The Case for Quality of Service On Demand
Empirical Evidence from the INDEX Project*

Jörn Altmann   Björn Rupp   Pravin Varaiya

1 Introduction

In order to ensure further Internet growth and efficiently support quality-differentiated network services, users' choice options have to go beyond different service plans that only reflect a rough market segmentation. To further subdivide these segments, choice options should provide the means to let users express their current needs by instantaneously selecting the service quality. This argument is strongly supported by empirical evidence from the Internet Demand Experiment (INDEX), a market trial for quality-differentiated Internet services. Arguing that the vast majority of Internet service plans for residential customers encourage waste and lead to user cross-subsidies, this chapter investigates user heterogeneity, activity heterogeneity, and acceptance of Quality of Service on demand to derive some consequences for Internet service provisioning and tariff design.

2 Background and Motivation

2.1 Quality and Waste

The economic model that the Internet has been using needs modifications to support new applications and different service qualities. However, current Internet service plans have made little progress towards supporting this goal: They are predominantly flat-rate and offer little, if any, choice of different service options. Currently, users who occasionally need high bandwidths are either forced to lease over-provisioned dedicated lines, risk the caprices of shared resources (best-effort quality), or forego the desired application altogether.

This leads to two kinds of inefficiencies: First, flat-rate pricing encourages waste. Whenever the marginal cost of network resource utilization is zero (like under a flat-rate pricing scheme), users do not have to optimize marginal utility and marginal cost. Inefficient over-utilization of resources occurs. Second, in a network with "best effort" shared resources like the current Internet, flat-rate pricing leads to noticeable quality deterioration for all users if those resources are over-utilized. Ultimately, if the network is consistently unable to fulfill the quality requirements that a certain application or individual user has, that user is effectively excluded from using the network for these purposes.

2.2 Integrated Services

Network externalities have fueled the trend towards an Integrated Services Internet. Different application classes (e.g., streaming video vs. file transfers) naturally have very different quality requirements in terms of packet loss, transmission delays, minimum required bit rate etc. Users of different application classes also have heterogeneous Quality of Service (QoS) preferences.

Two basic network designs can meet those re-

*This research was supported by grants from the National Science Foundation, Cisco Systems, SBC Communications, the California State MICRO Program, Hewlett-Packard and the German Research Society, Berlin-Brandenburg Graduate School of Distributed Information Systems (DFG grant no. GRK 316).

quirements: In case that connecting and subscribing to different networks is associated with very low costs, a market equilibrium with different network service providers offering different service qualities at different prices may prevail. With the emergence of a multitude of new services, however, it is at least questionable whether an increasing number of different networks tailored to support these services will not result in comparatively high costs. Subscribing to an integrated services network, in contrast, yields significant utility gains for every new network user. In addition, the potential penetration of new network services is increased significantly. New services can be accessed without the need for connecting and subscribing to another network. Therefore, there is an incentive for the integration of as many services as possible in one network.

The integration of new services will ensure further Internet growth and allow for its even wider dissemination among the general population. The division of services into quality-differentiated market segments and the design of appropriate pricing structures for each segment are therefore crucial for further proliferation of Internet services.

2.3 Demand

In an increasingly competitive environment, service providers have to offer combinations of quality and price that match user needs. While much effort in recent literature has been invested in the design of adequate pricing proposals (for a short overview of different approaches, see (Shenker et al. 1996)), the understanding of the demand structure for quality-differentiated network services is still in its infancy. However, such understanding is critical for future network provisioning decisions.

3 Empirical Evidence

We aim to bridge the gap between supply-side and demand-side analysis by supporting the theoretical argument for flexible QoS choices with empirical evidence from the INternet Demand EXperiment (INDEX). INDEX is a real-world market trial for quality-differentiated network services. It provides Internet access over 128 kbps ISDN lines to a group of users from the Berkeley campus community (students, faculty, staff). Users select network services from a menu of QoS-price offerings and pay for their usage. They control their usage of network resources by means of a Java application running on the user's computer (Figure 1). The subjects can choose a service quality instantaneously by clicking on a button and change the Quality of Service even during an active session. The application also provides usage feedback by displaying a summary of charges accumulated over the session, the day and the month. A detailed overview of the technology, experimental setup and design of INDEX can be found in (Rupp et al. 1998).

Figure 1: INDEX User Interface

While we will occasionally refer to other INDEX sub-experiments for comparison purposes, the reasoning of this paper is mainly based on data from INDEX' first sub-experiment, *Variable Symmetric Bandwidth*, in which users are given the choice of six different bandwidths (8, 16, 32, 64, 96, and 128 kbps). Subjects are charged a per-minute rate that depends on the selected connection speed. 8 kbps service is priced at zero cents per minute. Prices for all other bandwidths are randomly drawn from a set of prices ranging from a minimum of 0.1 cents/minute to a theoretical maximum of 20.94 cents/minute. Prices are varied during the experiment to measure the demand response for each individual. The data in this paper was collected by analyzing the usage patterns of 70 subjects. It covers the period from April 1998 to February 1999.

3.1 User Heterogeneity

The utility that is derived from the consumption of network resources depends on individual characteristics and preferences. This subsection highlights the extent of user heterogeneity by examining the budget variation for Internet usage, time spent on the Internet, and the weekly mean expenditure.

Figure 2 shows a histogram of the weekly mean expenditure of the subjects over three different experiments. Besides data from the *Variable Symmetric Bandwidth* experiment, we also use data from the *Variable Asymmetric Bandwidth* experiment (in which subjects can choose different bandwidths for traffic *from* the Internet and *to* the Internet separately) and the *Byte Volume* experiment (in which subjects pay for the number of transmitted bytes). That compensates for the effect of different pricing structures. Therefore, each subject is represented three times, once per experiment. The variation in subjects' weekly mean expenditure spans a range from \$0.20 to \$21.23. Although about 40 % of the subjects spend less than \$2 per week, there is a considerable variation in expenditure in the \$2–\$22 range.

Instead of charging for individual usage, we could impose a flat-rate tariff set to recover the same revenues (by dividing total expenditure by the number of users). In this case, we would however only address the needs of 25 % of the subjects, i.e. people spending between \$2 and \$4. We would deprive ourselves of the additional revenue that could be generated by the users with a higher willingness to pay. At the same time, we would exclude users with only very casual Internet activity from using the service at all.

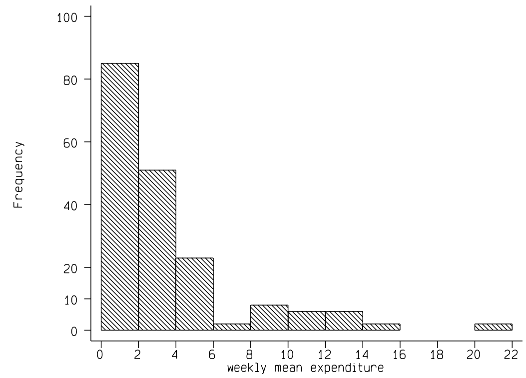

Figure 2: Expenditure Histogram

Looking at the differences between the three experiments, it becomes obvious that the disparities in expenditure patterns are not simply due to environmental or seasonal effects. They are rather an inherent characteristic of individual demand. Figure 3 shows this effect: The mean weekly expenditure of 48 % of all INDEX subjects varied only in a range of \$2. Considering that users were facing widely varying prices and disparate pricing schemes over the course of the three experiments, this result gives clear evidence that a significant percentage of users does have an exact idea of how much it intends to spend for Internet service in a given time period. Another 24 % of our subjects also set relatively tight constraints on their Internet budget, allowing for a maximum variation of \$4. Only the remaining 26 % displayed a significantly wider variation in their expenditure distribution.

Beside user expenditure, the time that users spend online is another indicator of user heterogeneity. Figure 4 compares per-minute pricing with flat-rate pricing. It demonstrates the high degree of cross-subsidy between light and heavy users under a flat-rate tariff. To analyze this effect, we aggregate and normalize the expenditure data on a per-user basis. We then rank the 70 subjects

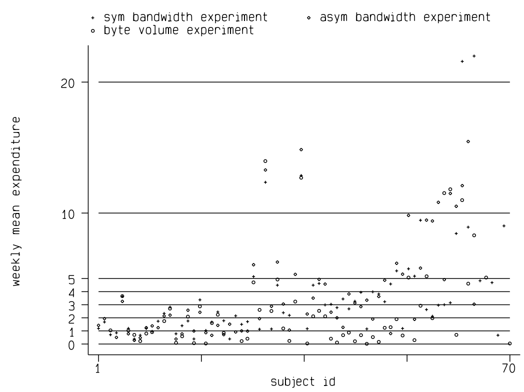

Figure 3: Budget

by the time they spent online (processing heaviest users first). The resulting upper curve of Figure 4 plots the actual cumulative expenditure from the *Variable Symmetric Bandwidth* experiment versus cumulative connect time. The curve starts out from the heaviest users close to the graph's origin and proceeds to the lightest users on the far right. Each dot represents one subject.

Each user is then imputed a flat-rate expenditure. Under this flat-rate pricing scheme, the imputed flat rate times the number of users equals the total revenue generated by the actual per-minute charges from the *Variable Symmetric Bandwidth* experiment. The lower curve in Figure 4 represents the cumulative expenditure under this flat-rate tariff.

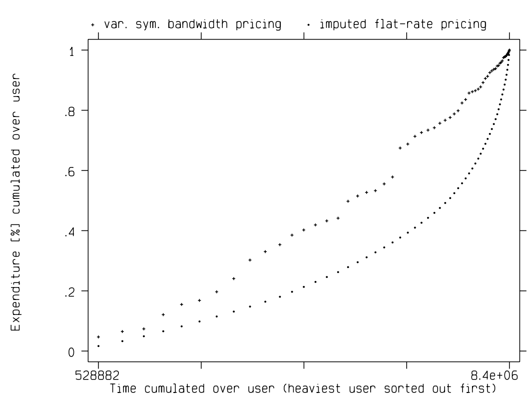

Figure 4: User Cross-Subsidy

The calculations yield results that relate well to other observations from the telecommunications field. 20 of the 70 subjects (28.6 % of the subject population) consume more than 75 % of network resources, measured in connect time to the Internet. Under flat-rate pricing, these heavy users would be subsidized by light users and would account for only about 30 % of overall expenditure. However, under per-minute pricing, these 20 users are charged in proportion to their usage. They have spent approximately 45 % of the actual overall expenditure. Therefore, we can conclude that usage-based pricing is a fair way to charge people and is significantly more equitable than the predominant flat-rate Internet service plans.

3.2 Activity Heterogeneity

After exposing the high degree of inter-user heterogeneity, we now turn to intra-user heterogeneity. In particular, we will look at two factors that show how individuals use the Internet for different purposes: time of day and activity type.

Classical peak-load pricing models heavily rely on time of day for allocating resources. Such pricing models assume that demand for a non-storable good such as bandwidth can be divided in different sub-periods in which the demand function does not change significantly. While we still see notable demand variations in terms of QoS choices made during a given period, we have been able to identify temporal patterns in network resource consumption. Figure 5 shows that network usage changes over the course of a day. Time of day (measured in minutes) is plotted versus frequency of use. One dot represents measured activity at the minute indicated, accumulated over the entire 6-week duration of the experiment. The underlying data is again taken from the *Variable Symmetric Bandwidth* experiment. The lower curve represents usage of paid services (16 kbps, 32 kbps, 64 kbps, 96 kbps, 128 kbps). The upper curve represents total usage (i.e. paid service plus free 8kbps service).

We can easily identify times of high traffic den-

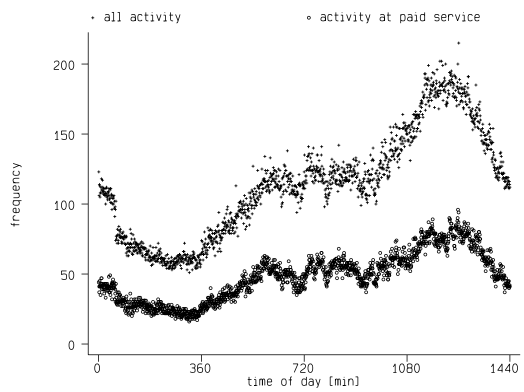

Figure 5: Time of Day Activity

sity, with a clearly visible peak time from 8 to 9 p.m. versus very little traffic in the early morning hours around 5 a.m. These findings are consistent with other data gathered on UC Berkeley modem pool usage.

While this data suggests that certain temporal patterns of residential Internet usage are relatively predictable, the detailed actual usage patterns show a wide variation in type and extent of usage. The same network infrastructure is used for very different tasks. Different applications and preferences lead to heterogeneity in individual network usage and resource consumption.

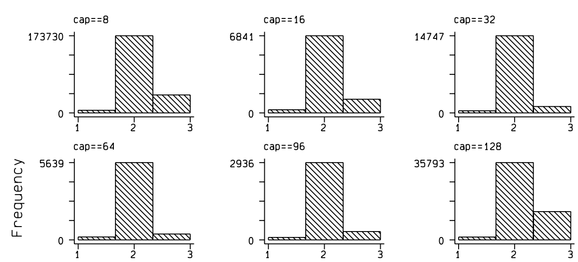

Figure 6: Activity Type

To illustrate this, we will focus on three different types of activity: *Bulk Traffic* (e.g. FTP, streaming data), *Burst Traffic* (e.g. World Wide Web) and *Interactive Traffic* (e.g. Telnet, X Windows).

In order to determine the activity type, we measure the average size of packets transferred during a minute. The classification is derived from this measure of traffic density and carried out for each minute of recorded in-bound traffic. As a consequence, we can only classify based on the main activity type in a given minute (i.e. either bulk, burst, or interactive traffic, but not a combination of different activity types). Activity is defined as bulk traffic if the number of packets is small compared to the number of bytes transferred, i.e. the average packet size is larger than 1000 bytes. If the average packet size is smaller than 45 bytes, then we infer that that minute was mainly used for interactive applications. All remaining minutes of recorded traffic are classified as representing bursty traffic. This classification method is of course limited in so far as it is based on proxies. Nevertheless, it can be used to point out some interesting properties of usage patterns.

Figure 6 shows the distribution of the three activity types for each choice of bandwidth (8 to 128 kbps). The activity type is represented by a number: 1 stands for bulk traffic, 2 for burst traffic, and 3 for interactive traffic. The overall percentage of burst traffic ranges from 66 % to 85 %, depending on bandwidth choice. Burst traffic clearly is the dominant activity type. An interesting observation is that interactive traffic strongly increases at 128 kbps. This illustrates that the subjects value fast response times although they are effectively wasting purchased capacity. These results verify the findings of a survey of residential Internet users in the midwestern USA about high speed Internet access (Hoag 1997).

4 Acceptance of QoS on Demand

After illustrating the heterogeneity of users and activities in the previous section, we now examine whether users appreciate the higher flexibility of being able to make QoS decisions at any time.

Figure 7 depicts the QoS choices of all INDEX subjects. The weighted two-way scatter plot graphs each subject against all priced bandwidth choices made by that subject over the entire duration of the *Variable Symmetric Bandwidth* experiment. The sum of all minutes spent at the same bandwidth is graphed as one circle, with the radius of each circle being proportional to the share of total time spent online by that subject and at that bandwidth.

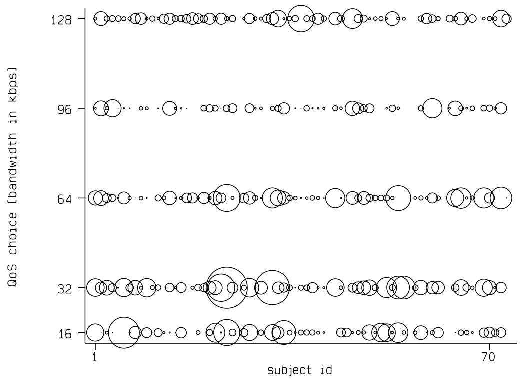

Figure 7: QoS Choices – Per Subject

The diversity of QoS choices pictured in this graph visualizes two important characteristics of user behavior: First, different users obviously have different valuations of network resources and consequently prefer different bandwidths. Second, QoS choices of individual users are not restricted to just one or two bandwidths. The subjects make use of a wide range of bandwidths. Almost all users purchase high quality at least sometimes. Figure 8 emphasizes this second characteristic even more. The histogram displays the number of different QoS choices (priced and unpriced) that were made over the entire duration of the experiment: 62.5 % of the subject population made use of the entire range of options (8kbps to 128kbps). 75 % used at least five different bandwidths.

These results make a strong case for persistently giving Internet users the choice of multiple service qualities. Our subjects clearly endorsed the higher flexibility associated with being able to make QoS decisions at any time. If different service qualities are only offered by means of subscribing to them under corresponding static flat-rate pricing regimes, then users who only occasionally demand high-quality services will be excluded from using such services. In contrast, if charges are set in proportion to actual usage, high-speed services are accessible for a much broader set of users. INDEX data suggests that the range of different quality choices will be fully exploited if the economic incentives to make use of these choices are not distorted by static service plans.

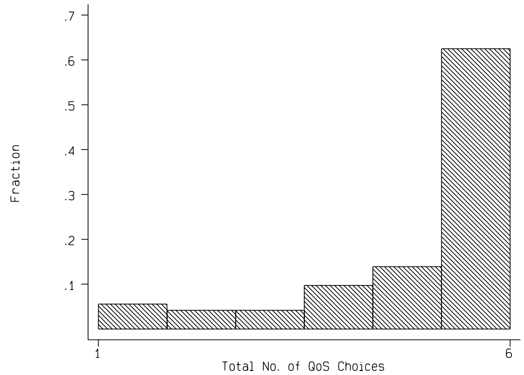

Figure 8: QoS Choices – Histogram

5 Implications for Service Provision and Tariff Design

The preceding analysis has shown: There is little intra-user variation in weekly mean expenditure; the inter-user distribution of individual budgets for Internet access and general usage intensity is very diverse; demand varies not only over time, but also depends on the activity type.

These findings lead us to conclude that user heterogeneity and activity heterogeneity call for a flexible system which enables users to choose Quality of Service on demand. Price discrimination that is based exclusively on customer type (i.e. different service plans for high-volume and low-volume users) is not sufficient to meet the demands of tomorrow's Internet users. In contrast, users should also be given the opportunity to switch between different service qualities more or less instantaneously. We have shown that users endorsed the higher flexibility associated with being able to make QoS decisions at any time. A wide range of QoS choices were made. In the *Variable Symmetric Bandwidth* experiment that we analyzed, 62.5 % of the subjects made use of *all* the QoS options that were offered to them.

At a general level, QoS on demand avoids waste and user cross-subsidy. Quality differentiation, combined with proper economic incentives, increases the overall value of a network and can pave the way towards an economically viable Integrated Services Internet.

References

[1] Hoag, Anne. 1997. Speed and the internet: The effects of high speed access on household usage. In: *Proceedings of the 25th Telecommunications Policy Research Conference*, section IX, 24-37.

[2] Rupp, Björn; Edell, Richard; Chand, Harish; Varaiya, Pravin. 1998. INDEX: A platform for determining how people value the quality of their internet access. In: *Proceedings of the 6th IEEE/IFIP International Workshop on Quality of Service*, 85-90.

[3] Shenker, Scott; Clark, David; Estrin, Deborah; Herzog, Shai. 1996. Pricing in computer networks: Reshaping the research agenda. *Telecommunications Policy*, 20 (3), 183-202.