

# A Model for Discovering Customer Value for E-Content

Srinivasan Jagannathan<sup>1</sup> Jayanth Nayak<sup>2</sup> Kevin Almeroth<sup>1</sup> Markus Hofmann<sup>3</sup>

Department of Computer Science<sup>1</sup>, Department of Electrical and Computer Engineering<sup>2</sup>  
University of California, Santa Barbara, CA 93106-5110

{jsrini@cs, jayanth@ece, almeroth@cs}.ucsb.edu

High-Speed Networks Research Department<sup>3</sup>  
Bell Laboratories, Holmdel, NJ 07733-3030

hofmann@bell-labs.com

## ABSTRACT

There exists a huge demand for multimedia goods and services in the Internet. Currently available bandwidth speeds can support sale of *downloadable content* like CDs, e-books, etc. as well as services like video-on-demand. In the future, such services will be prevalent in the Internet. Since costs are typically fixed, maximizing revenue can maximize profits. A primary determinant of revenue in such e-content markets is how much value the customers associate with the content. Though marketing surveys are useful, they cannot adapt to the dynamic nature of the Internet market. In this work, we examine how to *learn* customer valuations in close to real-time. Our contributions in this paper are threefold: (1) we develop a probabilistic model to describe customer behavior, (2) we develop a framework for pricing e-content based on basic economic principles, and (3) we propose a price discovering algorithm that learns customer behavior parameters and suggests prices to an e-content provider. We validate our algorithm using simulations. Our simulations indicate that our algorithm generates revenue close to the maximum expectation. Further, they also indicate that the algorithm is robust to transient customer behavior.

## 1. INTRODUCTION

The Internet is seeing an explosive growth in commercial activities, especially in content delivery services. The challenge is to envision realistic scenarios where customers can download music or movies for a price. For instance, Blockbuster could enter into an arrangement with the local cable modem or DSL services provider for on-demand movie rentals or music-video sales. Another example is how e-Books are being proposed as a viable alternative to paper books. In such markets, the content provider can maximize revenue by charging the customers as high an amount as they are willing to pay. The key to increasing revenues therefore lies in learning how much the customers are willing to pay for the content.

From the customer perspective, price affects whether to purchase the product or service. The customer can accept or reject a price based on his/her *capacity* to pay and *willingness* to do so. Choos-

ing the right price is therefore of great importance to maximize revenue. There are not many ways to choose the right price. Usually, content providers use years of experience with customers to guide their choice of prices. However, such approaches have some inadequacies in an Internet setting. First, these approaches are typically regional. Surveys can find customer valuations for a service in some geographical region. Such region based characterization may not be helpful in the Internet. For instance, it is not inconceivable that in the future, during prime-time, when servers in Europe are overloaded, customers in Europe can watch movies streamed by a content provider in North America. Prices based on North American market surveys (or based on experience with North American customers) would then severely affect revenues. Second, customer behavior can be time varying. A sudden news event can cause immediate changes in the Internet market. Experience-based prices and surveys cannot keep up with such dynamic behavior.

This paper develops an analytical framework for discovering the right price in an Internet setting. We answer two questions— (1) can the content provider learn customer valuations online? and (2) can the content provider adapt to changes in customer behavior? A simple approach to learn customer valuations is to charge different prices and observe how customers react. And by continuously observing how customers react, one can track changes in behavior. However, experimenting with different prices can result in loss of revenue. The challenge is to learn customer behavior with minimal loss in revenue due to experimentation. One can minimize loss in revenue if data acquired by making a small number of observations can be extrapolated based on a formal model of customer behavior. To this end, we develop a parameterized customer behavior model that takes into account how much a customer *can* spend as well as how *willing* she is to do so. We then develop an algorithm that ascertains these parameters online and then uses our analytical framework to suggest prices to the content provider. We perform simulations to validate our algorithm. We compare the revenue generated by our algorithm with the maximum expectation predicted by our analytical framework. We also compare the revenue generated by our algorithm with that generated using a simple pricing scheme. We simulate a variety of scenarios where customer behavior changes dynamically. The simulation results indicate that our algorithm is robust to such scenarios. Our work is based on a video-on-demand server, but it is sufficiently general to be applied to other forms of content and services in the Internet.

The rest of the paper is organized as follows. We describe our basic system model used in this paper in Section 2. We formulate the maximum expectation of revenue as a constrained optimization

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SIGKDD '02 Edmonton, Alberta, Canada

Copyright 2002 ACM 1-58113-567-X/02/0007 ...\$5.00.

problem in Section 3. In Section 4, we develop a price discovering algorithm that tracks customer behavior and suggests the price for the content. We perform simulations to validate and stress-test our algorithm in Section 5. We discuss related and motivational work in Section 6. We conclude the paper in Section 7.

## 2. SYSTEM AND CUSTOMER BEHAVIOR MODELS

We consider a system where requests are satisfied if resources are available and the customer agrees to pay the quoted price. Resources are modeled as *logical channels*. Every request which is satisfied occupies a channel for some finite amount of time. Requests are processed on a First-Come-First-Served basis. If there are no channels available when a request arrives, then the request is rejected. For a video-on-demand server we can think of the number of channels as the number of movies that can be served simultaneously. In this paper we do not focus on how the channel is allocated or how an allocated channel is managed. We mainly focus on the interaction between the system and the customer before a channel is allocated.

In our model, we assume that once the content provider makes the initial infrastructural investment, there are either negligible or fixed costs in maintaining the resources (caches, servers, bandwidth etc.), i.e., there are no additional costs based on number of requests served. This is a reasonable assumption because servers incur fixed costs and bandwidth is typically bought at a flat rate. If maintenance costs are negligible or fixed, profit maximization is equivalent to revenue maximization. We also assume that there is no distinction or classification in the type of service, i.e., all customers are provided the same quality of the product. We also assume that all the products/services being sold are similar. This implies that the content provider is selling only e-books or only CDs or a VoD service and not a combination of these. A content-provider who decides to combine these activities can pre-allocate some channels to each category. Our analysis would then hold true for each of these categories independently. Though our ultimate objective is develop learning algorithms for a competitive market, in this paper, we restrict ourselves to a monopolistic market where there is a single content provider providing the service.

Economic theory has established that there are a large number of customers with a small income and a very small number of customers with a very large income[2]. It is reasonable to assume that customers' capacities to spend will follow a similar behavior. Currently, two probability distribution models, *Pareto* and *log-normal*, are used to represent the distribution of incomes. In this paper, we use the Pareto distribution to represent the capacity to spend. Every customer has the capacity to pay based on a Pareto distribution with two parameters—shape  $\alpha$  and scale  $b$ . All customers have capacities at least as large as  $b$ . The shape  $\alpha$  determines how the capacities are distributed. The larger the value of  $\alpha$ , the fewer the people with a very large capacity to pay. The Pareto density function is defined as  $f_{\varphi}(x) = \frac{\alpha b^{\alpha}}{x^{\alpha+1}}$ , for  $x \geq b$ .

Figure 1 illustrates the Pareto density function for different values of shape  $\alpha$ , and scale  $b = 50$ . Let us consider an illustrative example to understand the Pareto distribution of capacities. Consider a video-on-demand server. We can expect all customers to have a capacity to pay at least some amount for the movie. We call the largest such amount that can be paid by all the customers as the scale of the distribution of their capacities and denote it as  $b$ . We would expect most of the customers to be able to pay only about this amount. There will be very few customers who can pay a significantly more than the scale. This information is captured by the

shape of the distribution, which we denote as  $\alpha$ . The greater the value of  $\alpha$ , the fewer the customers who can pay a lot more than  $b$ . For systems like video-on-demand servers, we would expect the shape to be very large. Another interpretation of the capacity to spend is to assume that it is the maximum valuation for that product. For common products, like movies or CDs, it is very likely that most people will have a similar maximum valuation and very few may have a higher maximum valuation. The Pareto distribution captures this behavior. Henceforth, we shall use the terms *capacity to pay* and *maximum valuation* interchangeably.

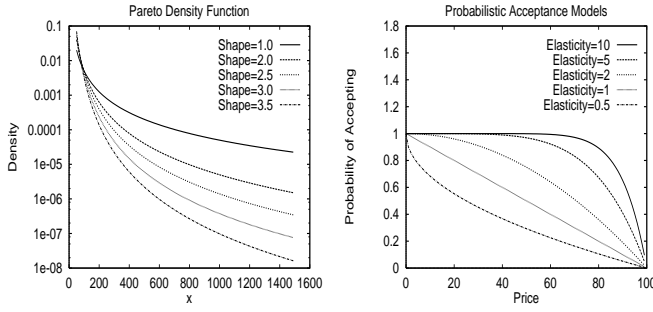
Even though customers *can* spend (i.e., have a maximum valuation of the product), they may not be *willing* to do so<sup>1</sup>. For instance, even though a customer might value watching a movie at \$10, how much he/she actually pays depends on the movie and other intangibles like “mood”. To adequately describe the willingness of customers to pay, we define a family of probability functions. Consider an arbitrary customer with capacity  $\chi$ . We denote his/her decision to purchase the service by the random variable  $\Upsilon$  which can take two values, 1 for accept and 0 for reject. The probability that the customer accepts the price  $\psi$ , denoted by  $P\{\Upsilon = 1 \mid \psi\}$  depends on his/her capacity (or maximum valuation)  $\chi$ , and the price  $\psi$ . In general, for rational customers, we would expect this probability to decrease with price and increase with capacity. In this paper, we work with a simple model, where  $P\{\Upsilon = 1 \mid \psi\}$  is defined as shown in Equation 1.

$$P\{\Upsilon = 1 \mid \psi\} = \begin{cases} 1 - \left(\frac{\psi}{\chi}\right)^{\delta} & , 0 \leq \psi \leq \chi \\ 0 & , \psi > \chi \end{cases} \quad (1)$$

The parameter  $\delta$  in Equation 1 is a product-specific variable that represents how willing a customer is to spend money. By varying the parameter  $\delta$ , we can make the willingness as *elastic* as desired. The higher the value of  $\delta$ , the more willing the customer is to spend money on that product. We show willingness models for a customer having capacity 100, with  $\delta$  values 0.5, 1, 2, 5 and 10 respectively in Figure 1. As can be seen, the model with  $\delta = 10$  makes the customer much more willing to spend money than in the case of the other models. In fact, as  $\delta$  increases, the willingness begins to resemble a “step-function”, i.e., the customer is willing to pay as much as his/her capacity to pay. In particular, when  $\delta$  is infinity, the model reduces to a quasi-linear utility function model, where customers will purchase at any price below their maximum valuation. Since different customers will have different degrees of willingness, we shall assume that for each product, the willingness parameter  $\delta$  of the customers is a random variable. While in reality, the customers' willingness can be totally arbitrary, to make our analysis tractable, we shall assume that it is exponentially distributed. Let the mean elasticity of willingness be  $\delta_0$ . Then, the probability density of  $\delta$  at  $x$ , denoted by  $f_{\Delta}(x)$ , is given by  $f_{\Delta}(x) = \frac{1}{\delta_0} e^{-\frac{x}{\delta_0}}$  for  $x \geq 0$

## 3. EXPECTATIONS OF REVENUE AND ACCEPTANCE

<sup>1</sup>Typically, the willingness of customers is modeled in terms of utility functions. This presupposes that the customer has associated a value with the good or service and that this value is known. In reality, the actual value a customer associates with the product is difficult to determine. We therefore propose a model where the customer probabilistically *reacts* to the quoted price.



**Figure 1: Pareto Distribution and Probabilistic User Willingness**

In this section, we discuss the dynamics of the customer behavior model and system resource constraints and how together they affect revenue. We state the following theorem without proof<sup>2</sup>.

**THEOREM 1.** Consider product  $p_i$ . Let customer willingness to pay for  $p_i$  be represented by Equation 1. For a Pareto distribution of customer capacities, with shape  $\alpha$  and scale  $b$ ,  $\alpha \geq 1$ ,  $b > 0$ , and exponentially distributed elasticity of willingness with mean  $\delta_{i0}$ , the expectation of the variable  $\Upsilon_i$  given price  $\psi_i$ ,  $E[\Upsilon_i | \psi_i]$  is as follows. In Equation 2,  $\Gamma(p, q)$  refers to the incomplete gamma function  $\int_q^\infty \frac{e^{-x}}{x^{1+p}} dx$ .

$$E[\Upsilon_i | \psi_i] = \begin{cases} 1 - \frac{\alpha}{\delta_{i0}} e^{\frac{\alpha}{\delta_{i0}}} \left(\frac{b}{\psi_i}\right)^\alpha \Gamma\left(0, \frac{\alpha}{\delta_{i0}} + \alpha \ln\left(\frac{b}{\psi_i}\right)\right), & 0 \leq \psi_i \leq b \\ \left(\frac{b}{\psi_i}\right)^\alpha \left(1 - \frac{\alpha}{\delta_{i0}} e^{\frac{\alpha}{\delta_{i0}}} \Gamma\left(0, \frac{\alpha}{\delta_{i0}}\right)\right), & \psi_i > b \end{cases} \quad (2)$$

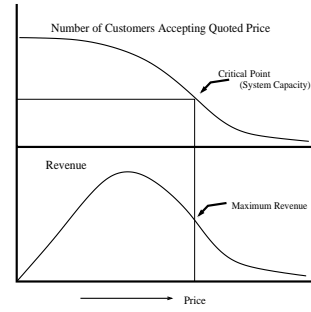
We now formulate an expression for the revenue. Let  $m$  be the total number of products and  $\lambda_i$  the arrival rate of requests for product  $p_i$ . The fraction of customers who accept the quoted price  $\psi_i$  for  $p_i$  per unit time is  $\lambda_i E[\Upsilon_i | \psi_i]$ . Therefore the expected revenue earned on  $p_i$  per unit time is given by  $\lambda_i \psi_i E[\Upsilon_i | \psi_i]$ . The expectation of revenue per unit time for all the products together is:  $\mathcal{R} = \sum_{i=1}^m \lambda_i \psi_i E[\Upsilon_i | \psi_i]$

One can show that there exists a maximum for the equation defined above. To do so we observe that each term in the expression for revenue is dependent only on one variable. Therefore, if each of the individual terms can be maximized, the sum can also be maximized. Moreover, the  $i^{\text{th}}$  term is a concave function<sup>3</sup> of the variable  $\psi_i$ . Therefore each term can be maximized and any local maxima is also a global maxima. Since the revenue function is well behaved, a good numerical package can be used to determine the optimal prices.

Though the optimal prices can be ascertained through numerical methods, resource constraints may prevent achieving the maximum expectation of revenue. To illustrate this phenomenon, let us consider a hypothetical system with one single product, 10 channels, and an arrival rate of 20 requests per minute. Further, assume that at the optimal price, the acceptance rate is 0.6. In effect, if the content provider quotes the optimal price, there will be 12 customers accepting the price. This clearly leads to an unstable system since only 10 requests can be accommodated. On the other hand, by

<sup>2</sup>The proof outline is presented in a detailed version of this work [13].

<sup>3</sup>We have omitted the proof for reasons of space.



**Figure 2: Maximizing Revenue Under Resource Constraints**

slightly increasing the price, the content provider can ensure that only as many customers accept the price as can be accommodated by the system. This is illustrated in Figure 2. In this figure, the revenue and the number of customers accepting the price are plotted as a function of price. At the price when revenue is maximized, the number of customers accepting the price is more than capacity of the system. By increasing the price to such an extent that only as many customers accept as can be accommodated by the system, maximum revenue is realized.

We formalize this idea by introducing the constraint of system utilization. System utilization is the relative fraction of time for which the channels are busy servicing requests. Let  $\lambda = \sum \lambda_i$  be the arrival rate of requests,  $n$  the number of channels, and the  $d$  the average time to serve a request<sup>4</sup>. Then, the system utilization,  $\rho$ , is defined as the ratio of the number of requests entering the system per unit time to the number of serviced requests exiting the system per unit time. The mathematical expression for system utilization, when we charge a price  $\psi_i$  for product  $p_i$ , is given by:  $\rho(\psi_1, \dots, \psi_m) = \frac{d}{n} \sum_{i=1}^m \lambda_i E[\Upsilon_i | \psi_i]$

For a stable system, the system utilization is bounded above by 1. This imposes a constraint on the revenue maximization problem. Combining the expressions for revenue and system utilization and the fact that prices should be positive, we formulate revenue maximization as a constrained optimization problem:

$$\text{Maximize : } \mathcal{R} = \sum_{i=1}^m \lambda_i \psi_i E[\Upsilon_i | \psi_i] \quad (3)$$

$$\frac{d}{n} \sum_{i=1}^m \lambda_i E[\Upsilon_i | \psi_i] \leq 1 \quad (4)$$

$$\forall i, 1 \leq i \leq m, \psi_i > 0 \quad (5)$$

## 4. PRICE DISCOVERING ALGORITHM

In the previous section, we formulated revenue maximization as a constrained optimization problem. To solve the optimization problem, the content provider would need to know the shape, scale, and mean customer willingness elasticity for all the products. In this section, we develop an algorithm that learns these parameters in real-time and suggests prices to the content provider.

Our algorithm is based on gauging the customers' reaction to different prices. Consider product  $p_i$ . By observing customers' decisions over a period of time for a trial price  $\psi_i$ , one can ascertain the mean rate of acceptance for that price. This observed rate is assumed to be the expectation of the decision to purchase, i.e.,  $E[\Upsilon_i | \psi_i]$ . Theorem 1, which quantifies  $E[\Upsilon_i | \psi_i]$ , depends on four variables: price ( $\psi_i$ ), shape ( $\alpha$ ), scale ( $b$ ), and mean customer

<sup>4</sup>We make the assumption that all products have similar service time. For other systems, only the formulation for system utilization will change. The constraint remains the same.

willingness elasticity ( $\delta_{i0}$ ). Of these we only know price. We assume a set of feasible pairs of values for shape and mean customer willingness. For each of these feasible pairs  $((\alpha, \delta))$ , we solve for scale using Equation 2. Thus, we now have a feasible set of values for the scale of the customer capacity distribution. We now set another trial price  $\psi_i^t$ . For each feasible tuple of shape, scale and mean willingness elasticity, we compute the expectation of the decision to purchase for price  $\psi_i^t$ . We then make a second round of observations of customer decisions, this time for the price  $\psi_i^t$ . By comparing actual customer reactions with the predicted reaction, we can estimate the closest feasible tuple for shape, scale and mean customer willingness elasticity. We can use these values in the constrained optimization problem to compute the optimal price. This new price is again used as a trial price and this process is repeated forever. Since there are continuous price experiments, any changes in customer behavior will be detected by the algorithm.

There are a few problems in the algorithm outlined above. First, customer preferences for products can be expected to follow a Zipf distribution [16]. This means that request arrival rates will be highly disproportionate. Therefore, the interval for observation to estimate the rate of acceptance could be very large. Because the trial price can be suboptimal, the greater the time spent charging the trial price, the greater the loss of revenue. Second, if the number of products is very large, the optimization problem can be fairly difficult to solve.

We therefore classify the set of products into categories. Each category consists of a set of products for which the content provider estimates that the mean customer willingness to pay is “similar”. The customer will be charged the same price for all products belonging to the same category. For instance, in a VoD system, the content provider can classify movies as *Hot*, *Cold* and *Special*. Hot movies are those that the content provider thinks are very popular. Cold movies are those for which customer may not pay a lot. Special movies are those that have a limited appeal, but those who prefer it have a high willingness to pay. Alternately, the content provider can classify the movies as *Action*, *Western*, *Romance*, etc. We believe that content providers can make classifications based on customer willingness to pay with a reasonable degree of accuracy. Since every product in a class is charged the same price, each class of products can be thought of as a single meta-product. There are two advantages of working with a set of meta-products: (1) the revenue optimization problem is more tractable since there are fewer variables, and (2) the intervals of observation for estimating the rate of acceptance are much smaller.

Another assumption we make is that the content provider sets a minimum price and a maximum price. Any prices suggested by the algorithm that are not in this range are ignored, and the nearest bound is set as the new price. Such bounds represent policy decisions that may be made by content-providers in real-life.

The modified algorithm makes a *round* of observations of customer reactions for each product. A round of observations is complete as soon as there are  $k_{req}$  requests for any of the meta-products. It is quite possible that one of the meta-products has received very few requests when the round is over. In this case, since the arrival rate for that meta-product is small, an erroneous estimate of the acceptance rate will not affect the revenue maximization. The specifics of the algorithm are outlined in a detailed version of this work [13].

In spite of the very large number of unknowns, our algorithm is able to produce consistently high revenues (when compared with the maximum expectation) in our simulations. We discuss our simulation scenarios in greater detail in the next section.

## 5. SIMULATION

We have implemented a simulator to model a content delivery system. Using simulations, we analyze the performance of the price discovering algorithm described in Section 4 and compare its performance with two other pricing schemes<sup>5</sup>. We describe the simulation scenario below.

**System Description:** We performed simulations with 1000 logical channels. We chose a fixed number of channels because the system capacity typically does not change very often. We chose request service times from a uniform distribution between 90 and 110 minutes. This closely models the typical length of movies. Channels were allocated based on a FCFS policy. Requests arriving when there are no free channels are rejected. There is no waiting queue.

**Customer Choice of Products:** In all our simulations we assume that there are 50 products for the customer to choose from. Customer choice of the products was assumed to follow a Zipf-like distribution with zipf-exponent<sup>6</sup>,  $\theta = 0.73$ . In a Zipf-like distribution, the  $i^{th}$  popular product in a group of  $m$  products is requested with probability  $\frac{1}{\sum_{j=1}^m \frac{1}{j^\theta}}$ .

**Customer Capacity and Willingness:** Since our system delivers only *similar* kinds of products<sup>7</sup>, we assume that for any given customer, his/her maximum valuation (or capacity to pay) for all the products is the same. Only the willingness to pay differs over products. A high request arrival rate for a specific product can be directly correlated to a high willingness to pay for that product. Alternately, there may exist no correlation between popularity and willingness to pay. We simulate both scenarios. In the first scenario, the mean customer willingness for the  $i^{th}$  most popular product was chosen to be  $100 \times \frac{1}{\sum_{j=1}^m \frac{1}{j^\omega}}$ , where  $\omega = 0.73$ . In the second scenario, the products were ranked according to customer willingness to pay. The mean customer willingness of the  $i^{th}$  ranked product was chosen in a similar way as in the first scenario. The set of mean customer willingness values so generated is highly skewed, with very few products having a high mean customer willingness and most other products having a low mean customer willingness.

**Request Arrival Process:** We simulated a variety of arrival-rate models, adapted from the work on arrival-rate based scheduling by Almeroth et al. [1]. The workloads, presented in Figure 5, are modeled based on a 24 hour period beginning from 8.00am of one day and running to 8.00am of the next. “Prime time” periods see a surge in demand. We have used a steady baseline workload, with no surges in demand, and three non-steady workloads. The arrival rates during prime time for the non-steady workloads was around five times greater than the normal rate—based on statistics reported by Little and Venkatesh [14]. We simulated both gradual as well as sudden increases in arrival rate. We also used a workload with hourly spikes in arrival rate during primetime. This type of workload is based on the belief that the workload for some systems may be synchronized with an external event like wall-clock time.

**Pricing Policies:** We assume that the content-provider will charge at least \$1 and not more than \$10 for serving the content. In all our simulations, we assume that our unit of currency is a dime (10

<sup>5</sup>While it would be interesting to use real data to validate our model, such data does not exist. The reason is that customers will seldom disclose their true valuations for products.

<sup>6</sup>Web-page accesses have been observed to obey a Zipf-like distribution with zip-exponent in the range 0.64 to 0.83 [5].

<sup>7</sup>For instance, on a VoD server, only movies are streamed and we can assume that a customer has a single maximum valuation for movies, even though his willingness to pay for specific movies is different.

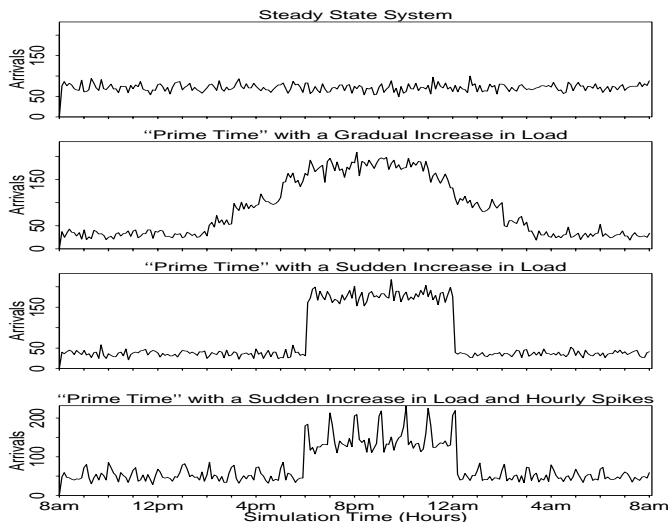


Figure 3: Workloads

dimes = \$1). To better understand the performance of our price discovering algorithm we compared its revenues with those of two other pricing policies: 1) optimal pricing and 2) fixed pricing. The optimal pricing policy is based on a *prescient* algorithm that knows all the parameters of the system. Using all these parameters, it computes the optimal price for each product using the constrained optimization package. The optimal pricing policy is clearly unrealistic because it is impossible to know all the customer behavior parameters. It is therefore an upper bound on the achievable revenue for a price discovering algorithm. The fixed pricing policy on the other hand is simplistic and practical. The content provider chooses some fixed price and charges all customers the same price irrespective of the product requested. The problem with this policy though is that the content provider does not know what price to charge. For the purposes of comparison, we repeated the simulations with different prices, starting with a price of 10 and going up to 100 in increments of 1.5. Assuming that all these prices are equally likely, we computed the mean revenue earned in the simulations for each of these prices. To get a better idea of how a particular fixed price impacts overall revenue, we also present results for simulations with prices 20, 40, 60 and 80. The fixed pricing scheme we outlined above is a representative of the class of fixed pricing algorithms. We do not claim that our fixed pricing scheme is the best representative of that class. It is not the aim of this paper to benchmark our algorithm with respect to the best fixed pricing algorithm. Instead, our objective is to understand the flexibility that a price discovering pricing scheme provides in adjusting to customer behavior and resource constraints.

## 5.1 Simulation Results

We ran simulations for a period of 1440 minutes (one day of simulated time). All the simulation results are averaged over 10 runs with different seed values for the random number generator. In case of the price discovering algorithm, in all our simulations, the products are classified into three categories. The same price is charged for all products in the same category. The metric we use in our simulations is called the *revenue-ratio*. The revenue-ratio of a pricing scheme for a particular simulation is the ratio of the revenue generated by that pricing scheme to the theoretically predicted maximum expectation for that simulation. The higher the revenue-ratio, the better the performance of the pricing scheme. Since the revenue-ratio is with respect to an expectation, it is possible for the

revenue-ratio to be greater than 1. Moreover, since the maximum expectation is computed using numerical methods, our results are accurate only within error bounds.

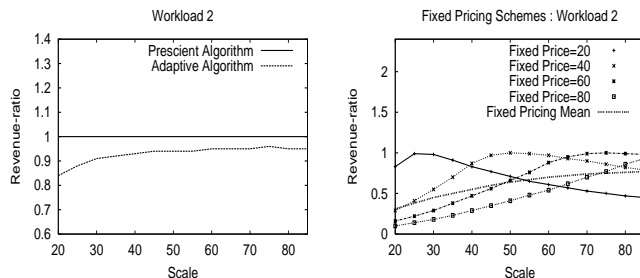
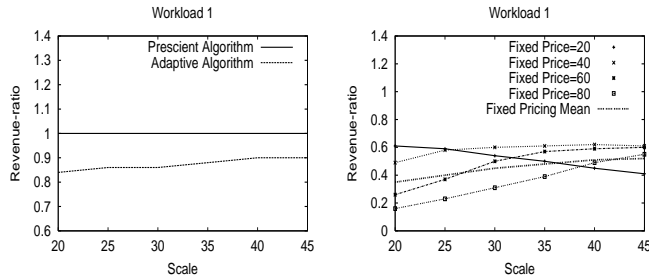


Figure 4: Revenue relative to maximum expectation for different pricing policies

Figure 4 shows the performance of the three pricing schemes for different customer capacities. The revenue-ratio is presented as a function of the scale of the Pareto distribution. The scale of the Pareto distribution is varied from 20 to 85. The results for the fixed pricing schemes are presented in a separate graph to improve readability. The revenue-ratio for the different fixed prices varies with the scale of the distribution. This indicates that, if the content provider does not know the customer behavior parameters, the revenue earned can be suboptimal. Notice that the mean revenue of the fixed pricing schemes increases with the scale. This is because as the scale increases, more customers could pay the prices charged in our simulations. The price discovering algorithm generates revenue within 80% of the optimal. The revenue generated by the price discovering algorithm is low for the lower values of scale because in all our simulations, we started with initial test price vectors of  $(30, 35, 35)$  and  $(50, 55, 55)$  respectively. Since these prices are much higher compared to the scale of the Pareto distribution, there is an initial loss in revenue. The prescient algorithm generates nearly the same revenue as the predicted maximum expectation. This validates our theoretical results. In this set of simulations, there is no correlation between popularity of a product and the mean customer willingness to pay for it. The results for the other scenario, where the popularity and willingness to pay are correlated, are very similar. We have presented results only for one of the workloads due to reasons of space. The revenue-ratios were similar for the other workloads. However, the actual revenue earned was higher in case of workloads with high request arrival rates.

In the simulations presented above, the customer behavior was invariant over time. But this may not be true in a real-world setting. For instance, one may expect a greater willingness to pay during “prime-time” than say, during “office-hours”. We performed many simulations to see how these behavioral changes may impact revenue. We present results for two scenarios—one where the scale of customer distribution increases during prime time and the other where the elasticity of willingness changes randomly during prime time. The results are presented in Figure 5. We also investigated how performance of the price discovering algorithm deteriorates when the frequency of customer behavior changes increases. For a good algorithm, this deterioration should be smooth and if possible negligible. In our simulations, we observed that the algorithm was able to adapt to changes in customer behavior. In fact, the performance (with respect to the optimal) does not deteriorate sig-

### Performance when scale is doubled



### Performance when mean willingness is changed randomly

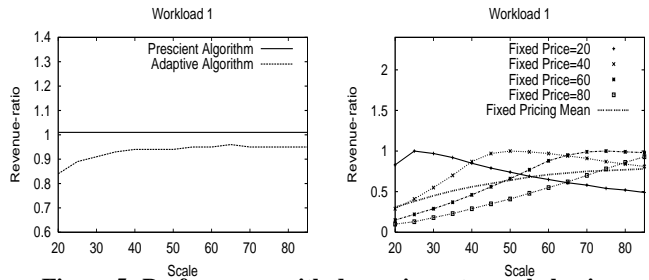


Figure 5: Performance with dynamic customer behavior

nificantly even when customer behavior changes after every hour of simulated time. We do not present the results for reasons of space<sup>8</sup>.

## 6. RELATED WORK

Our work has been motivated by the significant amount of research on bandwidth and congestion pricing. Dynamic pricing schemes [15, 7, 10] have been shown to be socially and economically efficient. The main thrust of such schemes has been to increase prices when resources are scarce. Such research has been “resource-centric”, where the primary focus has been on managing resources and not on revenue. Our work has also been motivated by research which has been “content-centric” [8, 4], and not concerned with resource constraints. Ignoring resource constraints can however lead to a situation where customers are rejected due to lack of server or bandwidth resources. We believe that our work strikes a balance between resource considerations and customers’ value for the content.

To the best of our knowledge there has been very little research on online price discovering algorithms [11, 12]. Most research has been focussed on agent based market economies [3, 6, 9].

## 7. CONCLUSIONS AND FUTURE WORK

We have developed an analytical framework for pricing of on-demand content. The framework models customer behavior as well as resource constraints. Based on this framework, we have developed an algorithm that suggests prices to the content-provider. We have analyzed the performance of our algorithm using simulations. We have used a simple fixed pricing scheme and an unrealistic optimal pricing scheme as a baseline for comparison. We have observed that the fixed pricing scheme can generate sub-optimal revenues if parameters governing customer behavior are not known. We have also observed that the relative performance of the pricing schemes is independent of the request arrival rate. We also observed that the fixed pricing scheme breaks down in the presence of dynamic customer behavior. Finally, we have observed that our

price discovering scheme performs consistently even with highly varying workloads and dynamic customer behavior.

Our work models a monopolistic market. Competition will affect customer behavior and hence the maximum expectation of revenue. Learning customer behavior in competitive markets needs to be studied in greater detail. Another avenue for research is to study the tradeoffs between a dynamic pricing scheme and a subscription-based pricing scheme and how the two can be integrated.

## 8. REFERENCES

- [1] K. Almeroth, A. Dan, D. Sitaram, and W. Tetzlaff. Long term channel allocation strategies for video applications. In *IEEE Infocom*, April 1997.
- [2] B. C. Arnold. *Pareto Distributions*. International Co-operative Publishing House, Burtonsville, Maryland, 1983.
- [3] J. Y. Bakos. Reducing buyer search costs: Implications for electronic marketplaces. *Management Science*, 43(12), December 1997.
- [4] P. Basu and T. Little. Pricing considerations in video-on-demand systems. In *ACM Multimedia Conference*, November 2000.
- [5] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker. Web caching and zipf-like distributions: Evidence and implications. In *Infocom*, pages 126–134, 1999.
- [6] A. Chavez and P. Maes. Kasbah: an agent marketplace for buying and selling goods. In *Proceedings of the First International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology*, April 1996.
- [7] R. J. Gibbens and F. P. Kelly. Resource pricing and the evolution of congestion control. *Automatica*, 35:1969–1985, 1999.
- [8] A. Goldberg, J. Hartline, and A. Wright. Competitive auctions and digital goods. Technical Report STAR-TR-99-01, InterTrust Technologies Corporation, November 2000.
- [9] A. Greenwald and J. Kephart. Shoptbots and pricebots. In *Sixteenth International Joint Conference on Artificial Intelligence*, August 1999.
- [10] A. Gupta, D. O. Stahl, and A. B. Whinston. The economics of network management. *Communications of the ACM*, 42(9):57–63, 1999.
- [11] S. Jagannathan and K. C. Almeroth. The dynamics of price, revenue and system utilization. In *Management of Multimedia Networks and Services*, Chicago, Illinois, USA, October 2001.
- [12] S. Jagannathan and K. C. Almeroth. Price issues in delivering e-content on-demand. *ACM Sigecom Exchanges*, 3(2), May 2002.
- [13] S. Jagannathan, J. Nayak, K. Almeroth, and M. Hofmann. E-content pricing: Analysis and simulation. Technical report, University of California Santa Barbara, November 2001. available at <http://www.nmsl.cs.ucsb.edu/papers/ECONTENTPRC.ps.gz>.
- [14] T. Little and D. Venkatesh. Prospects for interactive video-on-demand. *IEEE Multimedia*, pages 14–23, Fall 1994.
- [15] J. Mackie-Mason and H. Varian. Pricing the Internet. *Public Access to the Internet*, February 1994.
- [16] G. Zipf. *Human Behavior and the Principle of Least Effort, an Introduction to Human Ecology*. Addison-Wesley, 1949.

<sup>8</sup>Results are presented in a detailed version of this work [13].