# Fair Pricing of Video Transmissions using Best-Effort and Purchased Bandwidth

Stephan Bohacek
Department of Electrical and Computer Engineering
University of Delaware
Newark, DE 19716
bohacek@eecis.udel.edu

### Abstract

A model based approach to bandwidth pricing is developed. The focus is not on how much an ISP should sell bandwidth for, but rather, how much bandwidth a video service provider (VSP) will need to use beyond the bandwidth provided via best-effort. An algorithm is presented where the VSP sells the ability to transmit a movie. It is assumed that the end-user pays the VSP for this ability at the beginning of the download, whereas the VSP pays the ISP for the bandwidth at the end of the download. Hence, the VSP must predict how much bandwidth will be required.

There has been extensive research focused on how QoS guarantees can be accommodated in data networks [1], [2]. Typically, these approaches call for a single network to accommodate several classes of QoS. The idea behind this multitiered approach is that if users pay a premium they will be granted better service. While this research has reached advanced stages, there has been less work focusing on how these QoS guarantees should best be used. For example, if an ISP makes different levels of QoS guarantees available for different prices, how should end users decide if the extra prices should be paid, or should the "free" best-effort service is used.

In this paper, the fair price for bandwidth for video transmission is examined. We assume that the ISP provides QoS for a fixed and known price. The end user is in the market to watch a video in real-time. Thus, once the user decides to view the video, playback begins perhaps after a short period to fill receiving buffers. In this real-time or nearly real-time setting, it is not uncommon that best-effort service will not provide adequate service with a sufficiently high probability. Hence, the best effort service must be supplemented with purchased bandwidth. We envision a transaction model where the video service provider (VSP) negotiates with and pays the ISP for the extra bandwidth. The VSP then passes the price of the bandwidth on to the customer. However, the customer demands to pay up-front, while the ISP must be paid after the transfer is complete. Thus, the VSP must predict the amount that will eventually be paid to the ISP. Since it is assumed that the price of the bandwidth provided by the ISP is known to the video service provider and remains fixed during the transfer, the VSP must simply predict bit-rate to be purchased and the amount of time that the bit-rate is required.

To determine these values, the amount of available best-effort bandwidth must be predicted. For example, suppose that it is decided that at least 99% of the end users should receive the video. Then, if it is predicted that 99% of the time, the best-effort service can provide the require bandwidth, then there is no need to purchase any additional bandwidth. On the other hand, if, for example, the movie is 1.3GB and the best-effort service is predicted to only provide on-time transmission of 822MB to 99% of the users, then the 478MB short fall must be closed with purchased bandwidth. If, for example, the movie is 1.5 hours long, and 0.7Mbps is purchased, then 99% of the end users will receive the movie. However, the video service provider can typically change much less then the price for 0.7 Mbps time 1.5 hours. To see this, we present the following algorithm where the purchased bandwidth is terminated before the end of the movie.

# 1 Overview of Algorithm

The VSP (video service provider) estimates the cost of the bandwidth that must be purchased to support the transfer of the video. The VSP must determine the amount of bandwidth to buy and the length of time that the bandwidth is to be purchased for. The bandwidth to be purchased is the amount to ensure that with probability 0.99, the entire video will be delivered on time. Then, to minimize the amount of time that bandwidth is purchased, the VSP will no longer purchase bandwidth when there is a 0.99 probability that the entire video will be delivered on time. The VSP then charges enough money to cover the expected cost of this purchased bandwidth.

An important complication of the VSPs task is that the network conditions may vary. As will be discussed in later, the network conditions are summarized in by the state of single stochastic process $\theta_t$, the congestion level. Thus, the VSP must predict the bandwidth to be purchased in the face of changing network conditions. To simplify the presentation, it will be assumed that the congestion level at the start of the video transmission is known. In the rest of the paper, this assumption is lifted.

Let $M$ be the size of the video and the length of the video is $T_{Video}$ seconds. Let $P\left(D_{T_{Video}}|\theta_0\right)$ be the cumulative distribution of the amount of data sent over the best effort connection during the playing of the video. Let $Q_{0.99}$ be such that $P\left(D_{T_{Video}} < Q_{0.99}|\theta_0\right) = 0.99$. Then the purchased bandwidth is $(M - Q_{0.99})/T_{Video}$. At the start of the video transmission, both the best effort and the purchased bandwidth are used. Then, at every subsequent moment, a judgement is made whether the purchasing of the bandwidth should continue or not. Let $P\left(C_{T_{Video}-t}|\theta_t\right)$ be the cumulative distribution of the total data that can sent, $C$, in the remaining $T_{Video}-t$ seconds given the current congestion level $\theta_t$. Then the bandwidth is no longer purchased once $P\left(C_{T_{Video}-t} > D_t - M|\theta_0\right) > 0.99$, where $D_t$ is the amount of data sent up to time $t$. Let $S$ denote the time that the bandwidth purchased is ended. Then the VSP must compute the distribution of $S$, i.e., $P\left(S|\theta\right)$.

In all there are two related distributions to be computed, $P\left(C_{T_{Video}-t}|\theta_t\right)$ , the distribution of the data that could be send in the remaining time and $P\left(S|\theta\right)$ the distribution of the time that the bandwidth is purchased. In order to compute these distributions, the stochastic models are developed.

The paper proceeds as follows. Section 2 presents a model of the network and TCP running over the network. This section also develops the median TCP sending rate as a TCP friendly sending rate. The work presented in this section is only a brief review, a more details discussion can be found in [3]. Sections 3 - 5 use the models of Section 2 to determine the require probability distributions. Finally, Section 6 presents an example.

# 2 A Model of the Network and TCP

## 2.1 Models of Latency

There has been extensive work focused on understanding the distribution of latency, e.g. [4], [5], [6]. However, there has been less work on developing a dynamical model of latency. Work that has investigated the dynamics of the round-trip time includes [7] and [3]. In [3], diffusion models of round-trip delay are developed and will be briefly reviewed next.

Let $R_t$ be the time-varying part of the round-trip time experienced by a packet sent at time $t$. Thus, the actual round-trip time is $RTT_t = R_t + T$ where $T$ accounts for fixed delays such as propagation delay, transmission delay, etc. and depends on the size of the packet. On the other hand, $R_t$ is dominated by queuing delay but may also include effects such as address lookup. However, the effect of delays between network layers and transport or application layers is not modeled. The first model presented in [3] is a simple three parameter mean-reverting diffusion model

$$dR_t = \frac{\sigma_{\theta_t}^2}{2}\left(\lambda_{\theta_t} - \phi_{\theta_t} R_t\right) dt + \sigma_{\theta_t}\sqrt{R_t} dB_t, \tag{1}$$

where $B_t$ is Brownian motion, $\phi, \lambda$ and $\sigma$ are scalar parameters that are functions of $\theta$ which is a continuous time Markov chain. We assume that $\theta$ makes jump fairly infrequently, hence, much analysis is carried out assuming that $\theta$ is fixed. In the case that $\theta$ is fixed, (1) is known as CIR model of interest rates and has been widely studied and utilized in finance [8].

While (1) provides a good fit during periods of low to moderate congestion, during periods of high congestions, when the tail of the stationary distribution is large, the fit is not that good. While it might be possible to achieve better fitting by allowing $\theta$ to vary rapidly, the goal is to have $\theta$ only jump when there is a change in the level of congestion. Therefore, instead of varying $\theta$, during periods of high congestion, a better fitting model is had at the expense of adding more parameters. A six parameter model of round-trip time is

$$dR_t = \frac{\sigma_{\theta_t}^2}{2} R_t^{\rho_{\theta_t}-1}\left(\frac{\left(\delta_{\theta_t} R_t^{\delta_{\theta_t}} + \gamma_{\theta_t}\beta_{\theta_t}^{\gamma_{\theta_t}} R_t^{\gamma_{\theta_t}}\right)}{\left(R_t^{\delta_{\theta_t}} + \beta_{\theta_t}^{\gamma_{\theta_t}} R_t^{\gamma_{\theta_t}}\right)} + \rho_{\theta_t} - \phi_{\theta_t} 2\ln\left(R_t\right)\right) dt + \sqrt{\sigma_{\theta_t}^2 R_t^{\rho_{\theta_t}}} dB_t. \tag{2}$$

A single set of parameters does not provide a good fit for all times. For this reason, the parameters are permitted to vary and the purpose of $\theta$ is to account these variations. Since it is changes in congestion that lead to changes in the parameters $\theta$ is referred to as the congestion level. However, it is an abstract variable, so, for example, a large value of $\theta$ does not imply that there are more competing flows.

The variation of $\theta$ is modeled continuous Markov chain. Thus, the probability of $\theta$ making a jump[1] if $q$ and, given a jump occurs, the probability of jumping from $\theta_1$ to $\theta_2$ is $k\left(\theta_1, \theta_2\right)$. As discussed in [3], a useful way to compute $q$ and $k$ is by examining the parameter variation made over short time intervals. Also, as shown in [3], the parameter

[1] It is assumed that the rate of jumping from a state is independent of the state, i.e., $q$ does not depend on $\theta$.

variation over a short time interval $T$ is well modeled by a mixture of two Laplace

$$p_{a,b,p}\left(\phi_T|\phi_0\right) = \left(1-c\right)\alpha\exp\left(-2\alpha\left|\phi_T - \phi_0\right|\right) + cb\exp\left(-2b\left|\phi_T - \phi_0\right|\right). \tag{3}$$

With such densities, and correlations between parameters, it is possible to define a state space for $\theta$, define mappings $\lambda_\theta$, $\phi_\theta$, etc., and to define values of $q$ and $k$. While some progress has been made in finding compact expressions for the mapping and dynamics of $\theta$, much work remains. One surprising and useful observation is that the models for parameter variation do not change much from connection to connection.

## 2.2 Models of Loss Probability

There has been extensive work on modeling packet loss. In [10], a small network is considered and a deterministic model for packet drops is developed. In [11] and [12], drops are assumed to be highly correlated over short time scales and independent over longer time scales. In [13], drops are assumed to be bursty. In [14], drops are modeled as a renewal process with various distributions; deterministic, Poisson, I.I.D. and Markovian. A specific example of the model in [14] is developed in [15], where drops are modeled by a Poisson process. In [16] and [17], this approach is generalized and drops events are modeled as a Poisson process where the intensity depends on TCP's congestion window size. In [18], a dynamic model of loss is developed. The model developed by this effort is also dynamic. The difference between the two is that our model recognizes a strong dependence on the round-trip time. Because of this correlation, the dynamic models of loss probability are very efficient.

The objective is to find the loss probability. Since this loss probability depends on the latency we define the conditional loss probability

$$g\left(R_t\right) = P\left(\text{packet send at time } t \text{ is dropped } |R_t\right).$$

While many models of $g$ are appropriate, a spline representation has proven useful. There has been extensive work smoothing observed data with splines [19]. To this end, define

$$g\left(R_t, \theta_t\right) = \alpha_{0,\theta_t}T_0\left(R_t\right) + \alpha_{1,\theta_t}T_1\left(R_t\right) + \cdots + \alpha_{n,\theta_t}T_n\left(R_t\right),$$

where $\{T_i\}$ is a set of functions. For example, these could be Taylor series functions, 1, $x$, $x^2$,..., Chebshev polynomials, or, as we have chosen them, splines. Since the round-trip times of dropped packets is not observed, the conditional loss probability is not directly observable. However, after some elementary manipulation (see [3]), it is seen that the coefficients can be found by solving a system of linear equations.

## 2.3 A Diffusion Model of TCP

Given a model of the network, it is possible to develop a model the dynamics of TCP's congestion window. Following the ideas in [15], a stochastic differential equation model of the congestion window is

$$\begin{aligned} dX_t &= \frac{1}{T+R_t}dt - \frac{1}{2}X_t dN_t \\ dR_t &= \mu\left(R_t\right)dt + \sigma\left(R_t\right)dB_t, \end{aligned} \tag{4}$$

where $\mu$ and $\sigma$ are functions such as the ones given in Section 2.1 and $N$ is a Cox process that counts the number of drops where the drop occur at rate $g\left(R_t, \theta_t\right) \times X_t / \left(T + R_t\right)$.

While this model is similar to that in [15]. However, here the fact that round-trip time is not constant is embraced and that the drop rate depends on the sending rate. (In [15] the drop rate was assumed to be independent of sending rate, which implies that the drop probability decreases as the sending rate increases).

With these models, it is straight forward to determine the probability density functions of the congestion window. Let $p(x,r,t)$ be the probability density function for (4), i.e., let $p(x,r,t) = \frac{\partial^2}{\partial x \partial r} P(X_t < x, R_t < r)$ . We will assume that this and all necessary densities exists. Then $p(x,r,t)$ obeys

$$\begin{aligned}
&\frac{\partial}{\partial t} p(x,r,t) \\
&= -\frac{\partial}{\partial r}\left(\mu(r)\, p(x,r,t)\right) + \frac{1}{2}\frac{\partial^2}{\partial r^2}\left(\sigma^2(r)\, p(x,r,t)\right) \\
&- n(x,r)\, p(x,r,t) \\
&- \frac{1}{T+r}\frac{\partial}{\partial x} p(x,r,t) + 2n(2x,r)\, p(2x,r,t),
\end{aligned} \tag{5}$$

where $n(x,r)$ is the drop event rate discussed in subsection **??**. This partial differential equation representation of the density is known as Kolgomorov's forward equation and is a straight forward application of standard methods in stochastic differential equation [20]. It is not hard to show that the system is ergodic. Hence, with $p(x,r,t)$ found, the stationary density can be found by letting $t \to \infty$, i.e., $p(x,r) = \lim_{t\to\infty} p(x,r,t)$.

# 3 Distribution of the Total Data Sent

In order to decide how much bandwidth should be purchased and if the purchased bandwidth should continued to be purchased, the probability density function (PDF) of the total data that will be sent must be determined. Specifically, given the current state and the time remaining, we must determine the PDF of that total data that could be fairly sent during the remaining time. There are many approaches to take and some useful approximations can be made. First, the data sending rate must be determine. If a TCP-friendly approach is used and the median TCP sending rate is used, then the sending rate can be express by function $S(R,\theta)$, where $R$ is the current round-trip time and $\theta$ is the current congestion level. With this sending rate, the PDF of the total data can be found as follows.

Let $p_T(C,R,\theta|R_0,\theta_0)$ be the probability density of the cumulative data sent, $C_T$, in the remaining $T$ seconds with terminal round-trip time $R_T$ and congestion level $\theta_T$ and initial round-trip time $R_0$ and congestion level $\theta_0$. Then this PDF obeys

$$\begin{aligned}
\frac{\partial p_T}{\partial T}(c,r_T,\theta_T|r_0,\theta_T) &= -f(r_0,\theta_0)\frac{\partial p_T(c,r_T,\theta_T|r_0,\theta_T)}{\partial c} \\
&+ \mu(r)\frac{\partial}{\partial r_0} p_T(c,r_T,\theta_T|r_0,\theta_T) + \frac{1}{2}\sigma^2(r)\frac{\partial^2}{\partial r^2} p_T(c,r_T,\theta_T|r_0,\theta_T) \\
&+ \sum_{\phi} q(\theta) K(\theta,\phi)\, p_T(c,r_T,\theta_T|r_0,\theta_T) - q(\theta)\, p_T(c,r_T,\theta_T|r_0,\theta_T).
\end{aligned} \tag{6}$$

This can be solved backwards in time (as $T$ increases) with initial conditions

$$p_0(c,r_T,\theta_T|r_0,\theta_T) = \left\{ \begin{array}{c} \delta_{\{(r_0,\theta_0)=(r_T,\theta_T)\}} \text{ for} c=0 \\ 0 \text{ otherwise} \end{array} \right. .$$

At any time to go, $T$, the PDF of the total data that will be sent is given by

$$p\left(c|r_0,\theta_0\right)=\sum_{\theta_T}\int p_T\left(c,r_T,\theta_T|r_0,\theta_T\right)dR_T.$$

# 4 Distribution of the Purchased Bandwidth

Given the distribution of the amount of data that will be sent, it is possible to decide if bandwidth should continue to be purchased or not. The $D_t$ denote the total data sent up to time $t$. And let $M$ be the size of the movie, while the length of the movie is $T_{Movie}$ seconds. Hence, the data remaining to be sent at time $t$ is $M-D_t$. Let $Q_{0.99}$ dente the $99^{\text{th}}$ percentilem i.e.,

$$Q_{0.99}\left(\theta,T\right)=\min\left\{q:\int_0^q\sum_{\theta_T}P_T\left(x,\theta_T|\theta\right)dx>0.99\right\}.$$

Then, as described in Section 1, the purchasing of bandwidth ceases when

$$Q_{0.99}\left(\theta_t,T_{Movie}-t\right)\geq M-D_t. \tag{7}$$

We will purchase the bandwidth from the beginning of the transmission until the point at which (7) holds. Let $PUR_t$ denote whether the bandwidth is being purchased at time $t$. This $PUR_t$ switches from 1 to 0 when $Q_{0.99}\left(\theta,T\right)\geq M-D_t$. Then

$$dD_t=\begin{cases}f\left(R_t,\theta_t\right)dt+B_{purchased}dt & \text{if } PUR_t=1\\ f\left(R_t,\theta_t\right)dt & \text{otherwise}\end{cases}.$$

Finally, let $S$ denote the total time that the bandwidth has been purchased so far, i.e.,

$$dS_t=PUR_tdt$$

The PDF the total data sent, $d$, the round-trip time, $r$, congestion level, $\theta=\theta$, whether bandwidth is purchased or not, $pur$ and the total time the bandwidth is purchased at time $t$ asa  function of initial value of $\theta$ and $r$ obeys

$$\begin{aligned}\frac{\partial}{\partial t}p_t\left(d,r,\theta,pur,s|r_0,\theta_0\right)=&-f\left(r,\theta\right)\frac{\partial}{\partial d}p_t\left(d,r,\theta,pur,s|r_0,\theta_0\right) \\ &-\frac{\partial}{\partial s}p_t\left(d,r,\theta,pur,s_t|r_0,\theta_0\right)1_{\{pur_t=1\}} \\ &+p_t\left(d,r,\theta,1,s|r_0,\theta_0\right)1_{\{pur=0\}}\delta_{\{Q_{0.99}(\theta,T_{Movie}-t)\geq M-d\}} \\ &-p_t\left(d,r,\theta,1,s|r_0,\theta_0\right)1_{\{pur=1\}}\delta_{\{Q_{0.99}(\theta,T_{Movie}-t)\geq M-d\}} \\ &-\frac{\partial}{\partial r}\left(\mu\left(r\right)p_t\left(d,r,\theta,1,s|r_0,\theta_0\right)\right)+\frac{1}{2}\frac{\partial^2}{\partial r^2}\left(\sigma^2\left(r\right)p_t\left(d,r,\theta,1,s|r_0,\theta_0\right)\right),\end{aligned} \tag{8}$$

If the initial congestion level is exactly known, then the initial conditions are

$$p_0\left(d,r,\theta,pur,s|r_0,\theta_0\right)=\delta_{\{(d,r,\theta,pur,s)=(0,r_0,\theta_0,1,0)\}}.$$

However, if only a density of the initial congestion level is known, the initial conditions are

$$p_0\left(d,r,\theta,pur,s|r_0,\theta_0\right)=\delta_{\{(d,r,\theta,pur,s)=(0,r_0,\phi,1,0)\}}p\left(\phi\right),$$

where $p\left(\phi\right)$ is the density of the initial congestion level. In either case, $p_t$ is solved forward in time.

# 5 Price of Movie Transfer

Given the distribution on the amount of time that the bandwidth is needed purchased, the price of the transmission can easily be determined. In the case that the VSP is risk neutral, the price is just the average amount of time that the bandwidth will be purchased multiplied by the cost of the bandwidth per unit time. Specifically, if the initial congestion level is known, the cost of the bandwidth is

$$C := E\left(S_K|\theta_0\right) \times \text{price of purchased bandwidth/sec},$$

and if the initial congestion level is only estimated via PDF $p\left(\theta_0\right)$, the risk neutral price is

$$C := \sum_{s_K} \int s_K P\left(s_K|\theta_0\right) p\left(\theta_o\right) d\theta_o \times \text{price of purchased bandwidth/sec.}$$

A risk neutral setting is not realistic since there is may be a significant probability that the bandwidth will need to be purchased for much longer than its mean. For example, in the example is Section 6, the mean time that the bandwidth must be purchased is 31 minutes, but a significant number of viewers would need the bandwidth for the entire 1.5 hour movie. To reduce the VSP exposure to such large losses, a risk averse utility function should be used. Letting $U\left(w\right)$ to be the utility of gaining/losing wealth $w$, the cost $C$ is such that

$$\sum_{s_K} U\left(C - s_K \times \text{price of purchased/sec}\right) p\left(s_k|\theta_0\right) = 0,$$

or, if only an estimate of the initial condition is known, the cost is such that

$$\int \sum_{s_K} U\left(C - s_K \times \text{price of purchased/sec.}\right) p\left(s_k|\theta_0\right) p\left(\theta_0\right) d\theta_0 = 0,$$

# 6 Example

Here we use data collected from a Los Angeles - San Jose connection to illustrate the algorithm We assume that the initial state $\theta$ is known and the movie size is 1.3GB. The first step is to solve (6). The left hand plot in Figure 1 shows the probability density function of the total data sent over a 1.5 hour time interval. The left hand plot in Figure 1 also shows the movie size and the first percentile. Hence, to meet the objective that 99% of viewers successfully receive the movie, extra bandwidth is required. In this case, if 0.7Mbps is purchased for the 1.5 hours, then the 99% objective is met.

However, it is not required to purchase the bandwidth for the entire movie. The left hand plot in Figure 2 shows the cumulative data sent over the course of the movie for one realization. The black dotted line marks the size of the movie. Thus, in this realization, the entire movie was sent within the first half hour. Note that there are other realizations that require more time to send the movie. Indeed, 1% of the realization are not able to complete the transmission at all.

Clearly, there is no need to continuing to purchase bandwidth once the movie has been sent. The center plot in Figure 2 shows the cumulative amount to data sent if the purchased bandwidth is only utilized until the entire movie is sent. Note that after the purchased bandwidth is no longer used, this connection was able to receive a substantial

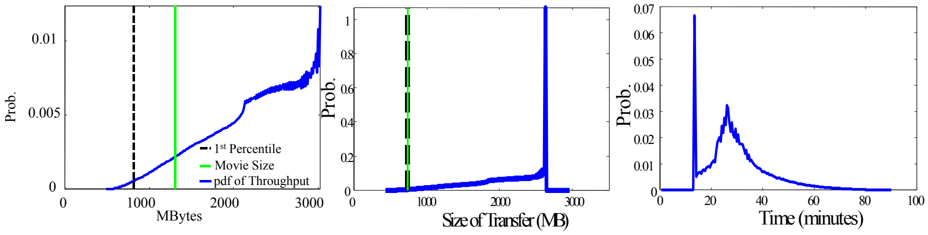

Figure 1: Left: Probability Density of the Total Data Transmitted Over the Course of the Movie. Center: The Probability Density Function of the Amount of Data That Will be Sent by the Best-effort Service in the Remaining 74 Minutes the Movie Plays. Right: The Probability Density Function of the Amount of Time the Purchased Connection is Used.

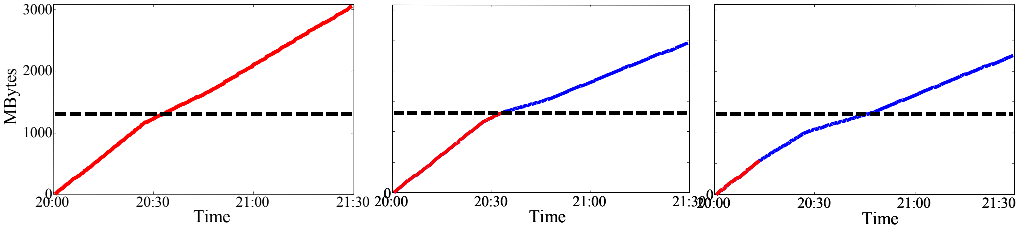

Figure 2: Sample Realizations. The left most plot shows the total data sent asa function of time if the purchased bandwidth is utilised for the entire duration of the movie. The center figure show the total data sent if the purchased bandwidth is used until the movie is completely transfered. The right most figure show the total data sent if the purchased bandwidth is utilized until it is decided that the transfer will complete with probability 0.99.

amount of data. Clearly, in this realization, this much purchased bandwidth was not required.

To conserve purchased bandwidth, the following algorithm is utilized. At time $t$, the total amount of data sent is determined along with an estimate of the current network state. The first percentile of the total data that can be sent over the best-effort connection in the remaining time is computed by solving (6). This quantity is the amount of data that will very likely be transmitted in the time remaining. If this quantity, combined with the total data sent so far, is greater than the movie size, then the purchased bandwidth is released and for the remainder of the movie is sent with only the best-effort service. A realization of this algorithm is shown in the right hand plot in Figure 2. In this case, the purchase bandwidth was only used until 8:16. Note that bandwidth is still wasted as nearly 1GB extra could have been sent beyond the size of the movie.

The center plot in Figure 1 shows the probability density function used to decide to stop purchasing the bandwidth at 8:16. The figure shows the total data sent in the remaining 74 minutes by the best-effort connection. The vertical green line is the amount of movie that remains to be sent and the vertical black line shows the first percentile. Hence, it can be concluded that 99% of the viewers will receive the rest of the movie utilizing only the best-effort connection and the purchase bandwidth is no longer required.

If this algorithm is followed, the probability density function of the amount of time the purchased bandwidth must be used can be found by solving (8). The right most plot in Figure 1 shows one such density function. Note that a very small number of users need to purchased bandwidth for the entire 1.5 hours. From this density function, the expected value of the time the bandwidth is required can be found. In this example, it was found to be 31 minutes.

# 7 Conclusion

This paper briefly shows how modeling can be used to determine the fair price of bandwidth purchased for video transmission. This approach requires accurate and easy to use models. Section 2 shows such modeling. While these models have been verified with real data, this modeling effort continues. With these models, the fair price can be found by solving (6) and (8). Unfortunately, these equations are extremely difficult to solve. However, efficient methods have been developed to approximate these equations. The resulting methods are fast enough to run in nearly real-time. These methods will be presented elsewhere.

While this paper focused pricing for bandwidth of video, there are many other application of this model approach. Essentially, any network aware application that requires prediction of bandwidth can be used methods developed here. Some such applications will be developed in future papers.

# References

[1] R. Braden, D. Clark, and S. Shenker, “Integrated services in the internet architecture: An overview,” 1994. RFC 1633.

[2] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, “An architecture for differentiated services,” 1998. RFC2475.

[3] S. Bohacek and B. Rozovskii, “A diffusion model of roundtrip time,” *Compuational Statistics and Data Analysis*, Accepted.

[4] J. Pointek, F. Shull, R. Tesoriero, and A. Agrawala, “Netdyn revisited: A replicated study of network dynamics,” Tech. Rep. CS-TR-3696, Dept. of Computer Science, University of Maryland, 1996.

[5] A. Mukherjee, “On the dynamics and significance of low frequency components of Internet load,” *Internetworking: Research and Experience*, vol. 5, no. 4, pp. 163–205, 1994.

[6] A. Acharya and J. Saltz, “A study of internet round-trip delay,” Tech. Rep. CS-TR-3736, Department of Computer Science, University of Maryland, 1996.

[7] M. S. Borella and G. B. Brewster, “Measurement and analysis of long-range dependent behavior of internet packet delay,” in *INFOCOM*, 1998.

[8] J. C. Cox, J. E. Ingersoll, and S. A. Ross, “A theory of the term structure of interest rates,” *Ecometrica*, vol. 53, pp. 385–407, 1985.

[9] L. Kleinrock, *Queueing Systems*. New York: Wiley, 1975.

[10] M. Mathis, J. Semke, J. Mahdavi, and T. Ott, “The macroscopic behavior of the TCP congestion avoidance algorithm,” *Computer Communication Review*, vol. 27, 1997.

[11] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose, “Modeling TCP throughput: A simple model and its empirical validation,” in *ACM Sigcomm*, Sept. 1998.

[12] N. Cardwell, S. Savage, and T. Anderson, “Modeling TCP latency,” in *INFOCOM 2000*, 2000.

[13] E. Altman, K. Avrachenkov, and C. Barakat, “TCP in the presence of bursty losses,” in *ACM SIGMETRICS*, June 2000.

[14] E. Altman, K. Avrachenkov, and C. Barakat, “A stochastic model of TCP/IP with stationary random losses,” in *ACM Sigcomm*, Sep. 2000.

[15] V. Misra, W. Gong, and D. Towsley, “Stochastic differential equation modeling and analysis of TCP-windowsize behavior,” in *Performance ’99*, (Istanbul, Turkey), 1999.

[16] S. Savari and E. Telatar, “The behavior of certain stochastic processes arising in window protocols,” in *IEEE Globecom*, Dec. 1999.

[17] A. Misra and T. J. Ott, “The window distribution of idealized TCP congestion avoidance with variable packet loss,” in *INFOCOM*, 1999.

[18] M. Yajnik, S. B. Moon, J. F. Kurose, and D. F. Towsley, “Measurement and modeling of the temporal dependence in packet loss,” in *INFOCOM*, pp. 345–352, 1999.

[19] R. L. Eubank., *Spline Smoothing and Nonparametric Regression.* New York: M. Dekker, 1988.

[20] B. Oksendal, *Stochastic Differential Equations : An Introduction with Applications.* New York: Springer, 1998.

[21] S. Floyd, “Connections with multiple congested gateways in packet-switched networks part 1: One-way traffic,” *Computer Communication Review*, vol. 21, pp. 30–47, 1991.

[22] M. Mathis, J. Semke, J. Mahdavi, and T. Ott, “The macroscopic behavior of the TCP congestive avoidance algorithm,” *Comput. Commun. Rev.*, vol. 27, 1997.

[23] J. Widmer and M. Handley, “Extending equation-based congestion control to multicast applications,” in *Proc of ACM SIGCOMM 2001*, 2001.

[24] M. Vojnovic and J. L. Boudec, “Some observations on equation-based rate control,” in *Proceedings of ITC-17, Seventeenth International Teletraffic Congress*, (Salvador da Bahia, Brazil), September 2001.

[25] S. Bohacek, “A stochastic model of TCP and fair video transmission,” in *Infocomm 2003.* Submitted.