*Proceedings of the 2003 Winter Simulation Conference*
*S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, eds.*

# A TWO-COMPONENT SPOT PRICING FRAMEWORK FOR LOSS-RATE GUARANTEED INTERNET SERVICE CONTRACTS

Aparna Gupta
Lingyi Zhang

Decision Sciences &
Engineering Systems
Rensselaer Polytechnic Institute
Troy, NY 12180, U.S.A.

Shivkumar Kalyanaraman

Electrical, Computer &
Systems Engineering
Rensselaer Polytechnic Institute
Troy, NY 12180, U.S.A.

## ABSTRACT

The technological advances in recent years are allowing Internet Service Providers (ISPs) to provide Quality of Service (QoS) assurance for traffic through their domains. This article develops a spot pricing framework for intra-domain expected bandwidth contract with a loss based QoS guarantee. The framework accounts for both the cost and the risks associated with QoS delivery. A nonlinear pricing scheme is used in pricing for cost recovery; a utility based options pricing approach is developed for risk related pricing. The application of options pricing in internet services provides a mechanism for fair risk sharing between the provider and the customer, and may be extended to price other uncertainties in QoS guarantees.

## 1 INTRODUCTION

The internet today mostly provides a *best-effort* service, i.e., it tries its best to push the data through from a source to a destination. In doing so it does not give any guarantees to its customers regarding the data actually reaching its destinations. Significant improvements in the network technology over the past few years are enabling the Internet Service Providers (ISPs) to incorporate better assurances on *Quality of Service (QoS)* for the traffic within their network domains. Mechanisms can then be developed so that the providers leverage on their network resources and improve utilization by pricing bandwidth appropriately and provide customers with assured services for their *inter-domain* traffic.

This article develops a spot pricing framework for *intra-domain* assured bandwidth service, specifically for expected bandwidth with a loss rate guarantee. The framework lays the foundation for pricing *inter-domain* guaranteed bandwidth for enterprise customers. A dynamic pricing scheme is employed, in which prices are generated that respond to customer demand characteristics and the current state of the network. In addition, a utility based options pricing is applied to evaluate the risks associated with the loss based QoS delivered to the customer. Deviations from the promised service is possible because the QoS experienced by each individual customer is affected by usage of network resources by other customers, over which the provider does not have complete control. An attractive feature of the framework is that it is implementable on the differentiated services architecture (*diff-serv*) and can be overlayed on schemes which are capable of providing intra-domain assured services, such as, Distributed Dynamic Capacity Contracting (Yuksel and Kalyanaraman 2002).

The article proceeds as follows. Section 2 provides a brief review of state-of-the-art for bandwidth pricing and relevant work in options pricing, as well as the advancements for supporting QoS towards the realization of assured bandwidth provision. In section 3 we describe the models for spot pricing in detail, followed by a discussion of pricing related network modeling in section4. Finally, discussions of the results and prospects for future research are given in section 5.

## 2 LITERATURE REVIEW AND BACKGROUND

### 2.1 Bandwidth Pricing

Internet pricing is a growing area of research. Until recently, *static pricing*, i.e., flat rate or time-of-the-day pricing schemes (Odlyzko 2000), has presided among providers. Despite their ease of implementation, these schemes do not react to the current state of the network, and therefore are not effective mechanisms for leveraging network resources. On the other hand, *dynamic pricing* schemes such as *Smart Market* (MacKie-Mason, Varian 1995), *Proportional Fair Pricing Schemes* (Kelly et al. 1998), *Priority Pricing* (Gupta et al.1997) takes the state of the network into account. However, these pricing strategies receive skepticism about the practicality of their implementation due to their fine time-granularity. There have also been propositions for dynamic pricing schemes on larger time scales (Gupta et al. 2002). Recently, an implementable *Pricing Over Congestion Control (POCC)* (Yuksel et al. 2002) for diff-serv architecture has been proposed which can be overlaid on the congestion control framework proposed by Harrison, et al. (2001) and provides a range of fairness in rate allocation by using pricing as a tool.

A closely related research in the past years has been under way in pricing for telecom bandwidth contracts. These contracts are usually longer term contracts than those discussed in the internet pricing literature, ranging in month durations. For end-to-end bandwidth pricing, the role of geographical arbitrage is investigated using different approaches, which include the application of compound option techniques (Cheliotis 2001; Keppo et al. 2002; Upton 2002).

Real options or contingent claim analysis (CCA) is used to address an increasing variety of problems in finance. In our work, real options concepts help to capture the stochastic nature of QoS guarantees, as often observed in the internet technologies. A great deal of theoretical work and practical application for real options analysis is found in valuation and decision making in various areas. Some examples, though far from exhaustive, are natural resources (Paddock, Siegel and Smith 1988), investment analysis and firm behavior (Dixit 1989; Pindyck 1991), R&D (Pennings and Lint 1997), manufacturing (Bengtsson 2001; Kamrad 1995), real estate and leasing (Paddock 1988; Grenadier 1995; Trigeorgis 1996). (See Lander et al. (1998) for a comprehensive review of real option valuation and their applications.) Real options have recently been used in the pricing for optional calling plan contracts in the telephone industry by valuing the uncertainty in accumulated call usage (Choi, Kim and Kim 2002).

In real options framework, since the underlying assets usually lack liquidity, the prices are often assumed to be exogenously driven by some associated liquid assets such as, output from a potential investment (Pindyck 1991) or products from a manufacturing facility (Kamrad 1995). Competitive equilibrium arguments are used to establish the value of the underlying assets (Grenadier 1995), which would require an implied assumption of the existence of competitive markets for the underlying assets. Another alternative is based on utility maximization assuming certain forms of utility functions. For example, Henderson and Hobson (2002) derive the values of options with a non-traded underlying asset added to the classical Merton's investment model (Merton 1969) for a power utility function.

### 2.2 Technology to Support QoS

In contrast with a leased line or a circuit-switching setting, traffic is not perfectly isolated in packet-switching due to the nature of scheduling mechanisms employed (Firoiu et al. 2002; Stoica et al. 1999; Zhang et al. 2000). Close monitoring and traffic engineering mechanisms are set in place to effect the QoS delivery in the internet.

QoS deployment in multi-domain, IP-based inter-networks has been an elusive goal partly due to complex deployment issues (Huston 2000). Therefore, from an architectural standpoint, contemporary QoS research has recognized the need to *simplify* and *de-couple* building blocks to promote implementation and inter-network deployment. The int-serv and RTP work (Schulzrinne et al. 1997; Braden et al. 1994) de-coupled end-to-end support from network support for QoS. RSVP de-coupled inter-networks signaling from routing. MPLS (Rosen et al. 2001) de-coupled forwarding mechanisms from the routing control plane, leading to traffic engineering capabilities (Awduche et al. 2002). The diff-serv services (Blake et al. 1998; Clark and Feng 1998) and core-stateless fair queuing (CSFQ) (Shenker et al. 1997) further simplified core architecture and moved data plane complexity to the "edges", and allowed a range of control plane options (Awduche et al. 2002; Durham et al. 2000). Therefore, concepts are being developed to address the challenge of provisioning QoS assurances at various levels – management of packets, configuration of inter-networks, and service delivery modes to customers (Giordano et al. 2003; Cortese et al. 2003; Engle et al. 2003; Mykoniati et al. 2003). Pilot studies are in progress that test these concepts (Roth et al. 2003). There has also been substantial empirical work in internet traffic monitoring and characterization and network performance analysis (Paxon 1999; Yajnik et al. 1999; SLAC; CAIDA; NLANR).

## 3 SPOT PRICING FRAMEWORK

Network performance can be defined in terms of a combination of its bandwidth, delay, delay-jitter and loss properties. Based on these performance measures, QoS guarantees can be stated in deterministic or probabilistic terms.

In this article, we will focus on pricing for an expected level of bandwidth with loss rate guarantees. Provision of QoS guaranteed contract is made at an access (edge) or exchange point. Such models implemented at the access and/or exchange points of different domains will allow the creation of inter-domain service assurance to the customers. Figure 1 shows the basic intra-domain bandwidth pricing setup.

Our spot pricing scheme consists of two major components. A nonlinear pricing scheme is employed to capture the cost factors in providing the expected bandwidth requirements within the contracted QoS. In order for a provider

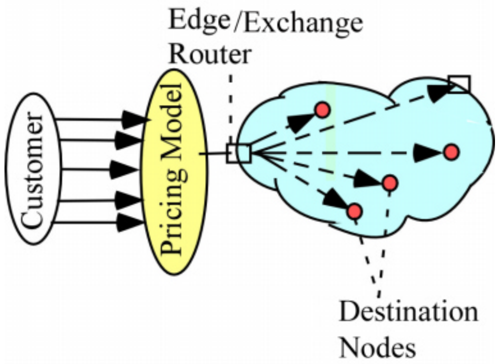


Figure 1: Basic Pricing Setup Implemented at an Access Point

to provide stronger QoS guarantees, he has to bear extra infrastructural cost for network monitoring, routers, more intelligent scheduling algorithms, etc., in addition to the cost of laying and lighting fiber. However, even with the necessary infrastructure in place, internet QoS can only be guaranteed in stochastic terms for the reasons described earlier (section 2.2). Therefore, an option-based approach is introduced to price the risk of deviations in the QoS experienced by the customers, specifically in terms of data losses.

### 3.1 Pricing to Recover Cost

Our first objective in pricing is cost recovery. We adopt a nonlinear pricing approach to this end. The term *nonlinear pricing* refers to a pricing scheme where the tariff is not proportional to the quantity purchased and the marginal prices for successive purchases decreases (Wilson 1993). Unlike a linear or uniform pricing scheme, in nonlinear pricing prices are chosen according to the inverse of the price elasticities for the incremental quantity purchased, and therefore are decreasing along the customer's demand for a typical demand function. Prices are also set above the marginal costs in order to recover the provider's full operating and capital expenses. Considerations of cost, competitive pressures and profits constitute the major motivations for favoring a nonlinear pricing scheme. Nonlinear pricing is particularly relevant in industries where large fixed cost is involved, as by favorable pricing a provider can attract customers with large demand and thus improve utilization of its capacity and sufficiently recover the fixed cost.

As a well known example of nonlinear pricing model, *Ramsey pricing*, has been widely popular in the telecommunication and power sectors (Wilson 1993; Dolan and Simon 1997). It produces an efficient tariff design in situations where due to either regulation or competition, revenue sufficient to only recover the provider's total costs are achievable. In particular, the price schedule obtained from Ramsey pricing maximizes a commonly used measure for the aggregated customers' benefits, i.e., the total *customer surplus*, given by

$$\mathrm{CS}(q) = \int_{p(q)}^{\infty} N(p,q)dp, \tag{1}$$

where $p(q)$ is the marginal price for the $q^{th}$ unit purchased, and $N(p,q)$ is the *demand profile* of a population, defined as the number or fraction of customer base that will buy at least $q$ units at the marginal price $p(q)$. The optimal price schedule $p(q)$ is then given by the following *Ramsey rule* (Wilson 1993):

$$\frac{p(q)-c(q)}{p(q)} = \frac{\alpha}{\eta(p(q),q)}, \tag{2}$$

where $c(q)$ is the marginal cost for the $q^{th}$ unit, and $\eta(p(q),q)$ is the elasticity of the demand profile. The Ramsey number $\alpha$ is the fraction of the monopoly profit margin common to all units of customers' purchases that is needed for cost recovery, and is an indicator of the monopoly power of the provider.

In our earlier work (Gupta et al. 2002), Ramsey pricing was applied to price expected bandwidth contracts. Different characteristics of demand profiles and competitive nature of the provider were considered, and prices were analyzed for different scenarios.

### 3.2 Pricing the Risk

Provision of loss based QoS guaranteed service is intrinsically risky due to the uncertainties caused by competing traffic in the internet. The final outcome of a service delivered to the customer may turn out to be in favor of or against the provider, i.e., the provider may or may not deliver the loss based QoS as promised. Consider a simple example of a service contract where the loss guarantee is defined as: "The total data loss over the contract duration of 1 hour starting from 9 : 00 a.m., June 13, 2003 does not exceed 10 MB." We say that the future outcome is in favor of the provider if at the end of the contract less than 10 MB of the customer's data is lost, and that it is against the provider otherwise.

Options pricing techniques appear as a natural tool for evaluating the risky nature of the service, as the value of the service is contingent on future outcomes. Pricing the risk appropriately will let the risks be fairly borne by the provider and the customer. In the above example, the service may be viewed as a simple "knock-out" type *barrier option* on the total data loss with an upper barrier of 10 MB. A *knock-out barrier option* is an option that only pays off when the prescribed barrier is *not* reached by an underlying uncertainty; the option becomes worthless if the underlying uncertainty reaches the barrier. The option is priced by a hedging portfolio argument, where the price is equal to the expectation of the payoff under a transformed risk neutral measure.

The underlying risks in our pricing framework are unhedgeable; therefore, utility based techniques for options pricing in incomplete markets need to be employed. In particular, we consider pricing from the provider's perspective using the concept of *state price density* (SPD), and evaluate the monetary "reward" for the favorable risks to the provider, which then becomes the second component of the price of the contract.

For pricing the risks in the loss processes, we construct a state price density to describe a representative provider's preferences for the future outcomes for losses. The state price density translates into a risk neutral measure, $Q$; if $Y_t$ is the *payoff* from the loss process at time $t$, the options price of the loss process is given by

$$V = E_Q[\int_0^T Y_t dt]. \quad (3)$$

$Y_t$ may take different forms depending on how the *payoff* is defined.

Following similar arguments, in scenarios when the provider does not deliver the loss based QoS as promised, a "penalty" oriented pricing may be developed from the customer's perspective. However, penalty oriented pricing would require inclusion of the customer's preferences, as well as negotiation power of the two parties.

# 4 MODEL DEFINITION AND ASSUMPTIONS

In this section we describe our spot pricing model for the intra-domain assured bandwidth contracts. We are interested in how the price for a contracted service for an individual customer is determined, which is driven by the interactions of traffic from the customer and the background traffic in the network from all other sources. We model the aggregate of the background traffic as a single process and define it as the *Aggregate*. An aggregate approach is used instead of the alternative of source based model due to issues regarding scalability and computational cost (Paxon et al. 2001). For pricing purpose the network is abstracted by a single link with a certain capacity.

A customer purchases bandwidth contracts of a fixed duration $T$ for simple and immediate file transfer applications. Upon arrival the customer announces its volume and loss rate requirements to the provider. The *Asked Capacity* is then obtained by dividing the expected total volume requested by the contract duration $T$. The customer is admitted into the network only when the *Asked Capacity* (in Kbps) is lower than the *Available Capacity* of the network at the time of arrival. When the customer is accepted, a contract for bandwidth with required service levels is created; the customer is assigned the Asked Capacity and after time $T$ the customer releases the capacity and leaves the system. The *Available Capacity* is updated with every entry and exit of customers. A price schedule $p(q)$ for the expected bandwidth is generated using the Ramsey pricing model described in section 3.1 based on the demanded capacity, $q$. $q$ is a scaled measure in the range $[0, 1]$ and is defined as the ratio of the *Asked Capacity* to the current *Available Capacity*. The first component of the price of the contract, $P(q)$, is computed by integrating over the marginal price.

## 4.1 Modelling the Loss Process

In our framework, options pricing technique is employed to price the risk related with losses of the customer's data. This is the second component of the price of the contract. The losses are essentially determined by the customer's traffic and its interaction with the *Aggregate*. We next describe our model for this interaction.

### 4.1.1 The Individual Traffic $I_t$

Traffic from the customer is modelled on a flow basis, described by its arrival rate and transfer parameters, including file sizes and transfer times. Literature on data analysis of internet traffic describes flow arrivals to follow a time dependent Poisson process, and file sizes and transfer times to be best represented by heavy-tailed distributions (Paxon 1995; Paxon et al. 2001; Crovella and Bestavros 1997). We model the arrivals of files from the customer by a Poisson process at a rate of $\lambda = 5/min$ averaged over a day. Arrivals are time dependent; based on historical data (NLANR 2002), we assume that 70% of the arrivals happen between 7 a.m. and 5 p.m., 20% between 5 p.m. and 11 p.m. and the rest 10% happen between 11 p.m. and 7 a.m. Pareto distribution with probability density function of the form

$$P(x) = \frac{ab^a}{x^{a+1}},$$

are used to model the heavy-tailed distributions of files sizes and transfer times, following the internet traffic data analysis literature (Paxon et al. 2001; SLAC; CAIDA; NLANR). The parameters for file size distribution are $a = 1.05, b = 1.2Kb$. For the transfer time distribution, $a = 1.2$, and the scale parameter $b$ is dependent on the size of file being transferred; for file sizes smaller than 2.3 KB, between 2.3 KB and 20 KB, and larger than 20 KB, $b$ takes the value of 0.01, 0.4 and 0.95 second, respectively. These parameters are kept fixed across customers for simplicity.

Combining the file arrival rates, file sizes and transfer times, an arrival curve and a service curve for the customer can be obtained (Figure 2a). At a given time $t$, we define *data in-transit*, $I_t$, as the difference between the arrival curve and the service curve. $I_t$ is the amount of the customer's data in the network, i.e., the data susceptible to loss, at time $t$ (Figure 2b).

#### 4.1.2 The Aggregate $A_t$

The *Aggregate* depicts the current state of the network. The modelling of the aggregate is intended to capture two significant characteristics of the aggregated internet traffic, *diurnal pattern* and *self-similarity* (Paxon et al. 2001; SLAC).

A clear diurnal pattern is observed in the internet traffic, which is believed to relate to human activities starting to rise around 8–9 a.m., peaking around 3–4 p.m. and declining around 5 p.m. when a business day ends. In addition, a relatively moderate peak is often observed at weekends than during weekdays. We use a sinusoidal curve with a period of 24 hours and an appropriate phase to model this diurnal pattern. The amplitude and the average of the sinusoidal curve for weekdays are chosen to be 5 GB and 5 GB, 3.5 GB and 4.25 GB for weekends, respectively.

Self-similarity in network traffic has been extensively discussed in the network literature (Paxon et al. 2001, Crovella and Bestavros 1997; Paxon 1995) for its significant influence on network performance and the consequent implications on network modelling and implementation. A class of so-called *fractional processes*, including for example, general fractional ARIMA (FARIMA) models, fractional Brownian motion, or fractional Gaussian noise (FGN), has been widely used to generate self-similar traffic in network simulation. We use the FGN in our model due to its simplicity of implementation among this class of self-similar processes. The FGN is usually generated based on its power spectrum given by

$$f(\lambda; H) \quad = \quad A(\lambda; H)[|\lambda|^{-2H-1} + B(\lambda; H)], \qquad (4)$$

for $0 < H < 1$ and $-\pi \leq \lambda \leq \pi$, where

$$A(\lambda; H) \quad = \quad 2sin(\pi H)\Gamma(2H+1)(1-cos\lambda), \qquad (5)$$

$$B(\lambda; H) \quad = \quad \sum_{k=1}^{\infty}[(2\pi k+\lambda)^{-2H-1} + (2\pi k-\lambda)^{-2H-1}],$$

where $H$ is the *Hurst parameter* which describes the degree of self-similarity of the process, and $0.5 < H < 1$ (Ledesma and Liu 2000). We use a linear approximation approach in generating the FGN introduced by Ledesma and Liu (2000), which according to them generates FGN with comparable accuracy as Paxon's method (Paxon 1995a), but at significantly less computational expense.

Therefore, at any given time $t$, we define the aggregate process $A_t$ as a sinusoidal function imposed with an appropriately scaled FGN process, i.e.,

$$A_t \quad = \quad Rsin(2\pi f t + \theta) + \overline{A_t} + Z_t, \qquad (6)$$

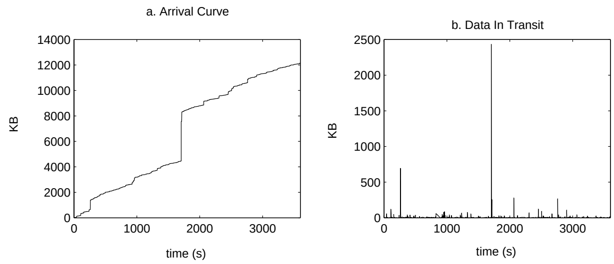


Figure 2: Customer Data Flow over Contract Duration: a. Arrival Curve; b. Data in Transit $I_t$

where $R$, $f$, $\theta$ and the average level of the aggregate $A_t$, $\overline{A_t}$ (Figure 3) are described above, and $Z_t$ is the scaled FGN process. Different values of the Hurst parameter $H$ of the FGN was simulated in the range of $0.7-0.95$ and the result shown here has $H = 0.8$.

#### 4.1.3 The Loss Process $L_t$

Data in-transit along with the state of the network are indicators of data loss. Given $I_t$ and $A_t$ as described above, the loss process is then modelled as a 2-state Markov process with 1 representing a state where losses happen and 0 representing a loss free state, the transition probabilities depending on $I_t$ and $A_t$. It is assumed that when the network is in a highly congested state, as indicated by a high value of $A_t$, and if there is sufficient amount of the customer's data in the network, losses will happen with certainty. On the other hand, when the network is extremely under utilized, there will be zero data loss. Between these two extremes losses happen with some nonzero probability. It is understood that although errors in data transmission and network failures may cause losses, losses of this nature are not accounted for in the contract (SLAC).

Two threshold $T^U$ and $T^L$ levels for the total amount of data in the network, i.e., $I_t + A_t$, as well as an upper threshold $T^U_{I_t}$ for $I_t$ are set. Therefore, the transition matrix $P_{ij}$, $(i, j = 0, 1)$ is given by

$$P_{ij} = \begin{cases} \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, & \text{if } A_t + I_t \leq T^L; \\ \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix}, & \text{if } T^L < A_t + I_t \leq T^U; \\ \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, & \text{if } A_t + I_t > T^U \text{ and } I_t \geq T^U_{I_t}, \end{cases} \qquad (7)$$

and $0 < p_{ij} < 1$ for all $i, j$. In our simulation, $p_{01} = 0.5\%$ and $p_{11} = 0.8\%$, respectively. The threshold values $T^H$ and $T^U$ are set as 1.2 and 0.5 times the peak value of the

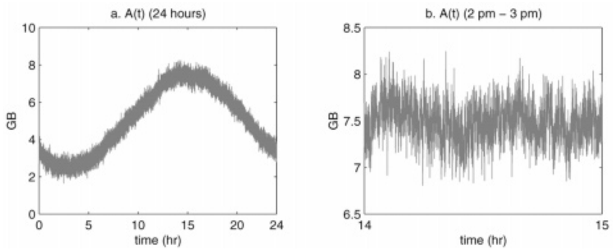


Figure 3: $A_t$: a. 24 Hours; b. 1 Hour (2 pm-3 pm)

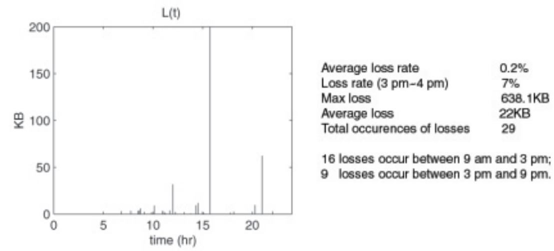


Figure 4: 24 Hour Variation of $L_t$

aggregate process given by the sinusoidal function of $A_t$ (Eqn. 6). It should be noted that the parameters used in our simulation are only representative values; other time variant choices can be easily accommodated in the framework. For simplicity, it is further assumed that when $L_t$ is in a loss state, the customer's data in transit, $I_t$, is lost, i.e., $L_t = I_t$ when $L_t$ is in state 1. A realization of the $L_t$ process in a 24 hour period is given in Figure 4. $L_t$ shows high burstiness, with the maximum size of loss much larger than the average loss. As expected, losses happen more frequent when $A_t$ is high; a comparison of $L_t$ and the corresponding $I_t$ indicates a positive correlation between $L_t$ and large values of $I_t$.

## 4.2 Pricing for Loss Guaranteed Service

In this section we apply the options based approach described in section 3.2 to price a demonstrative contract for a deterministically defined loss-rate guarantee. The contract mandates that the maximum loss rates monitored at minute intervals are less than 0.5% over the contract duration of 1 hour. This requires monitoring and testing at minute intervals. Since our system evolves at second intervals, aggregation per minute is required. This implies that the pricing will have a flavor of a combination of an Asian and a knock-out barrier option.

The per minute loss rate for the $t^{th}$ minute from the start of the contract, $l_t$, is obtained by

$$l_t = \frac{\sum_{i=1}^{60} L_{t,i}}{\sum_{i=1}^{60} I_{t,i}}, \tag{8}$$

where $I_t$ and $L_t$ were defined in the previous section. Let $S^u$ be the upper barrier for $l_t$ ($S^u = 0.5\%$), and $N$ is the total number of minutes within the contract duration $T$. Given $l_t$, the *payoff* of the service may be defined as

$$Y_t = I_{(0,1)}(l_t)|l_t - S^u|, \tag{9}$$

where $I_{(0,1)}(\cdot)$ is an indicator function defined as

$$I_{(0,1)}(l_t) = \begin{cases} 1, & \text{if } l_t < S^u; \\ 0, & \text{otherwise,} \end{cases} \tag{10}$$

for $t = 0, 1, \ldots, N$. Following Eqn.4 the price of the contract is given as

$$V = E_Q[\sum_0^N (I_{(0,1)}(l_t)|l_t - S^u|)], \tag{11}$$

where $Q$ is the risk neutral measure resulting from the provider's state price density. The state price density (SPD) is described next.

### 4.2.1 Definition of the Provider's SPD

A *state price*, $p_s$, is defined in financial terms as the price of one dollar to be obtained if state $s$ occurs in the future. The normalized state price constructed by

$$Q_s = \frac{p_s}{\sum_s p_s}, \tag{12}$$

is often referred to as the state price density (SPD). The SPD is a basic economic construct for a (representative) economic agent, and is used to describe the agent's subjective preferences for future outcomes. The basic construct of an SPD is used for pricing of assets governed by the selected sources of uncertainty. The pricing equation can be viewed as an expectation under a transformed measure defined by the SPD, termed as a risk neutral measure.

For our pricing purposes, we construct a representative provider's SPD based on the outcomes of the loss process, $L_t$. The loss process is taken to be the special rudimentary source of uncertainty, which the provider would be held responsible for. The SPD also plays the role of transforming the risks in the loss process into appropriate dollar values.

Without defining a specific form for the provider's utility function with regards to $L_t$, we infer the general properties of the SPD function based on the following observations.

1. The payoff, $Y_t$, to the provider can only be a decreasing function of $L_t$;
2. The provider would expect that there is no loss at all during most of the contract duration, and that losses would more likely happen within a small to moderate range, although there is a non-zero probability of extremely large losses to occur.

Table 1: Comparison of Price Variation During a Day

| Time | S1 | S2 | S3 | S4 |
|---|---|---|---|---|
| 12 a.m. | 1.488 | 1.500 | 1.490 | 1.465 |
| 3 a.m. | 1.490 | 1.500 | 1.495 | 1.485 |
| 6 a.m. | 1.485 | 1.500 | 1.495 | 1.485 |
| 9 a.m. | 1.437 | 1.470 | 1.455 | 1.460 |
| 12 p.m. | 1.425 | 1.443 | 1.431 | 1.416 |
| 3 p.m. | 1.444 | 1.433 | 1.420 | 1.395 |
| 6 p.m. | 1.464 | 1.458 | 1.450 | 1.427 |
| 9 p.m. | 1.471 | 1.500 | 1.475 | 1.460 |

An exponentially decaying SPD, with the mean of $\theta$ ($\theta = 0.2$), to make the SPD decay fast, would reward the low loss levels. The SPD may also take alternative forms besides being a strictly decreasing function of $L_t$. For example, the provider may consider it fit to be rewarded as long as $L_t$ is below a certain threshold level. In this case, the SPD may peak at a positive $L_t$ value before it starts to decay for moderately large values of losses. In practice, the SPD has to be estimated from the provider's, or an aggregate of providers', response to different scenarios of losses occurring.

### 4.3 Simulation Results and Discussion

We simulated the options based pricing for the above contract at different times of a day. A sample size of 20 is used for computing the expectations. A comparison of prices in different simulation scenarios is given in Table 1.

The following observations were obtained from the simulation results:

1. Due to the fast decay of the SPD, the price is dominated by zero to small losses, which is expected to occur more often when there is less traffic in the network. As a result, the price vary in a pattern reverse to that of $A_t$, with a low around 3 p.m. and peak around 3 a.m. [S1: baseline scenario with the original simulation parameters]. Similar explanation applies to the effects of increasing the thresholds, $T^L$ and $T^U$, which may correspond to a network upgrade [S2: $T^U = 6.25$GB, $T^L = 3.75$GB].
2. Lowering the loss threshold for $I_t$, $T^U_{I_t}$, strengthens the price variation, without changing its general pattern [S3: $T^U_{I_t} = 99^{th}$ percentile of $I_t$].
3. A larger $\theta$ of the SPD smoothes the price variation, as it gives lower weight to small losses [S4: $\theta$ of the SPD = 1]. Except for the price at 9 a.m., it generates lower prices than in the baseline scenario.

We developed the options based pricing for a simple deterministic loss guaranteed contract. However, the pricing approach can be enhanced to accommodate more realistic stochastic loss guaranteed contracts, such as a contract defined as "the one-way packet loss rates at minute intervals are guaranteed to be less than 0.05% for 95% of the contract duration $T$" (Bouras et al. 2002).

## 5 CONCLUSION AND FUTURE WORK

We have developed a two-component spot pricing framework for intra-domain expected bandwidth contracts with a loss based QoS guarantee. A nonlinear pricing scheme is used in pricing for cost recovery. By constructing a state price density for a representative provider, a utility based options pricing approach is developed to price the risky aspects of the loss based QoS guarantee.

QoS delivery in the internet has an inherent risky nature. The options based pricing approach was introduced to capture the risky aspects in loss based QoS assured service. The pricing approach described here can be applied to more complicated, stochastically defined loss assured contracts. In this article, the price is decided from the provider's perspective. A similar approach may also be used for penalty determination from the customer's side.

Further research would also follow different methods by which QoS guarantees in the internet can be defined. The options based pricing approach may be extended to cover other aspects of QoS, for example, delay and delay-jitter, and the price interactions when multiple QoS guarantees are present can be investigated. Forward contracts may be developed based on the spot pricing framework described here. Methods will need to be developed to use the spot pricing framework at an access/exchange point of the network to create inter-domain contracts.

## REFERENCES

Awduche, D., A. Chiu, A. Elwalid, I. Widjaja, and X. Xiao. 2002. Overview and principles of Internet Traffic Engineering. *IETF Internet RFC 3272*.

Bengtsson, J. 2001. Manufacturing flexibility and real options: A review. *International Journal of Production Economics*. 74: 213–224.

Blake, S., D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss. 1998. An architecture for differentiated services. *IETF Internet RFC 2475*.

Braden, R., et al. 1994. Integrated services in the internet architecture: an overview. *IETF Internet RFC 1633*.

Bouras, C., M. Campanella and A. Sevasti. 2002. SLA definition for the provision of an EF-based service. Available on line via `<http://www.dante.net/sequin/presentations/CQR_final_8-3-2002.pdf>` [accessed July 03, 2003].

CAIDA. 2003. `<http://caida.org/analysis>`. [accessed June, 2003].

SLAC. 2003. Stanford Linear Accelerator Center. <http://slac.stanford.edu>. [accessed June, 2003].

Cheliotis, G. 2001. *Structure and Dynamics of Bandwidth Markets*. PhD thesis. N.T.U. Athens.

Choi, H., I. Kim, and T. Kim. 2002. Contingent claims valuation of optional calling plan contracts in telephone industry. *International Review of Financial Analysis.* 11: 433–448.

Clark, D., and W. Feng. 1998. Explicit allocation of best-effort packet delivery service. *IEEE/ACM Transactions on Networking (TON)*. 6 (4): 362–373.

Cortese, G., R. Fiutem, P. Cremonese, S. D'Antonio, M. Esposito, S. P. Romano, and A. Diaconescu. 2003. CADENUS: Creation and deployment of end-user services in premium ip-networks. *IEEE Communications Magazine*. 41 (1): 54–60.

Crovella, M., and A. Bestavros. 1997. Self-similarity in world wide web traffic: Evidence and possible causes. *IEEE/ACM Transactions on Networking (TON)*. 5 (6): 835–846.

Dixit, A. 1989. Entry and exit decisions under uncertainty. *The Journal of Political Economy*. 97 (3): 620–638.

Dolan, R. J., and H. Simon. 1997. *Power Pricing - How Managing Price Transforms the Bottom Line.* New York: Free Press.

Durham, D., et al. 2000. The COPS (common open policy service) protocol. *IETF Internet RFC 2748*.

Edell, R. J., N. McKeown, and P. P. Varaiya. 1995. Billing users and pricing for TCP. *IEEE Journal on Selected Areas in Communications*. 13 (7): 1162–1175.

Engel, T., et al. AQUILA: Adaptive resource control for QoS using an ip-based layered architecture. *IEEE Communications Magazine*. 41 (1): 46–53.

Firoiu, V., J.-Y. Le Boudec, D. Towsley, and Z.-L. Zhang. 2002. Theories and models for internet quality of service. *Proceedings of the IEEE*. 90 (9): 1565–1591.

Floyd, S., and V. Paxon. 2001. Difficulties in simulating the Internet. *IEEE/ACM Transactions on Networking (TON)*. 9 (4): 392–403.

Giordano, S., et al.. 2003. Advanced QoS provisioning in the ip networks: The european premium ip projects. *IEEE Communications Magazine*. 41 (1): 30–36.

Grenadier, S. R. 1995. Valuing lease contracts: A real-options approach. *Journal of Financial Economics*. 38 (3): 297–331.

Gupta, A., S. Kalyanaraman, R. Pradhan, and M. Aboulfadl. 2002. A spot pricing framework to enable pricing and risk management of inter-domain assured bandwidth services. *Proceedings of the 2002 Winter Simulation Conference*. 1515–1523. E. Yilcesan, C. -H. Chen, J. L. Snowdon, and J. M. Charnes, eds.

Gupta, A., D. O. Stahl, and A. B. Whinston. 1997. *Internet Economics*, chapter Priority pricing of Integrated Services networks. 323–352. Boston, MA: MIT Press.

Harrison, D., S. Kalyanaraman, and S. Ramakrishnan. 2001. Overlay band-width services: Basic framework and edge-to-edge closed-loop building block. Poster in *SIGCOMM*.

Henderson, V., and D. G. Hobson. 2002. Real options with constant relative risk aversion. *Journal of Economic Dynamics & Control*, 27: 329–355.

Huston, G. 2000. *Internet Performance Survival Guide: QoS Strategies for Multiservice Networks*. Hoboken, NJ: John Wiley.

Kamrad, B. 1995. A lattice claims model for capital budgeting. *IEEE Transactions on Engineering Management.* 42 (2): 140–149.

Kelly, F. P., A. K. Maulloo, and D. K. H. Tan. 1998. Rate control in communication networks: Shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*. 49: 237–252.

Keppo, J., S. Rinaz, and G. Shah. 2002. Network pricing with continuous uncertainties and multiple QoS classes. Available on line via <http://www-personal.engin.umich.edu/˜keppo/QoSPricing.pdf>. [accessed March, 2003].

Lander, D. M., and G. E. Pinches. 1998. Challenges to the practical implementation of modelling and valuing real options. *The Quarterly review of Economics and Finance*. 38: 537–567.

Ledesma, S., and D. Liu. 2000. Synthesis of fractional Gaussian noise using linear approximation for generating self-similar network traffic. *ACM SIGCOMM Computer Communication Review*. 30 (2): 4–17.

Merton, R. C. 1969. Lifetime portfolio selection under uncertainty: The continuous time case. *Review of Economics and Statistics*. 51: 247–257.

Mykoniati, E., et al. 2003. Admission control for providing QoS in diffserv ip networks: The TEQUILA approach. *IEEE Communications Magazine*. 41 (1): 38–44.

NLANR. 2002. National laboratory for applied network research. <http://www.nlanr.net/NA/Learn/daily.html>. [accessed June, 2003].

Odlyzko, A. M. 2000. Internet pricing and history of communications. AT&T labs.

Paddock, J., D. Siegel, and J. Smith. 1988. Option valuation of claims on real assets: The case of offshore petroleum leases. *Quarterly Journal of Economics*. 103: 83–94.

Paxon, V. 1995a. Fast, approximate synthesis of fractional Gaussian noise for generating self-similar network traffic. *Computer Communications Review*. 27: 5–18.

Paxon, V., and S. Floyd. 1995. Wide area traffic: the failure of Poisson modelling. *IEEE/ACM Transactions on Networking (TON)*. 3 (3): 226–244.

Paxon, V. 1999. End-to-end Internet packet dynamics. *IEEE/ACM Transactions on Networking (TON)*. 7 (3): 277–292.

Pennings, E., and O. Lint. 1997. The option value of advanced R&D. *European Journal of Operational Research*. 103: 83–94.

Pindyck, R. S. 1991. Irreversibility, uncertainty, and investment. *Journal of Economc literature*. XXIX: 1110–1148.

Rosen, E., et al. 2001. Multiprotocol label switching architecture. *IETF Internet RFC 3031*.

Roth, R., et al. 2003. Ip QoS across multiple management domains: Practical experiences from pan-European experiments. *IEEE Communications Magazine*. 41 (1): 62–69.

Schulzrinne, H., et al. 1997. RTP: A transport protocol for real-time applications. *IETF Internet RFC 1889*.

Shenker, S., C. Partridge, and R. Guerin. 1997. Specification of guaranteed quality of service. *RFC 2212*.

Stoica, I., et al.1999. Per hop behaviors based on dynamic packet states. Available on line via `<http://www-2.cs.cmu.edu/~istoica/DPS/draft.txt>`. [accessed March, 2003].

Trigeorgis, L. 1996. Evaluating leases with complex operating options. *European Journal of Operational Research*. 91 (2): 315–329.

Upton, D. 2002. Modelling the market for bandwidth. Available online via `<http://www.statslab.cam.ac.uk/~dsu20/Papers/essay.pdf>`. [accessed March, 2003].

Wilson, R. B. 1993. *Nonlinear Pricing*. New York: Oxford University Press, Inc.

Yajnik, M., S. Moon, J. Kurose, and D. Towsley. Measurement and modelling of the temporal dependence in packet loss. 1999. *Proceedings of the INFOCOM 1999*. 1: 345–352.

Yuksel, M., and S. Kalyanaraman. 2002. Distributed dynamic capacity contracting: A congestion pricing framework for Diff-Serv. *Proceedings of IFIP/IEEE International Conference on Management of Multimedia Networks and Services (MMNS)*.

Yuksel, M., S. Kalyanaraman, and A. Goel. 2002. Congestion pricing overlaid on edge-to-edge congestion control. *GlobeCom (to appear)*.

Zhang, Z. L., Z. Duan, and Y. T. Hou. 2000. Virtual time reference system: A unifying scheduling framework for scalable support of guaranteed services. *IEEE Journal on Selected Areas in Communication*. 18 (12): 2684–2695.

## AUTHOR BIOGRAPHIES

**APARNA GUPTA** is an Assistant Professor in the Department of Decision Sciences and Engineering Systems at Rensselaer Polytechnic Institute.

**SHIVKUMAR KALYANARAMAN** is an Associate Professor in the Department of Electrical, Computer and Systems Engineering at Rensselaer Polytechnic Institute.

**LINGYI ZHANG** is a Ph.D. candidate in the Department of Decision Sciences and Engineering Systems at Rensselaer Polytechnic Institute.