Pricing Considerations in Video-on-Demand Systems[1]

P. Basu and T.D.C. Little

Department of Electrical and Computer Engineering
Boston University, Boston, Massachusetts 02215, USA
(617) 353-9877
tdcl@bu.edu

MCL Technical Report No. 10-09-2000

Abstract– Video-on-demand (VoD) has been an active area of research for the past few years in the multimedia research community. However, there have not been many significant commercial deployments of VoD owing to the inadequacy of *per user* bandwidth and the lack of a good business model. Several VoD field trials have been conducted in the US and elsewhere [5], but most of them have, so far, been reported to be unsuccessful from a business point of view. Significant research efforts have been directed towards reduction of network bandwidth requirements, improvement of server utilization, and minimization of start-up latency. In this paper, we investigate another aspect of VoD systems which has been largely neglected by the research community, namely, pricing models for VoD systems. We believe that the price charged to a user for an on-demand video stream should influence the rate of user arrivals into the VoD system and in turn should depend upon quality-of-service (QoS) factors such as initial start-up latency. We briefly describe some simple pricing models and analyze the tradeoffs involved in such scenarios from a profit maximization point of view. We further explore secondary content insertion (ad-insertion) which was proposed elsewhere [1] not only as a technique for reducing the resource requirements at the server and the network, but also as a means of subsidizing VoD content to the end user. We treat the rate of ad insertion as another QoS factor and demonstrate how it can influence the price of movie delivery.

Keywords: Video-on-demand, pricing models, service aggregation.

[1] *Proc. 8th ACM Intl. Multimedia Conf. 2000*, Los Angeles, CA November 2000. This work is supported by the NSF under grant No. NCR-9523958. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

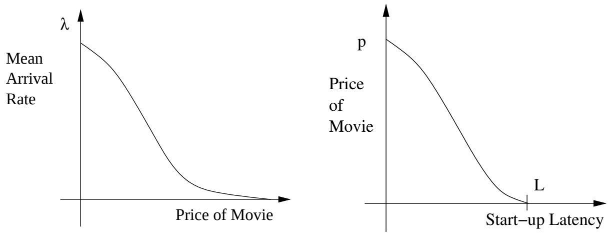
1 Introduction

With the explosive growth of the Internet and broadband cable networks in the mid- and late nineties, interactive video-on-demand (VoD) had been touted to be one of the most promising future applications for such networks. Unfortunately, inadequacy of bandwidth for serving a significant user population has stymied the growth in deployment of VoD on a large scale. Hence, lot of research efforts have been directed towards finding a scalable solution to the bandwidth problem using different techniques for serving *aggregates* of users [2, 3, 4]. These techniques use CATV broadcast or IP multicast as dissemination mechanisms. One aspect of VoD systems that has been largely neglected by the research community is pricing of VoD services. In this paper we briefly investigate several factors that can affect the price of a video stream delivered to the user, and also analyze some economic tradeoffs that exist in that context. Such an analysis can help in establishing a sound economic basis for a large scale deployment of VoD.

We develop models for pricing three different types of VoD service. First, in Sec. 2, we start with a simple interactive VoD system which provides each user with a dedicated channel and show how an optimal price can be calculated for maximization of profit. In Sec. 3, we describe how a QoS factor like start-up latency can affect the price of an on-demand movie in a staggered broadcast delivery system, and then we again show how profits can be maximized in such a setup. In Sec. 4, we describe how ads can be used to subsidize VoD to users, and speculate how varying degrees of ad insertion can affect the price and the interest of the users. We also demonstrate how profits can be maximized in such a scenario. Sec. 5 concludes the paper and gives some directions of future work.

2 Optimal Price Selection

A VoD system consists of a video server which can serve users on N channels at any point of time. Users arrive into the system with an average rate of λ per unit time. For simplicity, suppose that the VoD system is not aggregation based, i.e., each channel supports only one user, because the service is fully interactive. Suppose that the movie is L time units long, and that the users remain in the system for some more video time $T_{int} > 0$, on average, due to pauses and rewinds. (A large number of fast forward interactions may result in T_{int} being negative, but that is unlikely in a paid service.) We can treat the system as an $M/G/N$ queue, where M stands for Poisson arrivals, G stands for a general service time which has a



(a) λ vs. Price (b) Price vs. Latency

Figure 1: Interdependence of Parameters

deterministic part L and a stochastic part T_{int} ($T_s = L + T_{int}$), and N stands for the number of servers. For stability of the queue, i.e., for bounded waiting times, the condition $\lambda T_s < N$ must hold, hence N should be chosen properly.

Now suppose each user is charged a price p for viewing a movie. Intuitively, the mean arrival rate λ will depend on the price of the movie as depicted by Fig. 1(a). However, the exact function $\lambda(p)$ can only be known from marketing experiments such as user polls and surveys. Suppose the VoD service provider (VoDSP) incurs a cost b per channel per unit time. Therefore in time T, the cost incurred will be NbT. In that time the VoDSP will receive λT user requests. If no user is rejected (by bounding of waiting time due to appropriate selection of N), the revenue earned will be $p\lambda(p)T$. The profit per unit time is given by $P = p\lambda(p) - Nb$. Now, we have an optimization problem at hand given by:

Maximize: $P(p, N) = p\lambda(p) - Nb$

Subject to: (1) $p > 0$ and (2) $N > \lambda(p)T_s$

Optimality is achieved when the second constraint has an equality, and the optimal price p^* is given by the solution to the the equation: $(bT_s - p^*)\lambda'(p^*) = \lambda(p^*)$. We should point out here that there is an issue with the above analysis: when the second constraint is relaxed to an equality, the variability in arrivals may result in larger waiting times, and hence the assumption that the waiting times are low and do not affect user arrival, may fail. However, by over-engineering the system (i.e., by allocating a larger N than predicted by the optimal solution), we can approach a near-optimal solution. Also, if the function $P(p, N)$ is not *convex*, then first derivative techniques may be inadequate to solve the optimization problem. However, since the function is bounded, it will have a maxima, and that can be found using other more sophisticated (perhaps numerical) techniques.

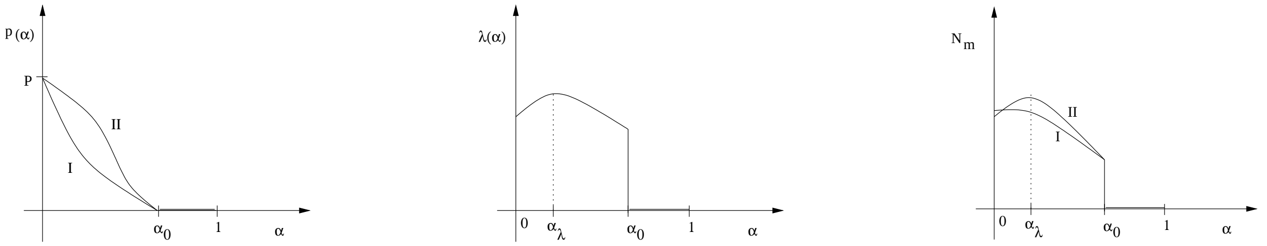
3 Broadcast Delivery Systems

Recently, staggered broadcast delivery systems have become popular with researchers in this area since they can be used to provide VoD service over CATV networks to a large user population with a *constant* number of “staggered” broadcast channels per movie. The basic idea is to divide the whole movie into a constant number of smaller segments and then broadcast each smaller segment repeatedly on a different channel. The client listens on one or more of these channels for downloading the video data before consumption. The simplest scheme which does not need any buffering on the clients is to distribute a movie of length L into an equal number of parts. In this case, the worst case start-up latency, t_{sl} is inversely proportional to the total number of broadcast channels, n ($n = \frac{L}{t_{sl}}$). The worst case start-up latency is a QoS parameter which can affect the price that is charged to a user by the VoDSP. We show by simple analysis how this parameter can be adjusted by the VoDSP for maximizing their profits. More sophisticated broadcast schemes like Skyscraper Broadcast [4] can reduce the startup latency by dividing the movie into segments of increasing size and by intelligent use of client buffering. However, the basic pricing model is similar in that case as n can be expressed as a function of t_{sl}.

As in the previous section, let us assume that the VoDSP incurs a cost b per video stream per unit time. Since they allocate $n = \frac{L}{t_{sl}}$ streams for a movie, the total cost incurred by the VoDSP in time T is $\frac{L}{t_{sl}}bT$. Suppose the mean arrival rate into the system is λ per unit time. Hence the total number of users that have arrived in the system in time T is λT. Suppose the price p charged to a user varies with t_{sl}. We depict the price by some function $p(t_{sl})$ which intuitively should be a decreasing function of t_{sl} (a possible function is depicted graphically in Fig. 1(b)). As in Sec. 2, the exact shape of this curve too can be characterized by marketing techniques such as polls and surveys. The revenue earned by the VoDSP in time T is $\lambda T p(t_{sl})$, and the profit per unit time, P is given by $P = \lambda p(t_{sl}) - \frac{L}{t_{sl}}b$. P is maximized at $t_{sl} = t_{opt}$ when $P'(t_{sl}) = 0$, i.e., $\lambda p'(t_{opt}) + \frac{L}{t_{opt}^2}b = 0$. If the equation has multiple roots, then the VoDSP can choose either of those operating points depending upon other factors like viewer utility, which we do not consider in this paper.

4 Ad-insertion Based Systems

Secondary content insertion (or ad insertion) has been proposed for reduction in server and network bandwidth requirements, and for subsidizing the cost of VoD to the end user [1].



(a) Price

(b) Arrival Rate

(c) Number of channels

Figure 2: Effect of Ad-Ratio

First, we consider ads only as a means of subsidizing the cost to the user. We consider a scenario where all users are uniformly shown ads at a rate α $(0 \leq \alpha \leq 1)$.($\alpha = \frac{A_{max}}{A_{max}+V_{min}}$, where A_{max} is the maximum continuous ad time and V_{min} is the minimum video time between ads.) Intuitively, the price of a movie, p to a user should be a decreasing function of α. If $\alpha \geq \alpha_0$, the movie is streamed to the users for free. Price p should be *maximum* when no ads are shown, i.e., $\alpha = 0$. An interesting topic to study is how the price varies with α in the interval $[0, \alpha_0]$. Two possible curves have been shown as I and II in Fig. 2(a).

Initially, let us suppose that the mean arrival rate λ is constant, and hence the number of movie channels needed to support that rate (N) is also constant. In this case, the VoDSP has two sources of revenue, namely, a broadcast advertisement channel, and the price of movies that it charges its users. If a video server can stream out N streams concurrently,(For bounded waiting time: $\lambda L < N$.) and if b is the cost per stream per unit time, the cost incurred in time T is NbT. If the VoDSP charges a per unit time to the advertisers, the revenue earned from the ad channel in time T is aT. Since the advertisers are willing to pay more money if they know that more people are watching ads at any point of time, a can be represented as an increasing function of α. With respect to the above model, the profit per unit time is given by $P = \lambda p(\alpha) + a(\alpha) - Nb$, and it is maximized at $\alpha = \alpha^*$ which satisfies the equation $\lambda p'(\alpha^*) + a'(\alpha^*) = 0$.

Now we consider the case when the mean arrival rate may be influenced by the price p and the quality of service factor α. Fig. 2(b) shows a tentative schematic of the effect of α on the arrival rate of users into the system in the steady state. ($a(.)$ may not be an increasing function in this case.) As mentioned previously, these functions can be characterized more exactly by conducting marketing experiments. As α increases, the price drops and hence more people are attracted to the system. But for $\alpha > \alpha_\lambda$, the high rate of ad insertion acts more as a deterrent, and λ decreases. Also, in this case, the number of channels that are needed to support a particular arrival rate is lower bounded by $N(\alpha) > \lambda(\alpha)L$. Therefore we have an optimization problem akin to the one discussed in Sec. 2:

Maximize: $P(\alpha) = \lambda(\alpha)p(\alpha) + a(\alpha) - N(\alpha)b$

Subject to: (1) $0 \leq \alpha < \alpha_0$ and (2) $N(\alpha) > \lambda(\alpha)L$

P is maximized for a particular value of $\alpha = \alpha^*$, where α^* is a root of the following equation:

$$p'(\alpha)\lambda(\alpha) + a'(\alpha) + (p(\alpha) - L)\lambda'(\alpha) = 0$$

Stream Merging Based Systems Constrained ad insertion has been proposed for reducing the server and network bandwidth requirements by reducing the temporal skews between adjacent streams receiving the same content [1]. Essentially, a leading stream is put onto a multicast ad channel for A_{max} time units and then is shown video for V_{min} time units, in a cyclic manner, while a trailing stream continuously receives video until it merges with the leading stream. The merging algorithm ensures that no stream receives ads at a rate greater than α.

In order to support users arriving into the system at a rate $\lambda(\alpha)$, in steady state, the number of channels needed is given by $N_m(\alpha)$ which is an decreasing function of α for a constant arrival rate λ. However, since the arrival rate is influenced by the ad ratio, $N_m(\alpha)$ may not be a strict decreasing function. For $\alpha > \alpha_\lambda$, $N_m(\alpha)$ drops due to the drop in the arrival rate, and due to greater ad insertion. But for $0 \leq \alpha \leq \alpha_\lambda$, $N_m(\alpha)$ is dictated by two opposing behaviors: increase in arrival rate (which tends to increase N_m) and greater ad insertion (tends to decrease N_m). Hence the cumulative effect of these two forces may result in either an increasing or a decreasing function of α. It is very hard to characterize $N_m(\alpha)$ analytically, but it can be characterized by simulations. Two curves that could possibly characterize $N_m(\alpha)$ are shown in Fig. 2(c). Again, we are faced with a profit maximization problem as before:

Maximize: $P(\alpha) = \lambda(\alpha)p(\alpha) + a(\alpha) - N_m(\alpha)b$

Subject to: (1) $0 \leq \alpha < \alpha_0$

P is maximized for $\alpha = \alpha^*$ where α^* is a root of the following equation:

$$p'(\alpha)\lambda(\alpha) + p(\alpha)\lambda'(\alpha) + a'(\alpha) - bN_m'(\alpha) = 0$$

5 Conclusion and Future Work

We investigated pricing related tradeoffs involved in different types of VoD system deployments. Three different settings were analyzed: (1) a simple interactive VoD system with a dedicated channel per user, (2) a staggered broadcast delivery system, and (3) an aggregation system based on ad-insertion. We demonstrated in this paper how a multitude of factors such as price, QoS level (start-up latency and ad ratio), and user arrival rates can influence each other, and how the optimal price of a VoD service can be obtained (from a profit maximization perspective) in each of the above settings. We believe that such an analysis of pricing models can help in establishing a sound economic basis for wide deployment of VoD in future.

There are some open issues that need investigation. Waiting times in VoD systems due to queueing (we have addressed waiting times due to batching) can be another QoS factor for pricing. (Large waiting times can result in loss of revenue due to *reneging*.) Also, a good model of user behavior with respect to factors such as price is needed (by extensive survey) for proper characterization of the functions mentioned in this paper.

References

[1] P. Basu, A. Narayanan, W. Ke, T. D. C. Little, and A. Bestavros. Optimal scheduling of secondary content for aggregation in video-on-demand systems. In *Proceedings ICCCN '99, Boston-Natick, MA*, pages 104–109, October 1999.

[2] A. Dan, P. Shahabuddin, D. Sitaram, and D. Towsley. Channel allocation under batching and VCR control in video-on-demand systems. *Journal of Parallel and Distributed Computing*, 30(2):168–179, November 1995.

[3] L. Golubchik, J. Lui, and R. Muntz. Adaptive piggybacking: A novel technique for data sharing in video-on-demand storage servers. *ACM/Springer Multimedia Systems*, 4:140–155, 1996.

[4] K. A. Hua and S. Sheu. Skyscraper broadcasting: A new broadcasting scheme for metropolitan video-on-demand systems. In *Proceedings ACM SIGCOMM '97, Cannes, France*, pages 89–100, September 1997.

[5] Interactive TV trials. URL – http://www.teleport.com/~samc/cable4.html.