Pricing for a Hybrid Delivery Model of Video Streaming

J.-Ch. Grégoire
Energie, Matériaux et Télécommunications
Institut National de la Recherche Scientifique
Montréal, Québec, Canada
gregoire@emt.inrs.ca

Angèle M. Hamel & D. Marc Kilgour
Department of Mathematics
Wilfrid Laurier University
Waterloo, Ontario, Canada
ahamel@wlu.ca, mkilgour@wlu.ca

ABSTRACT

Media streaming for video-on-demand requires a large initial outlay on server infrastructure. However, a peer-to-peer system whereby existing customers act as relays for new customers is a simple way to provide temporary capacity and gauge demand before committing to new resources. A customer who agrees to act as a relay should be provided the service at a discounted price, but at what price, and is this price affordable for the content provider? This paper investigates financial incentives for the hybrid model of video streaming services.

Keywords

Content Delivery, Hybrid P2P, Content Relay, Cost-Revenue Trade-off Analysis

1. INTRODUCTION

The challenge of delivering growing, and increasingly diverse, Internet content has led to the creation of dedicated or optimized delivery infrastructures. A large and dedicated infrastructure may be more cost-effective and efficient, but it poses a scalability problem when infrastructure limits are reached. At this stage, it may be difficult to justify a large outlay to acquire more infrastructure if demand and revenue are not apparent but merely anticipated. An ideal solution would be some way of testing the waters and gauging demand.

Among alternatives that ensure scalability, Peer-to-Peer (P2P) has often been touted as a natural complement to a fixed infrastructure [5, 6] and even deployed in such hybrid form [7]. It is usually taken for granted that the P2P alternative exists for a variety of reasons. Realistically, however, an existing subscriber who consents to act as a relay would demand a financial incentive, which should not be detrimental to the revenue. Thus, the benefit for the provider would depend on finding the right number of subscribers to act as relays.

However, most studies we are aware of on the hybrid model have focused on performance and there has been little evidence of work on its economics. For example, Balachandran *et al.* [1] looks at two models—a federation model with cross-region load balancing, and a P2P-CDN hybrid—and analyzes real data and real patterns to compare them. Zhao *et al.* [8] also explores a peer-assisted CDN, but their system does not have an incentive mechanism. Liu *et al.* [3] and Liu *et al.* [4] both consider the performance dimension using heterogeneous sources, but do not consider the economic dimension. Garmehi *et al.* [2] do consider an economic approach, but their approach differs from ours in that they consider traditional P2P where you can only get content if you are willing to share it, and they rebate bandwidth rather than price.

Our fundamental questions are: What level of incentive will motivate enough subscribers to act as relays? What is the maximum level of additional service that can be achieved using this strategy?

We take two practical approaches, considering first in Section 2 a baseline model without relays, followed by a consideration in Section 3 of the model with relays. Section 4 is a conclusion and future work.

2. MODEL WITHOUT RELAYS

In this section we establish notation and a baseline before turning to the effects of relays in Section 3. Let $\mathcal{N}$ be the set of potential users (i.e., the population of the region, city, country, etc. that can access the service). We suppose that the value of a subscription to the service is exponentially-distributed within the population, with mean m_v. This distribution is based on the probability density function

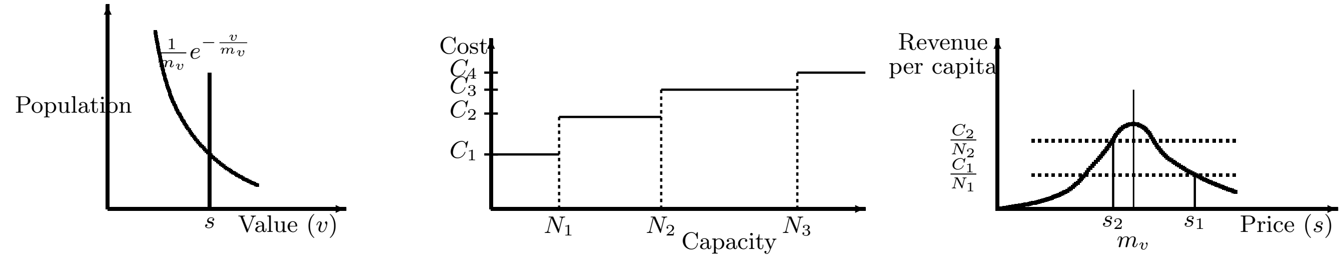
$$f_v(v) = \frac{1}{m_v} e^{-\frac{v}{m_v}} \tag{1}$$

defined for $v \geq 0$. See Figure 1a. Fix $s > 0$; the exponential distribution implies the probability a randomly chosen individual has value $v \geq s$ is $\int_s^\infty f_v(v)dv = e^{-\frac{s}{m_v}}$.

Now suppose that s is the subscription price of the service. Writing $N = |\mathcal{N}|$, we then have that, at equilibrium, there will be a large number of subscribers, namely $N \exp\{-\frac{s}{m_v}\}$. In Figure 1a, the set $\mathcal{S} \subseteq \mathcal{N}$ of subscribers consists of all those individuals whose value falls to the right of s along the horizontal axis. Note that our model of the relation of subscribers to price is an equilibrium model—if the price of the subscription were to rise or fall, the number of subscribers would fall or rise after a (fast) adjustment period.

In our model, the population $\mathcal{N}$ is fixed, and everyone in the population is potentially a subscriber—but only a certain fraction would value the service enough to pay the subscription price, s. Moreover, to increase the ranks of subscribers, the subscription price must be reduced, which of course affects revenues.

In the baseline situation, the provider's infrastructure consists of a fixed number of servers, serving (directly) a fixed



(a) Distribution of population vs. value (b) Total cost vs. number of subscribers. (c) Revenue & value as functions of s

Figure 1: Value and Cost Analysis

number of customers. To add capacity, a new server must be acquired. Thus the provider's cost, relative to the number of customers, is described by a step function; as each new server comes online, there is an increment to the overall cost, as shown in Figure 1b. Note that per-user administrative costs are considered negligible.

Of particular interest is what happens at the right-hand boundary of each step, as the number of users reaches the capacity of the current infrastructure. Deploying a new infrastructure raises the provider's total cost from C_i to C_{i+1}, a major increase, but permits an extra $N_{i+1} - N_i$ customers. Would the provider benefit by upgrading infrastructure?

Suppose that the price of the service is s, that M customers are being served, and that the provider's infrastructure lies on the first step of Figure 1b, where the cost is C_1 and at most N_1 customers can be served. Then the provider's profit equals $P = Ms - C_1$, provided that $M = Ne^{-\frac{s}{m_v}} \leq N_1$, which is equivalent to

$$s \geq m_v \ \ln\left(\frac{N}{N_1}\right) = s_1. \tag{2}$$

Thus, assuming (2), the provider's profit is

$$P = Nse^{-\frac{s}{m_v}} - C_1. \tag{3}$$

Consider in general a function of the form $x \exp\left\{-\frac{x}{m}\right\}$. Its derivative with respect to x is $\exp\left\{-\frac{x}{m}\right\}\left(1 - \frac{x}{m}\right)$. It follows that the function is increasing on $[0, m]$ and decreasing on $[m, \infty)$. Recalling that the expression (3) is applicable only when $s \geq s_1$, we have shown that the maximum profit occurs at $s = m_v$ if $s_1 < m_v$, and at $s = s_1$ otherwise. To interpret this finding, note that if m_v is relatively small (i.e., N is large relative to N_1), then profit is bound by the infrastructure constraint, and the optimal price (satisfying $s \geq s_1$) is $s = s_1$, which attracts N_1 customers, exactly matching the capacity of the infrastructure. In this case, profit is $P_1 = N_1 s - C_1 = Ns_1 e^{-\frac{s_1}{m_v}} - C_1$.

Now define

$$s_j = m_v \ \ln\left(\frac{N}{N_j}\right), \tag{4}$$

for $j = 1, 2, 3, \ldots$. It then follows from (2) and (4) that $s_1 > s_2 > s_3 > \ldots$. It is then easy to check that $s_{j+1} \leq s \leq s_j$ implies that $N_{j+1} \leq N \exp\left\{-\frac{s}{m_v}\right\} \leq N_j$. Thus subscription prices between s_{j+1} and s_j are appropriate to the j^{th} level of infrastructure illustrated in Figure 1b.

The earlier analysis shows that, if $s_{j+1} \leq s \leq s_j$, then a maximum of profit occurs at $s = m_v$ if $s_{j+1} \leq m_v \leq s_j$, and $s = s_{j+1}$ otherwise. In the former case, the maximum profit is $P_{m_v} = Ne^{-1} - C_{j+1}$; in the latter case, it is $P_{j+1} = Ne^{-\frac{s_{j+1}}{m_v}} - C_{j+1}$. Revenue and cost curves are shown in Figure 1c. For simplicity, the Figure shows per-capita revenue and cost, for example $\frac{P_{j+1}}{N} = e^{-\frac{s_{j+1}}{m_v}} - \frac{C_{j+1}}{N}$.

If $s_j > m_v$ and the curve at s_j lies below $\frac{C_j}{N}$, then it is never profitable to do business on the j^{th} step of the step function. But if $s_j > m_v$ and the height of the curve at s_j is above $\frac{C_j}{N}$, then values of s just above s_j will bring in approximately N_j subscribers, ensure that profit is positive, and, as Figure 1c illustrates, closely maximize profit, at least locally. Also, if $s_j < m_v < s_{j-1}$, a local maximum of profit occurs at $s = m_v$—again, profit may or may not be positive, depending on the value of C_j.

Figure 1c illustrates the problem of generating profit: the provider must choose a level of infrastructure (i.e., a value of j) and a value of s that maximizes the gap between the curve and $\frac{C_j}{N}$. We have shown that the best value of s will be either $s = m_v$ or $s = s_j$ for some value of j such that $s_j > m_v$. The maximum of profit will occur at whichever of these values of s corresponds to the greatest gap between the revenue curve and the cost per capita.

3. MODEL WITH RELAYS

Now we turn to the situation in which some subscribers act as relays. The step function describing costs shown in Figure 1b, and the hard limit it represents, explain why the relays might be valuable; a provider might increase the number of subscribers without incurring the immediate cost of acquiring more infrastructure. The problem is that in order to increase the total number of subscribers, the subscription price must fall, which may negatively affect revenue.

First we augment our model of the population so that it includes subscribers' costs for acting as a relay. Recall that $\mathcal{N}$ is the set of potential users. We have already assumed that the value of the service is exponentially-distributed within the population, with parameter $m_v > 0$.

Now we assume that the cost of acting as a relay is also exponentially distributed within the population, with parameter $m_c > 0$. Moreover, we assume that the distributions of cost and value are independent. Thus the probability density function of cost, c, is

$$f_c(c) = \frac{1}{m_c}e^{-\frac{c}{m_c}}. \tag{5}$$

where m_c and m_v are the means of the distributions.

Fix $c_0 > 0$. Then the probability that a randomly chosen individual has cost $c \le c_0$ for acting as a relay is

$$\int_0^{c_0} f_c(c)dc = 1 - e^{-\frac{c_0}{m_c}}. \qquad (6)$$

In particular, the expected number of subscribers whose value for the service is at least s, for whom the cost of acting as a relay is less than c_0, is $Ne^{-\frac{s}{m_v}}\left(1 - e^{-\frac{c_0}{m_c}}\right)$.

Assume that at the outset there are no relays and the price s_1 has been set to match *exactly* the number of subscribers that can be accommodated with the current infrastructure, N_1. As argued earlier: $N_1 = Ne^{-s_1/m_v}$.

Now consider a lower price $s < s_1$, which should lead to an increase in the number of subscribers, but at the same time we allow subscribers to become relays if they choose to do so. In our model, a relay is someone for whom the rebate (reduced subscription price) compensates the extra cost incurred for relaying traffic to another subscriber. The reduced price, r, is expressed as a percentage of s, so a subscriber acting as a relay will pay rs for the service.

Figure 2 illustrates the new situation with price s_1 chosen as described above. The vertical axis represents the cost of being a relay: some of the original subscribers are relays ($\mathcal{M}$), some remain subscribers but are not relays ($\mathcal{L}$) and the service has acquired new subscribers ($\mathcal{K}$). Our model is that a subscriber from the original group who becomes a relay must be paired with exactly one new subscriber, s, who must be one of the subscribers who was attracted to the service at the new price $s < s_1$. Note also that in this model new subscribers cannot be relays; only original subscribers can convert to relays. Thus the number of subscribers in $\mathcal{K}$ matches the number of relays in $\mathcal{M}$, or $K = |\mathcal{K}| = M = |\mathcal{M}|$.

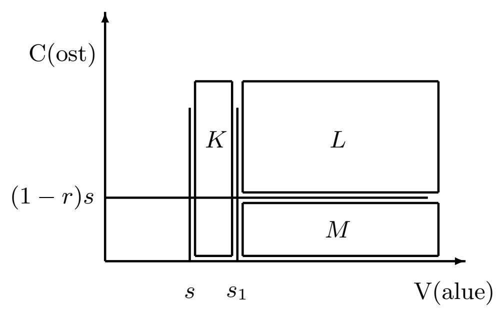


Figure 2: Subscribers with relays.

Thus, Figure 2 shows the number of original subscribers who become relays (in $\mathcal{M}$), $M = Ne^{-s_1/m_v}(1 - e^{-(1-r)s/m_c})$ and the number of new customers acquired, $K = N(e^{-s/m_v} - e^{-s_1/m_v})$. Clearly, the values of r must be set so that $K = M$. Equating the above equations and solving for r produces

$$r = 1 + \frac{m_c}{s} \ln\,(2 - e^{(s_1 - s)/m_v}). \qquad (7)$$

The total revenue is the sum of revenues from each of the three populations, $\mathcal{K}, \mathcal{L}$, and $\mathcal{M}$, i.e. $(K \times s) + (L \times s) + (M \times r \times s)$. Or, considering that $K = M$ and $L + M = N_1$ (see Figure 2), total revenue equals $(N_1 \times s) + (M \times r \times s) = sN_1 + rsN(e^{-s/m_v} - e^{-s_1/m_v})$. Substituting the value for N_1 and r above, total revenue equals

$$sNe^{\frac{-s_1}{m_v}} + sN(e^{-\frac{s}{m_v}} - e^{-\frac{s_1}{m_v}})\left(1 + \frac{m_c}{s}\ln\,\left(2 - e^{\frac{s_1 - s}{m_v}}\right)\right)$$

$$= sNe^{\frac{-s_1}{m_v}}\left(1 + \left(\frac{e^{\frac{-s}{m_v}}}{e^{\frac{-s_1}{m_v}}} - 1\right)\left(1 + \frac{m_c}{s}\ln\left(2 - e^{\frac{s_1 - s}{m_v}}\right)\right)\right).$$

Recall from (1) that the distribution of the population based on some value v and follows a negative exponential, whose rate of decrease is fixed by the value m_v. The smaller the value, the faster the decrease, meaning that the population is concentrated on smaller values of v—and we will say that the population is more sensitive to v. On the other hand, as m_v gets larger, the population spreads more evenly, and is less sensitive to the value of v. Similar statements can be made about the sensitivity to the value of c. We now look at the effect of the values of m_c and m_v on the relay population and the possibility of increased revenues.

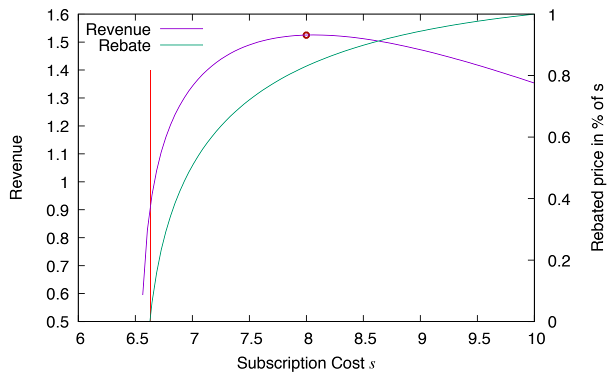
Figure 3 displays revenue and rebate functions for four different combinations of values for m_c and m_v. In all figures, the horizontal axis presents values of s. We assume that s_1 is 10 and study values of s in the 6 to 10 range. The left-hand vertical axis is the total revenue—note that its range varies between figures, for the sake of clarity. The right-hand vertical axis is the rebate r, with values between 0 and 1. Notice that, in all cases, the rebate curve passes through the point (10, 1). In these figures, the values of s_1, m_c and m_v have been chosen somewhat arbitrarily, but this is not an issue since we are interested in trends, rather than precise results. More specifically, we want to observe under which conditions it is possible to increase revenues using our scheme.

Figure 3a shows the curves for $m_c = 2$ and $m_v = 5$. To understand this figure, suppose that the price reduction starts at $r = 1$. Note that subscription price is $s = s_1 = 10$ at this point, and total revenue is approximately 1.4. As r moves down the right-hand vertical axis, the corresponding subscription price s to balance the new customers and the relays is the value of s at the corresponding point of the rebate curve. For example, at $r = 0.6, s = 7$ and revenue equals 1.3. Note also that when $r \to 0$, the subscription price that exactly balances new customers and relays approaches 6.6, and total revenue is about 0.8. The vertical line at this point shows what happens when all relays receive the service for free. Study of Figure 3a shows there is a peak in revenue, about 1.5, for a value of s around 8, labeled by a) on the plot, corresponding to reduced price $r = 0.85$.

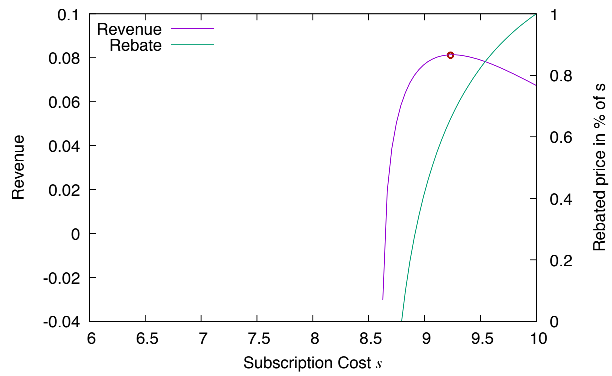
Lower sensitivity to value means that the revenue curve is flatter near the peak, which leads to stable results with a variation of ±5%. The matching Figure 3b, on the other hand, shows the results when the sensitivity is opposite. The rebate curve falls faster, and the peak is narrow.

To study the effect of discrepancy between sensitivity to cost and value, we examined larger differences in values: $m_c = 2$ and $m_v = 50$ and the opposite, $m_c = 50$ and $m_v = 2$ (not shown). Both cases have significantly different rates of descent for the rebate, but revenue is maximized at or near $s = s_1 = 10$ i.e., there is no benefit to using relays here.

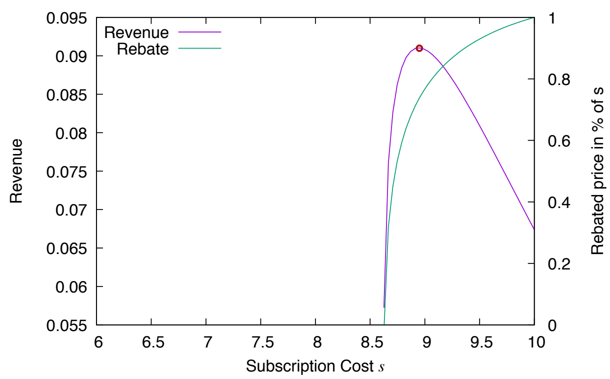
Finally, we look at the situation of same values for m_c and m_v, either both small in Figure 3c, or both large in Figure 3d. In both cases we have a peak value for revenue, and its spread is larger for large values of m_c and m_v, leading to more stable results. For larger values, we see that the start-



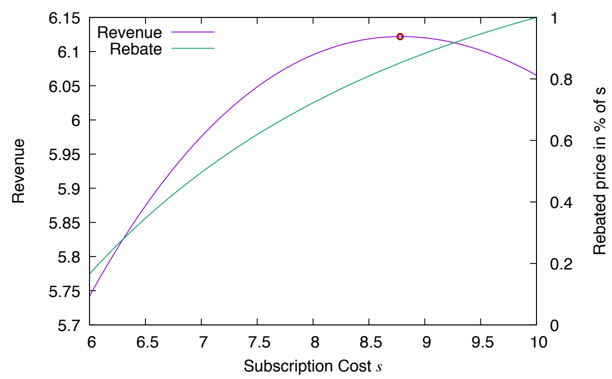
(a) $m_c = 2$ and $m_v = 5$



(b) $m_c = 5$ and $m_v = 2$



(c) $m_c = 2$ and $m_v = 2$



(d) $m_c = 20$ and $m_v = 20$

Figure 3: Revenue function for values of s at 10 and below.

ing point (at $r = 1$) is closer to the peak value, which is more stable for small changes in s, quite unlike what happens for small values of m_c and m_v.

4. CONCLUSION AND FUTURE WORK

We observe these trends: 1) It is possible to increase revenues with the introduction of relays; 2) This works best when there is not too large a discrepancy between sensitivity to cost and sensitivity to value; 3) Results are more stable for lower sensitivity to value. At a fundamental level we made assumptions about value and cost distributions, namely that they followed probability density functions, (1) and (5). These are reasonable assumptions, but actual data could lead to a model with more "realistic" distributions.

Moreover, our constructed model is not an equilibrium model, but could be adjusted to become more in line with stability considerations. We assumed only existing customers became relays, but what if some users were attracted by the (cheaper) relay option and who would not have been existing customers but would have been enthusiastic to act as relays? These questions will be explored in future work.

Acknowledgement: Work supported by a Wilfrid Laurier University Internal Grant and Canadian Tri-Council Research Support Fund. Authors were each supported by their own individual NSERC Discovery Grants.

5. REFERENCES

[1] A. Balachandran, V. Sekar, A. Akella, S. Seshan, Analyzing the potential benefits of CDN augmentation strategies for internet video workloads, IMC 2013.

[2] M. Garmehi, M. Analoui, M. Pathan, R. Buyya, An economic mechanism for request routing and resource allocation in hybrid CDN-P2P networks, *Int. J. of Network Management* 25 (2015), 375–393.

[3] H. Liu, Y. Wang, Y. R. Yang, A. Tian, and H. Wang. Optimizing Cost and Performance for Content Multihoming. In SIGCOMM, 2012.

[4] X. Liu, F. Dobrian, H. Milner, J. Jiang, V. Sekar, I. Stoica, and H. Zhang. A Case for a Coordinated Internet-Scale Video Control Plane. In Proc. SIGCOMM, 2012.

[5] Z. Lu, Y. Wang, Y.R. Yang, An analysis and comparison of CDN-P2P-hybrid content delivery system and model, *J. Comms*, 7 (2012), 232–245.

[6] Z. Shen, J. Luo, R. Zimmermann, A.V. Vasilakos, Peer-to-peer media streaming: Insights and new developments, *Proc. IEEE* 99 (2011), 2089–2109.

[7] H. Yin, X. Liu, T. Zhan, V. Sekar, F. Qiu, C. Lin, H. Zhang, B. Li, Design and deployment of a hybrid CDN-P2P system for live video streaming: experiences with LiveSky, ACM International Conference on Multimedia (SIGMM), 2009, 25–34.

[8] M. Zhao, P. Aditya, A. Chen, Y. Lin, A. Haeberlen, P. Druschel, B. Maggs, B. Wishon, M. Ponec, Peer-assisted content distribution in Akamai NetSession, IMC 2013.