect and disconnect times for all users in seconds for all dial-up sessions.

The administrators of the NERDC want to achieve a very high QoS level. The requirement is 99%, e.g., the probability that a user will get a line within one minute should be greater than or equal to .99. The one-minute figure is chosen to allow users' modem to make 3 attempts to establish a dial-up connection. The model can determine the worst-case QoS level if a certain capacity is provided given the existing distributions of busiest minute demands. The optimum capacity level lies somewhere between "excessive" capacity that provides high QoS level but drives away demand due to high average cost (therefore high prices) and "insufficient" but low-cost capacity that drives away demand due to unacceptable QoS level. To maintain the desired QoS level, our model predicts the future load of the network in terms of busiest minute demand using a regression model. The point where the future load implies a QoS level lower than the requirement is the time for the next capacity extension. Therefore, the next capacity addition should be made by this projected date considering the lead-time for installation and testing. Then, the model predicts the point in time when this projected expansion will be exhausted. Given the knowledge of the capacity, its total costs, and the duration it will last, average cost will be calculated. The administrators can then compare the average cost with other market prices and decide the price and whether or not the service should be continued.

Application of the Model and Future Research

The NERDC has been using the model suggested in this paper since 1997. A different decision rule had been used for capacity expansion before our model. The decision to add capacity was made based on ad-hoc forecast of the time when the high-water-mark, the highest number of connections at any instance, would hit the available capacity (Elnicki 1997).

Instead of an ad-hoc forecast of the high-water-mark our model uses an effective forecasting model. Our latest regression model has data for charge records for 1334 days. Since we are trying to forecast the busiest minute demand in future, we need to find the busiest minute demand figures for each of these 1334 days. The busiest minute is determined by using a charge vector of size 1440 (one entry for each minute of day) that represents the number of active dial-up connections. This number of connections during the busiest minute used as the dependent variable of the regression model. Our sole purpose of using the regression model is prediction.

The independent variables that are used in the final model to estimate the dependent variable include a time variable to account for general trend of demand (days 1 to 1334), dummy binary (1,0) variables to account for demand fluctuations due to vacations, fall-spring-summer semesters, and day-of-week (binary variables for 6 out of 7 days of week) and finally a dummy variable to account for fluctuations in supply of dial-up lines. Casual observation of data shows increase in busiest minute demand over time; therefore, we expect time variable to have positive effect on the busiest minute figure. We also expect lower figures during vacations and summer semesters. The final variable used for supply represents the anomalies in dial-up servers such as bad connections. Bandwidth capacity is not an issue for the NERDC so it is not included in the model. Vacation days should have a great impact on the demand and the capacity expansion decisions should definitely consider them before the capacity expansion is actually made. The time variable also has a positive impact on the busiest minute demand.

GLM procedure of SAS is used to find the best predictor for the dependent variable. Our latest model seems to provide a very good fit; however, there appears to be a structural change in the growth of demand. Aggregate demand in terms of total minutes per month and the number of sessions per month has been decreasing whereas minutes per session have been increasing since 1997. Therefore, the effect of this change on busiest minute connection figures may not be so obvious. It might be the case that even though aggregate demand is decreasing, busiest minute figures may not exhibit a similar pattern. If indeed the structure of the demand changed, a single regression model using all 1334 days would not be appropriate. The statistical tests to analyze this issue will be carried out soon.

References

Elnicki, R.A. “Dial-Up Server Capacity Planning Model,” *Proceedings of the Association for Information Systems 1997 (AIS97) Americas Conference*, Indianapolis, Indiana US, August 15-17, 1997, http://hsb.baylor.edu/ramsower/ais.ac.97/papers/elnicki.htm

Gupta, A., Stahl, D. O., and Whinston, A. B. *Pricing of services on the Internet,* Technical report, University of Texas at Austin, 1996, http://cism.bus.utexas.edu/alok/pricing.html

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