A

Pricing in Bandwidth Sharing Systems

SHARAD BIRMIWAL, University of Waterloo
RAVI R. MAZUMDAR, University of Waterloo
SHREYAS SUNDARAM, University of Waterloo

In this paper we study pricing when the bandwidth or server is shared according to processor sharing. Specifically, we analyze post-payment and pre-payment (or payment on arrival) schemes in three pricing frameworks: fixed-rate pricing, VCG (Vickrey-Clarke-Groves) auction based pricing, and congestion based pricing for users with logarithmic utilities. We show that in the absence of QoS constraints, the network operator can earn unbounded profits and thus there is a need to devise schemes where users are only charged if they are given a minimum rate. We obtain explicit characterizations for mean user payments and the operator's mean revenue for these frameworks. We also analyze price volatility via the second moments of the above implementations of arrival-based payments and post-payments. The volatility reflects the confidence in mean revenue for the operator and expected charges for a user. We present conditions under which a pre-payment mechanism is preferable over a post-payment mechanism. We also show that the same analysis can be applied to a scenario with admission control where each entering user is guaranteed a minimum service rate.

1. INTRODUCTION

The pricing of resource usage for users or jobs is critical for the deployment of sufficient resources in a server farm or for providing sufficient bandwidth on link. The desirable properties of pricing schemes is that they are easy to implement, transparent, predictable in revenue, provide incentive compatibility, and lead to an efficient utilization of resources. There are many pricing frameworks that can be considered. In this paper we restrict ourselves to three scenarios: 1) fixed rate pricing, 2) an incentive compatible framework that corresponds to a price set by a Vickrey-Clarke-Groves (VCG) auction mechanism, and 3) a congestion based framework where the prices correspond to the Lagrange multipliers. As we will show, when the users have logarithmic utilities for their allocated bandwidth, these schemes have the interpretation of being viewed as constant, linear, or quadratic (in the bandwidth) pricing structures and hence can be viewed as ways of approximating non-linear pricing schemes [Wilson 1997] that can be tailored to different criteria for performance or revenue sensitivity.

The way capacity of a resource is shared plays an important role in determining performance and robustness. A key metric is that of fairness and this involves some notion of objective functions or utilities. In this paper we assume that the server capacity or bandwidth is allocated according to processor sharing. The processor sharing discipline is an egalitarian bandwidth sharing mechanism where the server's capacity is shared equally between all users present in the system; there is no queueing involved and a new arriving user starts service immediately. A buffer is only needed to store the job while it is being processed. Processor sharing arises naturally in the case of competing TCP flows with similar round trip times or in round-robin scheduling of jobs on a fast server [Schassberger 1984]. Processor sharing also has a nice interpretation as being the bandwidth allocation strategy that is socially optimal for $\log$-utility maximization in a single server system (see [Kelly 1997; Yaiche et al. 2000]). Thus processor sharing is a Nash Bargaining solution that is Pareto efficient.

Processor sharing also has important ramifications from the probabilistic viewpoint. The resulting stationary system (when it exists) has a stationary distribution that is *insensitive* to the service time distribution. In other words the stationary distribution is completely characterized by the (Poisson) arrival rates and the mean of the service times. Another nice feature is that the stationary distribution has a product form that

can readily be used to calculate various performance quantities of interest that allows us to measure the revenue effects and user costs using techniques from Palm calculus.

The analysis of the pricing model brings out the implicit connection between the expected user payment and the different moments of the total number of users in the system. The results on a post-payment implementation of the pricing model, where a user pays the charge accrued during its service duration, motivate introducing a quality of service (QoS) constraint in the system where a user only pays for the sojourn duration when a minimum service rate is offered. The QoS constraint discourages the operator from profiting by causing congestion artificially by installing small network capacity.

Another payment mechanism, which we call the pre-payment mechanism, is proposed where the operator charges a user upfront on arrival based on the state of the system. This mechanism however requires the operator to have second moment information of job sizes. The analysis also carries over to systems with admission control based on the global server occupancy. Admission control basically means that any admitted user is guaranteed a minimum service rate by blocking new users under congestion.

Pricing of resources has been studied in many contexts. For example, pricing in communication networks has been extensively studied, see [Songhurst 1999; DaSilva 2000] where the goal was to combine QoS while accounting for burstiness in traffic. In [Kelly et al. 1998] they showed how the Lagrange multipliers as prices for logarithmic utilities could be obtained through primal-dual algorithms in what is often referred to as a Network Utility Maximization (NUM) framework. The role of VCG pricing has been used in many contexts such as spectrum auctions and the idea of implementing VCG prices through NUM techniques was first discussed in [Yang and Hajek 2007]. In the above approaches the results have been derived assuming a fixed population size.

In this paper we study the pricing issue in the context of a processor sharing server with Poisson arrivals of jobs. It is assumed that the prices are instantaneous prices that would be charged according to the appropriate pricing model for a given state of the system. Of course whether these correspond to the optimal prices under the various schemes in some ergodic or time-averaged sense is not clear. Moreover even the structure of the prices over an interval are not known because of the infinite dimensional variational problem that needs to be solved. Furthermore, they would be very difficult to implement without more knowledge of how jobs arrive and their sizes due to the variation of resources offered to a given job over its random sojourn.

We analyze pricing of server bandwidth allocation based on fixed, VCG, and Lagrange pricing frameworks. We derive explicit formulae for the mean payment and operator revenue. We also characterize the second moments that are used to study the volatility of the implementation of the pricing schemes as pre- or post-payment mechanisms. Recent work [Guillemin and Mazumdar 2013] studies the volatility of pre- and post-payment pricing and their tail distributions, and confirms the preference of pre-payment over post-payment due to a smaller coefficient of variation (as was observed in [Birmiwal et al. 2012b] for fixed rate pricing and VCG based pricing in processor sharing systems). The second moment of payment under congestion based pricing is also shown to be smaller for pre-payment than post-payment for small and large values of server capacity, as illustrated in our simulation results.

The organization of this paper is as follows. We present the model in the next section. Useful functionals of the number of users in the system are derived in Section 3. Section 4 presents details on the pricing models, and post-payment and pre-payment mechanisms. Section 5 shows how the framework extends to the situation with admission control. Simulation results are provided in Section 6. All detailed proofs are relegated to an appendix at the end.

2. SYSTEM MODEL

The system is modeled as a single server of capacity C with M/G inputs using the processor sharing discipline. Each arriving user represents a flow, and belongs to one of K classes indexed by the set $\{1, 2, \ldots, K\}$. A class is characterized by its arrival and the service requirement statistics. Class k arrivals are modeled as a Poisson process with rate $\lambda_k \in \mathbb{R}_+$. Each class k user brings a random amount of work, independent and identically distributed, with a common general distribution with mean $\nu_k \in \mathbb{R}_+$. At an instant t, let $\vec{x}(t) = (x_1(t), \ldots, x_K(t))$ denote the system state where $x_k(t) \in \mathbb{Z}_+$ denotes the number of class k users present. Let $|\vec{x}(t)| = \sum_{k=1}^{K} x_k(t)$ denote the total number of users in the system at time t; we will drop the dependence on time from the notation when the meaning is clear. The allocation to a user under processor sharing is $C/|\vec{x}|$ and the collective allocation to class k is $x_k C/|\vec{x}|$. Processor sharing results from maximizing social welfare when each user's utility is logarithmic in its assigned share of the resource. The $\log$ utility function is also of interest since its solution coincides with the Nash bargaining solution and since it is the only scale invariant utility function. We will assume this utility function when we derive the payments under VCG based pricing and congestion-based pricing.

Define $\alpha_k = \lambda_k \nu_k$ and $\vec{\alpha} = (\alpha_1, \ldots, \alpha_K)$. The traffic intensity of class k is denoted by $\rho_k = \alpha_k / C$. Let the total traffic intensity, ρ, be given by $\rho = \sum_{k=1}^{K} \rho_k$. We use the notation $\vec{\alpha}^{\vec{x}} = \prod_{k=1}^{K} \alpha_k^{x_k}$ for convenience.

Let π be the stationary distribution of the underlying Markov process, $\mathbb{P}_N$ be the Palm probability associated with any stationary point process N, and $\mathbb{E}_N$ be the expectation with respect to the Palm probability, which in this case, is the same as the stationary measure due to the Poisson Arrivals See Time Averages property (PASTA). With a slight abuse of notation, let $\mathbb{E}_{\vec{x}}$ be the expectation conditioned on the arrival state $\vec{x}$. Define $\chi(\vec{x}) = \Phi(\vec{x})\vec{\alpha}^{\vec{x}}$ to be an invariant distribution under π where $\Phi(\vec{x})$ is the balance function satisfying

$$\Phi(\vec{x}) = \frac{1}{C} \sum_{m=1}^{K} \Phi(\vec{x} - \vec{e}_m),$$

and $\vec{e}_m$ is a K-dimensional unit vector with a 1 at the mth element (see [Bonald and Proutiere 2002; Birmiwal et al. 2012a; 2012b] for a discussion). The stationary distribution $\pi(\vec{x})$ is then given by

$$\pi(\vec{x}) = \frac{\chi(\vec{x})}{\sum_{\vec{y}} \chi(\vec{y})}. \tag{1}$$

3. PERFORMANCE METRICS

This section will show that the mean user payments and the operator's revenue are characterized by the following statistics of the number of users present in the system.

$$t(n) = \sum_{\vec{x}:|\vec{x}|=n} \chi(\vec{x}) \tag{2}$$

$$s_k(n) = \sum_{\vec{x}:|\vec{x}|=n} x_k \chi(\vec{x}) \tag{3}$$

$$\bar{s}_k(n) = \sum_{m>n} s_k(m) \tag{4}$$

$$s_{i,j}(n) = \sum_{\vec{x}:|\vec{x}|=n} x_i x_j \chi(\vec{x}). \tag{5}$$

The above terms are useful in evaluating different moments of the number of users in the system. For example, the mean number of class k users is obtained as

$$\mathbb{E}[x_k] = \sum_{\vec{x}} x_k \pi(\vec{x}) = \frac{\sum_{\vec{x}} x_k \chi(\vec{x})}{\sum_{\vec{y}} \chi(\vec{y})} = \frac{\sum_{n=0}^{\infty} s_k(n)}{\sum_{n=0}^{\infty} t(n)}.$$

The following lemmas evaluate $t(n), s_k(n), \bar{s}_k(n)$, and $s_{i,j}(n)$. The proofs of these and subsequent results are provided in the Appendix.

LEMMA 3.1. *Let $t(n)$ be defined as in* (2)*. Then, $t(n) = \Phi(\vec{0})\rho^n$ and $\sum_{n=0}^{\infty} t(n) = \sum_{\vec{y}} \chi(\vec{y}) = \frac{\Phi(\vec{0})}{1-\rho}$.*

LEMMA 3.2. *Let $s_k(n)$ and $\bar{s}_k(n)$ be defined as in* (3) *and* (4)*. Then*

$$s_k(n) = n\rho^{n-1}\rho_k \Phi(\vec{0}) \quad \textit{and} \quad \bar{s}_k(n) = \frac{\Phi(\vec{0})}{1-\rho}\rho^n \rho_k \left(n + \frac{1}{1-\rho}\right).$$

LEMMA 3.3. *Let $s_{i,j}(n)$ be defined as in* (5)*. Then,*

$$s_{i,j}(n) = \begin{cases} n(n-1)\rho_i\rho_j\rho^{n-2}\Phi(\vec{0}) & \textit{if } i \neq j \\ n\left((n-1)\rho_i^2 + \rho_i\rho\right)\rho^{n-2}\Phi(\vec{0}) & \textit{if } i = j. \end{cases} \tag{6}$$

The above three lemmas compute the zeroth, the first, and the second moment of the number of users in the system under the invariant distribution $\chi(\vec{x})$. Note that Lemma 3.1 gives the expression for the normalizing term in (1) for obtaining the stationary distribution from the invariant distribution. We also define the terms

$$u(n) := \sum_{\vec{x}:|\vec{x}|>n} (|\vec{x}| - 1/2)\,\pi(\vec{x}) \tag{7}$$

$$v(n) := \sum_{\vec{x}:|\vec{x}|=n} |\vec{x}|^2 \chi(\vec{x}) = n^2 t(n) \tag{8}$$

$$g_k(n) := \sum_{\vec{x}:|\vec{x}|=n} \frac{x_k}{|\vec{x}|}\chi(\vec{x}) = s_k(n)/n, \tag{9}$$

which are needed for evaluating prices.

PROPOSITION 3.4. *Let $u(n)$ be defined as in* (7)*. Then,*

$$u(n) = \rho^{n+1}\left(n + \frac{1}{1-\rho} - \frac{1}{2}\right).$$

The proposition is useful for evaluating revenue under VCG pricing.

4. PRICING MODELS

We now apply the expressions derived in the previous section for the moments of the number of users in a processor sharing system to analyze three pricing models, viz, fixed rate pricing, VCG-based pricing, and congestion based pricing. We start by introducing some notation and preliminary details of the pricing models. Let $R_F(\vec{x}), R_V(\vec{x})$, and $R_L(\vec{x})$ denote the revenue per unit-time collected by the operator under fixed rate pricing, under VCG pricing, and under congestion-based pricing, respectively, when

the number of users is given by $\vec{x}$. Similarly, let $c_k^F(\vec{x}), c_k^V(\vec{x})$, and $c_k^L(\vec{x})$ denote the payment per unit-time by each class k user under the three pricing models in state $\vec{x}$.

The first pricing model is *fixed rate pricing* where the user pays a fixed price of β per unit-time per unit-resource, i.e., if a user is allocated Λ resource for time T, the user pays $\Lambda\beta T$. Thus,

$$c_k^F(\vec{x}) = \frac{\beta C}{|\vec{x}|}, \quad \text{if } |\vec{x}| \geq 1. \tag{10}$$

The operator's revenue is the aggregate of user payments, i.e.,

$$R_F(\vec{x}) = \beta C, \quad \text{if } |\vec{x}| \geq 1.$$

The $\log$ utility assumption is important for the next two pricing models. Under *VCG-based pricing*, a user pays the decrease in maximum social welfare caused by it entering the system (see [Vickrey 1961; Yang and Hajek 2007]). Let s index over the set of users. Let Λ_s denote the allocation to user s. If $|\vec{x}| \geq 2$, the per unit-time price paid by user r of class k is

$$\begin{aligned} c_k^V(\vec{x}) &= \max_{\vec{\Lambda}} \sum_{s \neq r | \Lambda_r = 0} \log(\Lambda_s) - \sum_{s \neq r} \log(\Lambda_s^{PS}) \\ &= (|\vec{x}| - 1) \log \frac{C}{|\vec{x}| - 1} - (|\vec{x}| - 1) \log \frac{C}{|\vec{x}|} \\ &= (|\vec{x}| - 1) \log \frac{|\vec{x}|}{|\vec{x}| - 1}, \end{aligned}$$

where $\Lambda_s^{PS} = C/|\vec{x}|$ is the allocation to user s under processor sharing. If $|\vec{x}| < 2$, then $c_k^V(\vec{x}) = 0$. The aggregate per unit-time revenue collected by the operator is given by

$$R_V(\vec{x}) = \begin{cases} |\vec{x}|(|\vec{x}| - 1) \log \frac{|\vec{x}|}{|\vec{x}|-1} & \text{if } |\vec{x}| \geq 2 \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

To gain further insights in the revenue and payment problem, the following approximation is shown to hold.

PROPOSITION 4.1. *$R_V(\vec{x}) \approx |\vec{x}| - \frac{1}{2}$ and the approximation error is $O(1/|\vec{x}|)$. Furthermore, $|\vec{x}| - \frac{1}{2}$ is an upper bound on $R_V(\vec{x})$ for $|\vec{x}| > 0$.*

The above approximation for VCG revenue is used throughout. Thus, the price paid by the user is approximated as

$$c_k^V(\vec{x}) = \begin{cases} \left(1 - \frac{1}{2|\vec{x}|}\right) & \text{if } |\vec{x}| \geq 2 \\ 0 & \text{otherwise,} \end{cases} \tag{12}$$

and the revenue earned by the operator is given by

$$R_V(\vec{x}) = \begin{cases} |\vec{x}| - 1/2 & \text{if } |\vec{x}| \geq 2 \\ 0 & \text{otherwise.} \end{cases}$$

In *congestion-based pricing*, the Lagrange dual variable (shadow price) of the social welfare maximization problem in (13) is charged to the users, e.g., see [Kelly 1997].

$$\begin{aligned} &\max \sum_{\text{users } r} \log(\Lambda_r) \\ &\text{subject to} \sum_{\text{users r}} \Lambda_r = C, \text{ and } \Lambda_r \geq 0. \end{aligned} \tag{13}$$

This shadow price has the advantage of leading the system to social welfare in a distributed implementation. The shadow price under processor sharing (given by the above constraints) is $|\vec{x}|/C$. Thus, the payment per unit-time made by a class k user is given by

$$c_k^L(\vec{x}) = \begin{cases} \frac{|\vec{x}|}{C} & \text{if } |\vec{x}| \geq 1 \\ 0 & \text{otherwise,} \end{cases} \tag{14}$$

and the aggregate per unit-time revenue collected by the operator is

$$R_L(\vec{x}) = \begin{cases} \frac{|\vec{x}|^2}{C} & \text{if } |\vec{x}| \geq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Fundamentally, the price per unit-time charged to users under fixed rate pricing, VCG-based pricing, and congestion-based pricing is proportional to $1/|\vec{x}|$, to ≈ 1, and to $|\vec{x}|$. The consequence is that a user is charged less (offered a discounted price) at high loads under fixed rate pricing, offered an approximately constant price under VCG pricing, and is charged more (penalized for contributing to congestion) under congestion-based pricing.

4.1. QoS Constrained Pricing

It will be shown that under VCG pricing and congestion-based pricing, an operator can collect arbitrarily large revenue by installing small capacity, leading to longer sojourn times and greater accrued payments. To overcome this, the following QoS requirement is imposed. A class k user pays the operator only if the rate allocated at time t, $C/|\vec{x}(t)|$, is equal to or greater than r_k. For ease of exposition, all r_k are assumed to be identical, i.e., $r_{\min} = r_k$. The $C/|\vec{x}| \geq r_{\min}$ condition is equivalent to a $|\vec{x}| \leq n^*$ condition where $n^* = \lfloor C/r_{\min} \rfloor$. This motivates the operator to install sufficient bandwidth for the users.

4.2. Post-payments vs. Pre-payments

The three pricing models charge a user based on the number of users in the system. A change in the number of users is reflected in the instantaneous per unit-time price. For a given user, the exact charge accrued is evaluated by tracking arrivals and departures during the user's sojourn. The mean of this exact payment incurred by a user is provided using sample path arguments from Palm probability. The mean of the operator's revenue is independently derived. Since the total charge accrued is only known at the end of sojourn, we refer to such an implementation as the *post-payment mechanism* or the *post-payment scheme*.

After deriving the mean revenue and post-payment expressions, a *pre-payment mechanism* (or scheme) is investigated where a user is charged an up-front fee based on the system load and the expected sojourn time observed on arrival. Prices are adjusted to ensure the same mean payment by each class as in the post-payment scheme.

A pre-payment scheme has several benefits. The user is aware of the payment up-front unlike the post-payment scheme where a user may be billed a large fee caused by sudden high loads during its sojourn. Based on the second moment of payments, we also see that user payments have smaller variability under certain conditions and thus the operator's revenue is predictable with greater confidence. Computing pre-payment price however requires greater information about the traffic statistics of each class, which is not required for the post-payment mechanism.

4.3. Mean Operator Revenue

The mean operator revenue is derived under the three pricing models. The expressions hold under both pre-payment and post-payment schemes since the mean payment under both schemes is the same (by design) from each class.

PROPOSITION 4.2. *The operator's mean revenue per unit-time under the three pricing models when the QoS constraint is imposed is given by*

$$\bar{R}_F = \beta C \rho (1 - \rho^{n^*}) \tag{15}$$

$$\bar{R}_V = \frac{\rho^2}{2}\left(1 + \frac{2}{1-\rho}\right) - \rho^{n^*+1}\left(n^* - \frac{1}{2} + \frac{1}{1-\rho}\right) \tag{16}$$

$$\bar{R}_L = \frac{1-\rho}{C}\sum_{n=1}^{n^*} n^2 \rho^n. \tag{17}$$

COROLLARY 4.3. *The operator's mean revenue per unit-time under the three pricing models in the absence of the QoS constraint is given by*

$$\tilde{R}_F = \beta C \rho, \;\; \tilde{R}_V = \frac{\rho^2}{2}\left(\frac{3-\rho}{1-\rho}\right), \; \textit{and} \;\; \tilde{R}_L = \frac{\rho(1+\rho)}{C(1-\rho)^2}.$$

The proof is provided in the Appendix. The proof highlights that the mean revenue earned by the operator for fixed rate pricing, VCG pricing, and congestion-based pricing is respectively related to the zeroth, first, and the second moment of the total number of users in the system, i.e.,

$$\begin{aligned}
\bar{R}_F &\propto \mathbb{E}[|\vec{x}|^0 \mathbf{1}_{(1 \le |\vec{x}| \le n^*)}] \\
\bar{R}_V &\propto \mathbb{E}[(|\vec{x}| - 1/2)\mathbf{1}_{(2 \le |\vec{x}| \le n^*)}] \\
\bar{R}_L &\propto \mathbb{E}[|\vec{x}|^2 \mathbf{1}_{(1 \le |\vec{x}| \le n^*)}].
\end{aligned}$$

This key insight is attributed to the inherent structure of the pricing models identified in Section 4. The corollary shows that the mean operator's revenue grows indefinitely as $\rho \uparrow 1$ under VCG based pricing and congestion-based pricing.

4.4. Post-payments: Exact charge accrued by users

PROPOSITION 4.4. *The mean total payment by a class k user under the three pricing models with the QoS constraint imposed is given by*

$$\begin{aligned}
\bar{c}_k^F &= \nu_k \beta (1 - \rho^{n^*}) \\
\bar{c}_k^V &= \frac{\nu_k}{C}\left(\rho\left(\frac{1-\rho^{n^*-1}}{1-\rho}\right) + \frac{\rho}{2} - \left(n^* - \frac{1}{2}\right)\rho^{n^*}\right) \\
\bar{c}_k^L &= \frac{\nu_k(1-\rho)}{C^2}\sum_{n=1}^{n^*} n^2 \rho^{n-1}.
\end{aligned}$$

While $c_k^{(\cdot)}(\vec{x})$ denotes the exact payment per unit time by a class k user given the system state $\vec{x}$, note that $\bar{c}_k^{(\cdot)}$ is the expected total payment by a user of class k (not per unit time).

COROLLARY 4.5. *The mean total payment by a class k user under the three pricing models in the absence of the QoS constraint is given by*

$$\tilde{c}_k^F = \nu_k \beta, \;\; \tilde{c}_k^V = \frac{\nu_k \rho}{2C}\left(\frac{3-\rho}{1-\rho}\right), \; \textit{and} \;\; \tilde{c}_k^L = \frac{\nu_k(1+\rho)}{C^2(1-\rho)^2}.$$

Note that the metering required by the operator is at the time-scale at which users enter and leave the system. The next proposition characterizes the second moment of user payments for the special case of M/M inputs. Define

$$\mu(\vec{x}) = \sum_{m=1}^{K} \frac{x_m C}{|\vec{x}|\nu_m}, \tag{18}$$

the rate at which users depart from the system. Also define $\omega(\vec{x}) = |\vec{\lambda}| + \mu(\vec{x})$, the rate at which the underlying Markov process leaves the state $\vec{x}$.

PROPOSITION 4.6. *Let $\vec{\eta}_k^X(\vec{x})$ and $\xi_k^X(\vec{x})$ respectively denote the first and second moment of the total payment made by class k users when the state immediately after arrival is $\vec{x}$, under the pricing model X. Then, the second moment of payments by class k users, ξ_k^X, is given by*

$$\xi_k^X = \sum_{\vec{x}:x_k \geq 1} \xi_k^X(\vec{x}) \frac{\pi(\vec{x} - \vec{e}_k)}{\sum_{\vec{y}:y_k \geq 1} \pi(\vec{y} - \vec{e}_k)}, \tag{19}$$

where $\xi_k^X(\vec{x})$ is obtained by solving

$$\omega(\vec{x})\eta_k^X(\vec{x}) = \left(\sum_{m=1}^{K} \lambda_m \eta_k^X(\vec{x} + \vec{e}_m) \right) + \sum_{m=1}^{K} \frac{(\vec{x} - \vec{e}_k)_m C}{|\vec{x}|\nu_m} \eta_k^X(\vec{x} - \vec{e}_m) + c_k^X(\vec{x}) \cdot \mathbf{1}_{(1 \leq |\vec{x}| \leq n^*)}$$

and

$$\begin{aligned} \omega(\vec{x})\xi_k^X(\vec{x}) = &\left(\sum_{m=1}^{K} \lambda_m \xi_k^X(\vec{x} + \vec{e}_m) \right) + \sum_{m=1}^{K} \frac{(\vec{x} - \vec{e}_k)_m C}{|\vec{x}|\nu_m} \xi_k^X(\vec{x} - \vec{e}_m) \\ &+ 2c_k^X(\vec{x})\eta_k^X(\vec{x}) \cdot \mathbf{1}_{(1 \leq |\vec{x}| \leq n^*)}. \end{aligned} \tag{20}$$

The proof of this proposition is omitted as it is a simple variation of the proof of Proposition 5.3 without state space contraction and arrival rate constraints.

4.5. Pre-payments: Freezing Prices on Arrival

A pre-payment scheme is devised next where the user is charged a price up-front on arrival. For each of the three pricing models (fixed rate, VCG, congestion based), the price charged to a given user now depends only on the number of users in the system when it arrived and is based on the expected sojourn time at arrival. The payment is adjusted so that the mean payments by class k users remain the same as in Section 4.4.

Let W_k be the random variable denoting the sojourn time of the class k arrival. Under the pricing model X, where X is a placeholder for F, V, or L, let $\gamma_k^X(\vec{x})$ be the per unit-time price fixed on class k user's arrival when the arrival observes the system state as $\vec{x}$. Under the pre-payment scheme, the price charged to any class k user is given by

$$p_k^X(\vec{x}) = \gamma_k^X(\vec{x} + \vec{e}_k)\mathbb{E}_{\vec{x}}[W_k]. \tag{21}$$

Based on the structure of fixed rate pricing, VCG pricing, and congestion-based pricing, we define $\gamma_k^X(\vec{x})$ as

$$\gamma_k^F(\vec{x}) = \sigma_k^F |\vec{x}|^{-1}, \ \gamma_k^V(\vec{x}) = \sigma_k^V, \text{ and } \gamma_k^L(\vec{x}) = \sigma_k^L |\vec{x}|.$$

Note that $\gamma_k^X(\vec{x})$ is not zero for $|\vec{x}| > n^*$. The constants σ_k^F, σ_k^V, and σ_k^L are determined by equating the mean payments by class k users to the payments in Section 4.4, i.e.,

$$\mathbb{E}[p_k^X(\vec{x})] = \bar{c}_k^X.$$

PROPOSITION 4.7. *For a processor sharing server with multiclass M/G inputs,*

$$\mathbb{E}_{\vec{x}}[W_k] = A_{k,0} + \sum_{m=1}^{K} A_{k,m} x_m,$$

for some positive coefficients $A_{k,i}$, $0 \le i \le K$.

The proposition is an immediate consequence of [Rege and Sengupta 1994, Theorem 6] which shows that the sojourn time of a new arrival can be decomposed into the sum of random variables for each pre-existing user and the new arrival. The proposition follows by taking the expectation of this sum.

PROPOSITION 4.8. *The constants σ_k^F, σ_k^V, and σ_k^L are*

$$\sigma_k^F = \frac{\nu_k \beta (1-\rho^{n^*})}{(1-\rho)} \left(\frac{A_{k,0}}{\rho} \log \frac{1}{1-\rho} + \frac{1}{\rho^2} \left(\frac{\rho}{1-\rho} - \log \frac{1}{1-\rho} \right) \sum_{m=1}^{K} A_{k,m} \rho_m \right)^{-1}$$

$$\sigma_k^V = \rho(1-\rho^{n^*-1}) + \frac{\rho(1-\rho)}{2} - (1-\rho)\left(n^* - \frac{1}{2}\right)\rho^{n^*}$$

$$\sigma_k^L = \frac{\nu_k (1-\rho)^2}{C^2} \frac{\sum_{n=1}^{n^*} n^2 \rho^{n-1}}{A_{k,0} + \frac{2}{1-\rho} \sum_{m=1}^{K} A_{k,m} \rho_m}.$$

The outline of the proof is provided in the Appendix.

Since the mean payments by class k users remain the same as under the post-payment scheme, the mean revenue collected by the operator also remains the same as in Section 4.3. The evaluation of the performance metrics in Section 3 allows the explicit characterization of the second moment (and hence the standard deviation) of class k payments. Define $\mathrm{Li}_2(\rho) = \sum_{n=1}^{\infty} \frac{\rho^n}{n^2}$.

PROPOSITION 4.9. *The second moment of pre-payments by class k users is given by*

$$\begin{aligned}
\mathbb{E}[(p_k^F(\vec{x}))^2] = {} & (\sigma_k^F)^2 (1-\rho) A_{k,0}^2 \frac{\mathrm{Li}_2(\rho)}{\rho} + (\sigma_k^F)^2 (1-\rho) \left(\frac{\rho + 3\log(1-\rho) - 3\rho\log(1-\rho)}{\rho^3(1-\rho)} + \right. \\
& \left. + \frac{2\,\mathrm{Li}_2(\rho)}{\rho^3} \right) \sum_{m=1}^{K} A_{k,m}^2 \rho_m^2 + (\sigma_k^F)^2 (1-\rho)(-\mathrm{Li}_2(\rho) - \log(1-\rho)) \sum_{m=1}^{K} \frac{A_{k,m}^2 \rho_m}{\rho^2} \\
& + \frac{2A_{k,0}(\sigma_k^F)^2(1-\rho)}{\rho^2} (-\mathrm{Li}_2(\rho) - \log(1-\rho)) \sum_{m=1}^{K} A_{k,m} \rho_m \\
& + \frac{(\sigma_k^F)^2}{\rho^3} [\rho + 3(1-\rho)\log(1-\rho) + 2(1-\rho)\,\mathrm{Li}_2(\rho)] \sum_{i=1}^{K} \sum_{j=1:j\neq i}^{K} A_{k,i} A_{k,j} \rho_i \rho_j
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[(p_k^V(\vec{x}))^2] = {} & (\sigma_k^V)^2 A_{k,0}^2 + \frac{2(\sigma_k^V)^2}{(1-\rho)^2} \sum_{m=1}^{K} A_{k,m}^2 \rho_m^2 + \frac{(\sigma_k^V)^2}{(1-\rho)} \sum_{m=1}^{K} A_{k,m}^2 \rho_m \\
& + \frac{2(\sigma_k^V)^2}{(1-\rho)} A_{k,0} \sum_{m=1}^{K} A_{k,m} \rho_m + \frac{2(\sigma_k^V)^2}{(1-\rho)^2} \sum_{i=1}^{K} \sum_{j=1:j\neq i}^{K} A_{k,i} A_{k,j} \rho_i \rho_j
\end{aligned}$$

$$\mathbb{E}[(p_k^L(\vec{x}))^2] = \frac{(\sigma_k^L)^2 A_{k,0}^2(1+\rho)}{(1-\rho)^2} + (\sigma_k^L)^2 \sum_{m=1}^{K} A_{k,m}^2 \frac{2\rho_m(2+9\rho_m+3\rho_m\rho-\rho-\rho^2)}{(1-\rho)^4}$$
$$+ 2A_{k,0}(\sigma_k^L)^2 \frac{2\rho+4}{(1-\rho)^3} \sum_{m=1}^{K} A_{k,m}\rho_m + \frac{(\sigma_k^L)^2 6(3+\rho)}{(1-\rho)^4} \sum_{i=1}^{K} \sum_{j=1:j\neq i}^{K} A_{k,i}A_{k,j}\rho_i\rho_j.$$

The outline of the proof is provided in the Appendix. We note that the proofs rely on further metrics such as $\mathbb{E}[x_i|\vec{x}|]$, $\mathbb{E}[x_i x_k|\vec{x}|^2]$ (higher moments), and $\mathbb{E}[x_i x_j/|\vec{x}|^2]$.

When employing the post-payment mechanism, the operator only requires the current number of users in the system to evaluate user prices per unit-time. On the contrary, prices under the pre-payment scheme depend on the expected sojourn time and thus require traffic statistics for each class, including the coefficients $A_{k,m}$ a priori.

5. ADMISSION CONTROL: A SPECIAL CASE OF QOS CONSTRAINT

So far we analyzed a situation where all users are allowed to enter but only pay for the periods when they receive their minimum bandwidth or rate requirements. This prevents the operator from installing small capacity because users would encounter congestion more often and have long processing times. This may not be very satisfactory to an operator or users.

We now show that it is easy to use the previous analysis to consider a slightly different model where QoS is enforced by admission control. Here users do not enter the system if the rate available on arrival is below $r_{\min}$. This is a simple example of state dependent arrivals, i.e., arrivals stop if $|\vec{x}| = n^*$. This mechanism is effectively one of admission control where instead of the server blocking new arrivals, the new arrival itself refuses to enter the system. Under congestion based pricing, this may also be considered as a reaction to high instantaneous price.

This mechanism has a structure similar to that of post-payments with QoS constraints described in Section 4.1. The following propositions provide the operator's mean revenue per unit-time and mean total payments by a class k user. The proofs for results in this section are provided in the Appendix.

PROPOSITION 5.1. *The operator's mean revenue per unit-time with admission control under the three pricing models is given by*

$$\hat{R}_F = \frac{\bar{R}_F}{1-\rho^{n^*+1}},\ \hat{R}_V = \frac{\bar{R}_V}{1-\rho^{n^*+1}},\ \textit{and}\ \hat{R}_L = \frac{\bar{R}_L}{1-\rho^{n^*+1}}.$$

PROPOSITION 5.2. *The mean total payment made by a class k user with admission control under the three pricing models is given by*

$$\hat{c}_k^F = \frac{\bar{c}_k^F}{1-\rho^{n^*}},\ \hat{c}_k^V = \frac{\bar{c}_k^V}{1-\rho^{n^*}},\ \textit{and}\ \hat{c}_k^L = \frac{\bar{c}_k^L}{1-\rho^{n^*}}.$$

Even when $\rho \uparrow 1$, the mean payments and revenues do not become arbitrarily large. This is because for ρ close to unity, the system will have at most $C/r_{\min}$ number of users (and thus the per unit-time price is finite) and the users will exit in finite time since a minimum service rate of $r_{\min}$ is guaranteed.

The second moment of payments is evaluated next under admission control for the special case of M/M inputs. Define $\hat{\omega}(\vec{x}) = |\vec{\lambda}| \cdot \mathbf{1}_{(|\vec{x}|<n^*)} + \mu(\vec{x})$ for convenience, where $\mu(\vec{x})$ is defined in (18).

PROPOSITION 5.3. *Consider an admission controlled system with M/M input. Suppose $n^* > 1$. Let $\eta_k^X(\vec{x})$ and $\xi_k^X(\vec{x})$ respectively be the first and second moment of payment*

by a class k user under the pricing model X when the system state immediately after the user's arrival is $\vec{x}$. Define $\mathsf{X}^k = \{\vec{x} \in \mathbb{Z}_+^K : |\vec{x}| \leq n^* \text{ and } x_k \geq 1\}$. Then, the second moment of payments by class k users, $\hat{\xi}_k^X$, is given by

$$\hat{\xi}_k^X = \sum_{\vec{x} \in \mathsf{X}^k} \xi_k^X(\vec{x}) \frac{\hat{\pi}(\vec{x} - \vec{e}_k)}{\sum_{\vec{y} \in \mathsf{X}^k} \hat{\pi}(\vec{x} - \vec{e}_k)}, \tag{22}$$

where $\xi_k^X(\vec{x})$ is the solution of

$$\hat{\omega}(\vec{x})\eta_k^X(\vec{x}) = \left(\sum_{m=1}^K \lambda_m \eta_k^X(\vec{x} + \vec{e}_m)\right) \cdot \mathbf{1}_{(|\vec{x}|<n^*)} + \sum_{m=1}^K \frac{(\vec{x} - \vec{e}_k)_m C}{|\vec{x}|\nu_m} \eta_k^X(\vec{x} - \vec{e}_m) + c_k^X(\vec{x}), \quad \textit{for } \vec{x} \in \mathsf{X}^k, \tag{23}$$

and

$$\hat{\omega}(\vec{x})\xi_k^X(\vec{x}) = \left(\sum_{m=1}^K \lambda_m \xi_k^X(\vec{x} + \vec{e}_m)\right) \cdot \mathbf{1}_{(|\vec{x}|<n^*)} + \sum_{m=1}^K \frac{(\vec{x} - \vec{e}_k)_m C}{|\vec{x}|\nu_m} \xi_k^X(\vec{x} - \vec{e}_m) + 2c_k^X(\vec{x})\eta_k^X(\vec{x}), \quad \textit{for } \vec{x} \in \mathsf{X}^k. \tag{24}$$

All $\eta_k^X(\vec{x})$ and $\xi_k^X(\vec{x})$ terms where $\vec{x} \notin \mathsf{X}^k$ are ignored.

Remark 5.4. The system of equations in (23) for obtaining the conditional mean payment $\eta_k^X(\vec{x})$ and in (24) for obtaining the conditional second moment $\xi_k^X(\cdot)$ in Proposition 5.3 have unique solutions. This is easily seen by writing the equations as $\mathbf{A}\vec{x} = \vec{b}$ and identifying the eigenvalues of $\mathbf{A}$ to lie outside the origin by the Gershgorin disc theorem (see [Horn and Johnson 1985]).

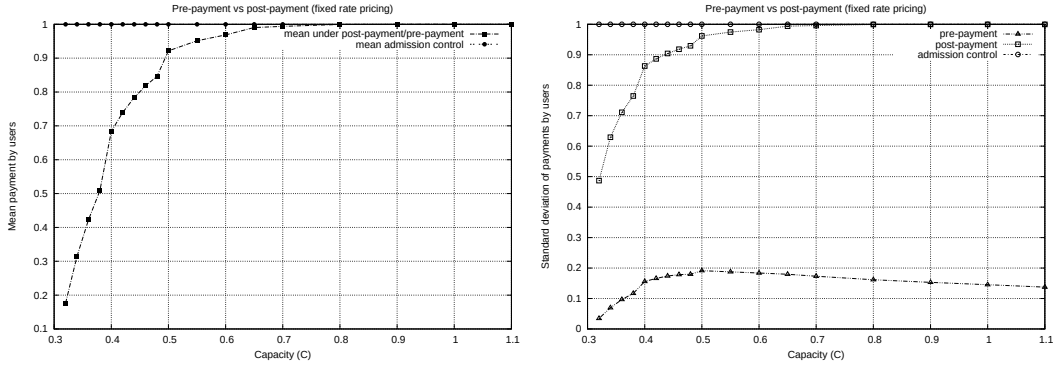
Remark 5.5. Note that $c_k^V(\vec{e}_k) = 0$, i.e., a sole user in the system does not pay the operator under VCG based pricing.

The outline of the proof for the fixed rate pricing model is presented in the Appendix. The steps for the other two pricing models are similar.

6. SIMULATION RESULTS

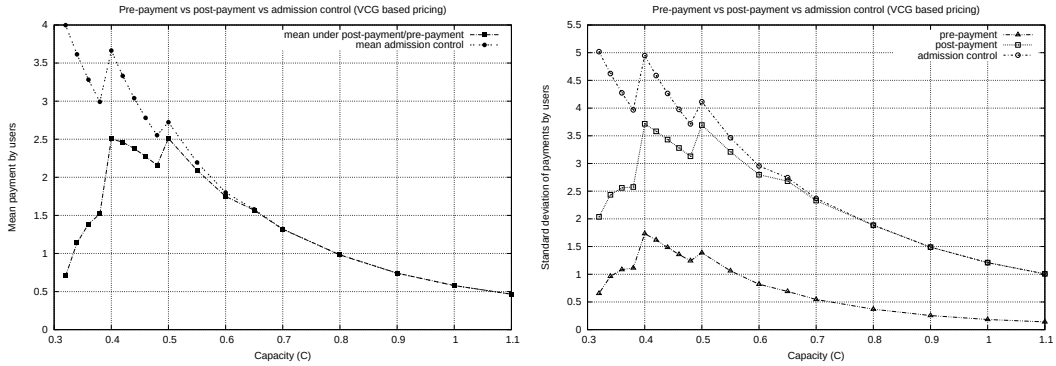
This section presents representative numerical results obtained from simulations. The simulation data is in close agreement with theoretically obtained results. Figure 1, Figure 2, and Figure 3 provide a comparison of mean user payments and standard deviation under the post-payment mechanism, the pre-payment mechanism, and admission control for fixed rate pricing, VCG based pricing, and congestion-based pricing. The results are generated for a system with a single class of traffic with Poisson arrivals and exponentially distributed service requirements ($\lambda_0 = 0.3, \nu_1 = 1.0$). For fixed rate pricing, $\beta = 1$. For QoS constraint and admission control, $r_{\min}$ is 0.1. The $A_{k,m}$ constants are estimated from simulations. Standard deviation of payments under admission control is obtained from Proposition 5.3.

The plots corroborate our analytical findings. Under fixed-rate pricing, mean post payment increases to $\beta\nu_k$ as capacity increases because a greater proportion of the users' sojourn time is spent in the regime where the QoS constraint is satisfied. Under VCG and congestion based pricing however, mean post payment decreases with increasing capacity as the payments are increasing in congestion. Mean pre-payments are designed to be the same as mean post-payments. The mean payments under pre-payment and post-payment mechanisms are over the same probability space $\{\vec{x} \in \mathbb{Z}_+^K\}$. The mean payments under the admission control scheme exists on a different, smaller



(a) Mean payment (b) Standard deviation

Fig. 1: Fixed rate pricing



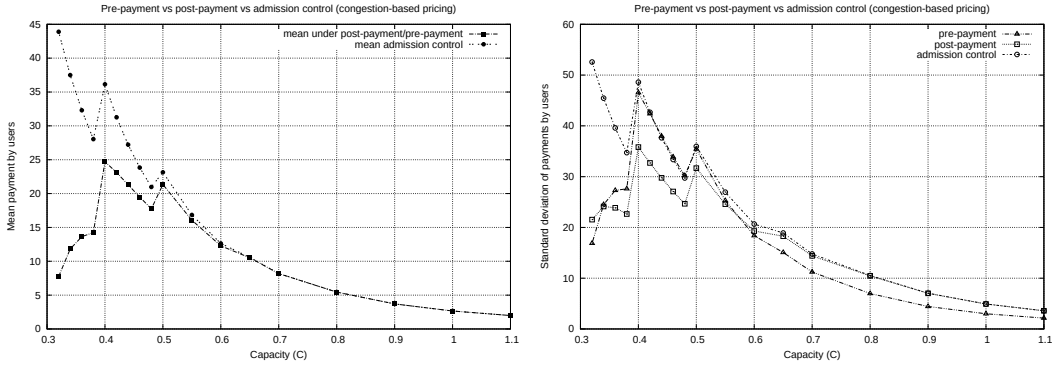
(a) Mean payment (b) Standard deviation

Fig. 2: VCG based pricing

probability space $\{\vec{x} \in \mathbb{Z}_+^K : 0 \leq |\vec{x}| \leq n^*\}$. The mean payment under admission control is the mean of the same expression as under post-payment with QoS but with re-normalized probabilities.

The standard deviations of payments are also provided. Since mean-payments are the same under pre and post pricing mechanisms, the standard deviations indicate that pre-payments have smaller variability as compared to post-payment mechanism under fixed rate, VCG based pricing, and for small and large values of capacity under congestion based pricing. An analytical explanation for this is provided in [Guillemin and Mazumdar 2013]. Standard deviation under admission control is also provided. For low ρ (high capacity rates), the admission control mechanism is effectively the same as the post-payment mechanism and thus payments and second moment results are identical under those two mechanisms.

The discontinuities in the figures are due to the discontinuities in $n^* = \lfloor C/r_{\min} \rfloor$.



(a) Mean payment (b) Standard deviation

Fig. 3: Congestion based pricing

7. CONCLUSION

In this paper we analyzed pricing for a model with a single server that allocates its service rate between jobs of multiple classes via the processor sharing discipline. Processor sharing occurs naturally in several resource allocation problems, e.g., CPU sharing on a server. Bandwidth distribution between competing TCP flows with the same round trip time are also known to share equal bandwidths.

Relevant performance metrics were derived for the processor sharing mechanism and used for characterizing operator revenue and user payments under three pricing models: fixed rate pricing, VCG pricing, and congestion-based pricing. The analysis highlights a link between the pricing schemes and different moments of the number of users in the system. The method developed in the paper can be extended to analyze nonlinear pricing schemes based on other moments of users present in the system.

An interesting observation is that there is a difference in volatility between charging upfront based on the system state and charging at job completion. It is seen that the pre-payment mechanism is preferable in terms of revenue and cost predictability. However, the pre-payment method requires information about the second moments of the jobs while the post-payment method requires us to maintain information not only of sojourn times but also the periods of congestion when the rate contract is not fulfilled. Thus one trades off more information about the nature of jobs vs. bookkeeping at the server.

Future directions will be to see how such pricing schemes can be applied in a network of processor sharing servers.

APPENDIX

PROOF OF LEMMA 3.1. We have that $t(0) = \chi(\vec{0}) = \Phi(\vec{0})$. For $n \geq 1$,

$$t(n) := \sum_{\vec{x}:|\vec{x}|=n} \chi(\vec{x}) = \sum_{\vec{x}:|\vec{x}|=n} \frac{1}{C} \sum_{m=1}^{K} \Phi(\vec{x} - \vec{e}_m)\vec{\alpha}^{\vec{x}}$$

$$= \sum_{m=1}^{K} \rho_m \sum_{\vec{x}:|\vec{x}|=n-1} \Phi(\vec{x})\vec{\alpha}^{\vec{x}} = \rho \cdot t(n-1).$$

Also, $\sum_{\vec{x}} \chi(\vec{x}) = \sum_{n=0}^{\infty} t(n) = \frac{\Phi(\vec{0})}{1-\rho}$. $\square$

PROOF OF LEMMA 3.2. We start with

$$\begin{aligned} s_k(n) &= \sum_{\vec{x}:|\vec{x}|=n} \frac{x_k}{C} \sum_{m=1}^{K} \Phi(\vec{x}-\vec{e}_m)\vec{\alpha}^{\vec{x}} \\ &= \sum_{m=1}^{K} \rho_m \left[\sum_{\vec{x}:|\vec{x}|=n-1} x_k \chi(\vec{x}) + \sum_{\vec{x}:|\vec{x}|=n-1} (\vec{e}_m)_k \chi(\vec{x}) \right] \\ &= \sum_{m=1}^{K} \rho_m s_k(n-1) + \sum_{m=1}^{K} \rho_m \sum_{\vec{x}:|\vec{x}|=n-1} (\vec{e}_m)_k \chi(\vec{x}). \end{aligned}$$

$$s_k(n) = \rho s_k(n-1) + \rho_k t(n-1) = \rho s_k(n-1) + \rho_k \rho^{n-1} \Phi(\vec{0}). \tag{25}$$

It is easily shown that the result is a solution to the recursion in (25), starting with $s_k(0) = 0$. The first part of the result follows. Next,

$$\bar{s}_k(n) = \rho_k \Phi(\vec{0}) \sum_{m>n} m\rho^{m-1} = \rho_k \Phi(\vec{0}) \frac{\rho^n}{1-\rho} \left(n + \frac{1}{1-\rho} \right).$$

□

PROOF OF LEMMA 3.3. The proof relies on establishing a recursive expression for $s_{i,j}(n)$ as in (6).

$$\begin{aligned} s_{i,j}(n) &= \sum_{m=1}^{K} \frac{\alpha_m}{C} \sum_{\vec{x}:|\vec{x}|=n} x_i x_j \Phi(\vec{x}-\vec{e}_m)\vec{\alpha}^{\vec{x}-\vec{e}_m} \\ &= \sum_{m=1}^{K} \rho_m \sum_{\vec{y}:|\vec{y}|=n-1} (\vec{y}+\vec{e}_m)_i (\vec{y}+\vec{e}_m)_j \Phi(\vec{y})\vec{\alpha}^{\vec{y}} \\ &= \rho s_{i,j}(n-1) + \rho_j s_i(n-1) + \rho_i s_j(n-1) + \mathbf{1}_{(i=j)} \rho_i t(n-1). \end{aligned}$$

Note that $s_{i,j}(0) = 0$ for any i, j and that $s_{i,j}(1) = 0$ if $i \neq j$. The expression in (6) is the solution to this recursion. □

PROOF OF PROPOSITION 3.4.

$$\begin{aligned} u(n) &= \sum_{\vec{x}:|\vec{x}|>n} \left(|\vec{x}| - \frac{1}{2} \right) \pi(\vec{x}) = \sum_{\vec{x}:|\vec{x}|>n} |\vec{x}|\pi(\vec{x}) - \frac{1}{2} \sum_{\vec{x}:|\vec{x}|>n} \pi(\vec{x}) \\ &= \sum_{\vec{x}:|\vec{x}|>n} \frac{(x_1 + \cdots + x_K)\chi(\vec{x})}{\sum_{\vec{y}} \chi(\vec{y})} - \frac{1}{2} \sum_{\vec{x}:|\vec{x}|>n} \frac{\chi(\vec{x})}{\sum_{\vec{y}} \chi(\vec{y})}. \end{aligned}$$

Using $\sum_{\vec{x}} \chi(\vec{x}) = \sum_{n\geq 0} t(n)$,

$$u(n) = \frac{\sum_{k=1}^{K} \bar{s}_k(n)}{\sum_{m\geq 0} t(m)} - \frac{\sum_{m>n} t(m)}{2\sum_{m\geq 0} t(m)}.$$

Using Lemma 3.1 and Lemma 3.2, we get

$$u(n) = \rho^{n+1} \left(n + \frac{1}{1-\rho} \right) - \frac{\rho^{n+1}}{2}.$$

□

PROOF OF PROPOSITION 4.1. From (11), we have

$$R_V(\vec{x}) = |\vec{x}|(|\vec{x}|-1)\sum_{n=1}^{\infty}\frac{1}{n|\vec{x}|^n} = (|\vec{x}|-1)\left(1+\frac{1}{2|\vec{x}|}+\frac{1}{3|\vec{x}|^2}+\dots\right)$$
$$= |\vec{x}| - \frac{1}{2} - \sum_{m=2}^{\infty}\frac{1}{m(m+1)|\vec{x}|^{m-1}} < |\vec{x}| - \frac{1}{2},$$

which shows the required result. □

PROOF OF PROPOSITION 4.2. Let the system be in state $\vec{x}$. The mean revenue per unit time under fixed rate pricing is given by

$$\bar{R}_F = \sum_{\vec{x}:1\le|\vec{x}|\le n^*}\beta C\pi(\vec{x}) = \beta C\frac{1}{\sum_{\vec{x}}\chi(\vec{x})}\sum_{\vec{x}:1\le|\vec{x}|\le n^*}\chi(\vec{x}) = \beta C\frac{\sum_{n=1}^{n^*}t(n)}{\sum_{n=0}^{\infty}t(n)}.$$

Using Lemma 3.1,

$$\bar{R}_F = \beta C\frac{1-\rho}{\Phi(\vec{0})}\frac{\Phi(\vec{0})(\rho-\rho^{n^*+1})}{1-\rho} = \beta C\rho(1-\rho^{n^*}).$$

The mean revenue under VCG pricing is

$$\bar{R}_V = \sum_{\vec{x}:2\le|\vec{x}|\le n^*}\left(|\vec{x}|-\frac{1}{2}\right)\pi(\vec{x}) = u(1)-u(n^*),$$

where $u(n)$ is given by Proposition 3.4. The result for $\bar{R}_V$ follows by simplification.

The mean revenue under congestion-based pricing is given by

$$\bar{R}_L = \sum_{\vec{x}:1\le|\vec{x}|\le n^*}\frac{|\vec{x}|^2}{C}\pi(\vec{x}) = \frac{1}{C}\sum_{\vec{x}:1\le|\vec{x}|\le n^*}|\vec{x}|^2\frac{\chi(\vec{x})}{\sum_{\vec{y}}\chi(\vec{y})} = \frac{1}{C}\frac{\sum_{1\le n\le n^*}v(n)}{\sum_{n\ge 0}t(n)}.$$

Using $v(n) = n^2t(n)$ defined in (8) and Lemma 3.1,

$$\bar{R}_L = \frac{1-\rho}{C}\sum_{n=1}^{n^*}n^2\rho^n.$$

□

PROOF OF PROPOSITION 4.4. To evaluate the mean payment by a class k user under fixed rate pricing, consider the following integral where A_k is the arrival process for class k users and W_0^k is the random variable denoting the sojourn time of the class k arrival at time 0:

$$\bar{c}_k^F = \mathbb{E}_{A_k}\left[\int_0^{W_0^k}\frac{\beta C}{|\vec{x}(t)|}\mathbf{1}_{(1\le|\vec{x}(t)|\le n^*)}dt\right].$$

Applying the Swiss Army formula (see [Brémaud 1993]), equation (9), and Lemma 3.2,

$$\bar{c}_k^F = \frac{\beta C}{\lambda_k}\mathbb{E}\left[\frac{x_k}{|\vec{x}|}\mathbf{1}_{(1\le|\vec{x}|\le n^*)}\right] = \frac{\beta C}{\lambda_k}\sum_{n=1}^{n^*}\sum_{\vec{x}:|\vec{x}|=n}\frac{x_k}{n}\pi(\vec{x})$$
$$= \frac{\beta C}{\lambda_k}\frac{1}{\sum_{n=0}^{\infty}t(n)}\sum_{n=1}^{n^*}g_k(n) = \nu_k\beta\rho(1-\rho^{n^*}).$$

Similarly, for VCG pricing, the mean payment for a class k user is given by

$$\bar{c}_k^V = \mathbb{E}_{A_k}\left[\int_0^{W_0^k}\left(1-\frac{1}{2|\vec{x}(t)|}\right)\mathbf{1}_{(2\le|\vec{x}(t)|\le n^*)}dt\right].$$

Again, using the Swiss Army formula,

$$\begin{aligned}\bar{c}_k^V &= \frac{1}{\lambda_k}\mathbb{E}\left[x_k\left(1-\frac{1}{2|\vec{x}|}\right)\mathbf{1}_{(2\le|\vec{x}|\le n^*)}\right]\\ &= \frac{1}{\lambda_k}\mathbb{E}[x_k\mathbf{1}_{(2\le|\vec{x}|\le n^*)}] - \frac{1}{\lambda_k}\mathbb{E}\left[\frac{x_k}{2|\vec{x}|}\mathbf{1}_{(2\le|\vec{x}|\le n^*)}\right]. \qquad (26)\end{aligned}$$

Let J_1 and J_2 be the first and the second term respectively in (26). Then,

$$J_1 = \frac{1}{\lambda_k}\sum_{n=2}^{n^*}\sum_{\vec{x}:|\vec{x}|=n} x_k\pi(\vec{x}) = \frac{1-\rho}{\lambda_k\Phi(\vec{0})}\sum_{n=2}^{n^*}s_k(n) = \frac{\nu_k(1-\rho)}{C}\sum_{n=2}^{n^*}n\rho^{n-1},$$

and,

$$J_2 = \frac{1}{2\lambda_k}\sum_{n=2}^{n^*}\sum_{\vec{x}:|\vec{x}|=n}\frac{x_k}{|\vec{x}|}\pi(\vec{x}) = \frac{1-\rho}{2\Phi(\vec{0})\lambda_k}\sum_{n=2}^{n^*}g_k(n) = \frac{\nu_k\rho}{2C}(1-\rho^{n^*-1}).$$

Using the identity

$$\sum_{n=1}^{m}n\rho^{n-1} = \frac{1-\rho^{m+1}-(m+1)(1-\rho)\rho^m}{(1-\rho)^2},$$

and simplifying provides the required result. For congestion-based pricing, the mean payment by a class k user is given by

$$\bar{c}_k^L = \mathbb{E}_{A_k}\left[\int_0^{W_0^k}\frac{|\vec{x}(t)|}{C}\mathbf{1}_{(1\le|\vec{x}(t)|\le n^*)}dt\right].$$

Applying the Swiss Army formula gives,

$$\begin{aligned}\bar{c}_k^L &= \frac{1}{\lambda_k C}\mathbb{E}[x_k|\vec{x}|\mathbf{1}_{(1\le|\vec{x}|\le n^*)}] = \frac{1}{\lambda_k C}\sum_{n=1}^{n^*}\sum_{\vec{x}:|\vec{x}|=n}x_k n\pi(\vec{x})\\ &= \frac{1-\rho}{\lambda_k C\Phi(\vec{0})}\sum_{n=1}^{n^*}ns_k(n) = \frac{\nu_k(1-\rho)}{C^2}\sum_{n=1}^{n^*}n^2\rho^{n-1},\end{aligned}$$

which shows the required result. □

PROOF OF PROPOSITION 4.8. Suppose a class k arrival sees the system state as $\vec{x}$ on arrival. The fixed rate, pre-payment price charged is

$$p_k^F(\vec{x}) = \frac{\sigma_k^F}{|\vec{x}+\vec{e}_k|}\left(A_{k,0}+\sum_{m=1}^{K}A_{k,m}x_m\right).$$

It is required that the mean payment by a class k user equal $\bar{c}_k^F$, i.e.,

$$\mathbb{E}[p_k^F(\vec{x})] = \bar{c}_k^F. \qquad (27)$$

Starting with the left hand side (LHS) of (27),

$$LHS = \sum_{\vec{x}} \frac{\sigma_k^F}{|\vec{x}+\vec{e}_k|}\pi(\vec{x})\left(A_{k,0} + \sum_{m=1}^{K} A_{k,m}x_m\right)$$
$$= \sigma_k^F(1-\rho)\left[\frac{A_{k,0}}{\rho}\log\frac{1}{1-\rho} + \left(\frac{\frac{\rho}{1-\rho} - \log\frac{1}{1-\rho}}{\rho^2}\right)\sum_{m=1}^{K} A_{k,m}\rho_m\right].$$

Equating this to $\bar{c}_k^F$ gives σ_k^F. Similarly, under VCG pricing,

$$p_k^V(\vec{x}) = \sigma_k^V\left(A_{k,0} + \sum_{m=1}^{K} A_{k,m}x_m\right),$$

and

$$\mathbb{E}[p_k^V(\vec{x})] = \sigma_k^V \sum_{\vec{x}}\left(A_{k,0} + \sum_{m=1}^{K} A_{k,m}x_m\right)\pi(\vec{x}) = \sigma_k^V\,\mathbb{E}[W_k] = \sigma_k^V\frac{\nu_k}{C(1-\rho)}.$$

Equating this to $\bar{c}_k^V$ gives σ_k^V. Last, under congestion-based pricing,

$$p_k^L(\vec{x}) = \sigma_k^L|\vec{x}+\vec{e}_k|\left(A_{k,0} + \sum_{m=1}^{K} A_{k,m}x_m\right).$$

Taking the expectation gives

$$\mathbb{E}[p_k^L(\vec{x})] = \sigma_k^L\sum_{\vec{x}}|\vec{x}+\vec{e}_k|\left(A_{k,0} + \sum_{m=1}^{K} A_{k,m}x_m\right)\pi(\vec{x}) = \frac{\sigma_k^L}{1-\rho}\left[A_{k,0} + \frac{2}{1-\rho}\sum_{m=1}^{K} A_{k,m}\rho_m\right],$$

and equating this to $\bar{c}_k^L$ gives σ_k^L. □

PROOF OF PROPOSITION 4.9. The steps for deriving the second moment under congestion-based pricing are outlined here. The proof for the other two pricing models is similar.

$$\mathbb{E}[(p_k^L\ \ (\vec{x})\ \ )^2] = \sum_{\vec{x}}(p_k^L(\vec{x}))^2\pi(\vec{x})$$
$$= (\sigma_k^L)^2\sum_{\vec{x}}|\vec{x}+\vec{e}_k|^2\left(A_{k,0} + \sum_{m=1}^{K} A_{k,m}x_m\right)^2\pi(\vec{x})(\sigma_k^L)^2\sum_{\vec{x}}|\vec{x}+\vec{e}_k|^2A_{k,0}^2\pi(\vec{x})$$
$$+(\sigma_k^L)^2\sum_{\vec{x}}|\vec{x}+\vec{e}_k|^2\sum_{m=1}^{K}A_{k,m}^2x_m^2\pi(\vec{x}) + (\sigma_k^L)^2 2\sum_{\vec{x}}|\vec{x}+\vec{e}_k|^2A_{k,0}\sum_{m=1}^{K}A_{k,m}x_m\pi(\vec{x})$$
$$+(\sigma_k^L)^2\sum_{\vec{x}}|\vec{x}+\vec{e}_k|^2\sum_{i=1}^{K}\sum_{j=1:j\neq i}^{K}A_{k,i}A_{k,j}x_ix_j\pi(\vec{x})$$
$$:= S_1 + S_2 + S_3 + S_4$$

$$S_1 = (\sigma_k^L)^2\sum_{\vec{x}}|\vec{x}+\vec{e}_k|^2A_{k,0}^2\pi(\vec{x}) = \frac{(\sigma_k^L)^2A_{k,0}^2(1-\rho)}{\Phi(\vec{0})}\sum_{\vec{x}}(|\vec{x}|+1)^2\chi(\vec{x})$$
$$= \frac{(\sigma_k^L)^2A_{k,0}^2(1+\rho)}{(1-\rho)^2}.$$

$$S_2 = (\sigma_k^L)^2 \sum_{\vec{x}} |\vec{x} + \vec{e}_k|^2 \sum_{m=1}^{K} A_{k,m}^2 x_m^2 \pi(\vec{x}) = \frac{(\sigma_k^L)^2(1-\rho)}{\Phi(\vec{0})} \sum_{m=1}^{K} A_{k,m}^2 \sum_{n=0}^{\infty} (n+1)^2 s_{m,m}(n)$$
$$= 2(\sigma_k^L)^2 \sum_{m=1}^{K} A_{k,m}^2 \frac{\rho_m(2 + 9\rho_m + 3\rho_m\rho - \rho - \rho^2)}{(1-\rho)^4}.$$

$$S_3 = (\sigma_k^L)^2 \sum_{\vec{x}} |\vec{x} + \vec{e}_k|^2 2A_{k,0} \sum_{m=1}^{K} A_{k,m} x_m \pi(\vec{x})$$
$$= \frac{2A_{k,0}(\sigma_k^L)^2(1-\rho)}{\Phi(\vec{0})} \sum_{m=1}^{K} A_{k,m} \sum_{n=0}^{\infty} (n^2 + 2n + 1) s_m(n)$$
$$= 2A_{k,0}(\sigma_k^L)^2 \frac{(2\rho+4)}{(1-\rho)^3} \sum_{m=1}^{K} A_{k,m}\rho_m.$$

$$S_4 = (\sigma_k^L)^2 \sum_{\vec{x}} |\vec{x} + \vec{e}_k|^2 \sum_{i=1}^{K} \sum_{j=1:j\neq i}^{K} A_{k,i}A_{k,j} x_i x_j \pi(\vec{x})$$
$$= \frac{(\sigma_k^L)^2(1-\rho)}{\Phi(\vec{0})} \sum_{i=1}^{K} \sum_{j=1:j\neq i}^{K} A_{k,i}A_{k,j} \sum_{n=0}^{\infty} (n+1)^2 s_{i,j}(n)$$
$$= \frac{6(\sigma_k^L)^2(3+\rho)}{(1-\rho)^4} \sum_{i=1}^{K} \sum_{j=1:j\neq i}^{K} A_{k,i}A_{k,j}\rho_i\rho_j.$$

Combining S_1, S_2, S_3 and S_4 gives the result. □

PROOF OF PROPOSITION 5.1. Let $\hat{\pi}(\vec{x})$ be the stationary distribution of the system under admission control. Noting that the underlying process is a truncation of a reversible process (see [Kelly 1979, Corollary 1.10]), the stationary distribution is given by

$$\hat{\pi}(\vec{x}) = \pi(\vec{x}) \frac{\sum_{\vec{y}} \pi(\vec{y})}{\sum_{\vec{y}:|\vec{y}|\leq n^*} \pi(\vec{y})} = \pi(\vec{x}) \left(\frac{1}{1-\rho^{n^*+1}} \right).$$

The mean revenue is calculated as in Proposition 4.2 under $\hat{\pi}(\vec{x})$. □

PROOF OF PROPOSITION 5.2. Let $\hat{A}_k$ be the arrival process of class k users. Note that $\hat{A}_k$ is Poisson distributed for $|\vec{x}| < n^*$ and there is no new arrival if $|\vec{x}| = n^*$. The mean payment under the admission control system is

$$\hat{c}_k^{(\cdot)} = \mathbb{E}_{\hat{A}_k} \left[\int_0^{W_k} c_k^{(\cdot)}(\vec{x}(t))dt \right] = \frac{1}{\hat{\lambda}_k} \mathbb{E}\left[x_k c_k^{(\cdot)}(\vec{x}) \right].$$

The second expectation above is under $\hat{\pi}$ instead of π in Proposition 4.4 and the stochastic intensity $\hat{\lambda}_k$ of $\hat{A}_k$ is

$$\hat{\lambda}_k = \sum_{n=0}^{n^*-1} \lambda_k \pi(|\vec{x}| = n) = \lambda_k \left(\frac{1-\rho^{n^*}}{1-\rho^{n^*+1}} \right).$$

Thus, the mean payment under admission control is

$$\hat{c}_k^{(\cdot)} = \frac{1-\rho^{n^*+1}}{1-\rho^{n^*}} \frac{1}{1-\rho^{n^*+1}} \bar{c}_k^{(\cdot)}.$$

□

PROOF OF PROPOSITION 5.3. Let $P_k^F(\vec{x})$ be the random variable denoting the mean payment made by a given class k user when the state is $\vec{x}$. Let $Y(\vec{x})$ be the random variable indicating the payment made by this user in state $\vec{x}$ until the next event (arrival of a new user or departure of an existing user) occurs. Then,

$$\begin{aligned} P_k^F(\vec{x}) = {} & \mathbf{1}_{(\text{next event = arrival})} \sum_{m=1}^{K} [Y(\vec{x}) + P_k^F(\vec{x}+\vec{e}_m)] \\ & + \mathbf{1}_{(\text{next event = other user's departure})} \sum_{m=1}^{K} [Y(\vec{x}) + P_k^F(\vec{x}-\vec{e}_m)] \\ & + \mathbf{1}_{(\text{next event = tagged user's departure})} Y(\vec{x}) \quad \text{for } |\vec{x}| < n^*. \end{aligned}$$

With a slight abuse of notation, let $Y(p;\vec{x})$ and $P_k^F(p;\vec{x})$ respectively denote the probability density function of $Y(\vec{x})$ and $P_k^F(\vec{x})$. Then,

$$Y(p;\vec{x}) = \hat{\omega}(\vec{x}) \frac{|\vec{x}|}{\beta C} e^{-\hat{\omega}(\vec{x})\frac{|\vec{x}|}{\beta C}p},$$

and

$$\begin{aligned} P_k^F(p;\vec{x}) = {} & \sum_{m=1}^{K} \frac{\lambda_m}{\hat{\omega}(\vec{x})} \int_0^p Y(q;\vec{x}) P_k^F(p-q;\vec{x}+\vec{e}_m) dq \\ & + \sum_{m=1}^{K} \frac{(\vec{x}-\vec{e}_k)_m C}{|\vec{x}|\nu_m \hat{\omega}(\vec{x})} \int_0^p Y(q;\vec{x}) P_k^F(p-q;\vec{x}-\vec{e}_m) dq \\ & + \frac{C}{|\vec{x}|\nu_k \hat{\omega}(\vec{x})} Y(p;\vec{x}), \quad \text{for } |\vec{x}| < n^*. \end{aligned}$$

Let $P_k^F(s;\vec{x})$ be the Laplace-Stieltjis Transform (LST) of $P_k^F(p;\vec{x})$. Then, taking the LST of the above gives

$$\beta C P_k^F(s;\vec{x}) = \frac{1}{s + \frac{\hat{\omega}(\vec{x})|\vec{x}|}{\beta C}} \left[\sum_{m=1}^{K} \lambda_m(\vec{x}) P(s;\vec{x}+\vec{e}_m) + \sum_{m=1}^{K} \frac{(\vec{x}-\vec{e}_k)_m C}{\nu_m} P(s;\vec{x}-\vec{e}_m) + \frac{C}{\nu_k} \right],$$

for $|\vec{x}| < n^*$. Taking the derivative of the above once and twice and using

$$P_k^F(s;\vec{x})\big|_{s=0} = 1, \; -\left(P_k^F\right)'(s;\vec{x})\Big|_{s=0} = \eta_k^F(\vec{x}), \text{ and } \left(P_k^F\right)''(s;\vec{x})\Big|_{s=0} = \xi_k^F(\vec{x}),$$

gives (23) and (24). To obtain the second moment of payments by class k users, $\mathbb{E}[\xi_k^F(\vec{x})]$ is evaluated. Note that the state before arrival is $(\vec{x}-\vec{e}_k)$ when a payment of $P_k^F(\vec{x})$ is made.

For $|\vec{x}| = n^*$, the equations are similar except for $\lambda_m = 0$ since no arrivals take place in this state. □

REFERENCES

S. Birmiwal, R. R. Mazumdar, and S. Sundaram. 2012a. Predictable Revenue under Processor Sharing. In *Proceedings of CISS 2012*.

S. Birmiwal, R. R. Mazumdar, and S. Sundaram. 2012b. Processor Sharing and Pricing Implications. In *Proceedings of 24th International Teletraffic Congress*.

T. Bonald and A. Proutiere. 2002. Insensitivity in processor-sharing networks. *Performance Evaluation* 49, 1/4 (2002), 193–209.

P. Brémaud. 1993. A Swiss Army Formula of Palm Calculus. *Journal of Applied Probability* 30 (1993), 40–51.

L.A. DaSilva. 2000. Pricing for QoS-enabled networks: A survey. *Communications Surveys Tutorials, IEEE* 3, 2 (2000), 2–8.

F. M. Guillemin and R. R. Mazumdar. July 2013. Volatility of payment schemes in bandwidth sharing networks. *Preprint, available from www.ece.uwaterloo.ca/∼mazum* (July 2013).

R. A. Horn and C. R. Johnson. 1985. *Matrix Analysis*. Cambridge University Press.

F. P. Kelly. 1979. *Reversibility and stochastic networks*. John Wiley & Sons Ltd., Chichester. Wiley Series in Probability and Mathematical Statistics.

F. P. Kelly. 1997. Charging and rate control for elastic traffic. *European Transactions on Telecommunications* 8 (1997), 33–37.

F. P. Kelly, A. K. Maulloo, and D. K. H. Tan. 1998. Rate Control for Communication Networks: Shadow Prices, Proportional Fairness and Stability. *Journal of Operational Research Society* 49, 3 (1998), 237–252.

K. M. Rege and B. Sengupta. 1994. A Decomposition theorem and related results for the discriminatory processor sharing queue. *Queueing Systems* 18 (1994), 333–351.

R. Schassberger. 1984. A new approach to the $M/G/1$ processor-sharing queue. *Adv. in Appl. Probab.* 16, 1 (1984), 202–213.

D. Songhurst. 1999. *Charging communication networks: from theory to practice*. Elsevier Science B.V.

W. Vickrey. 1961. Counterspeculation, auctions, and competitive sealed tenders. *J. Finance* 16 (1961), 8–37.

R. W. Wilson. 1997. *Nonlinear Pricing*. Oxford University Press. Published in association with the Electric Power Research Institute.

H. Yaiche, R. R. Mazumdar, and C. P. Rosenberg. 2000. A Game Theoretic Framework for Bandwidth Allocation and Pricing in Broadband Networks. *IEEE/ACM Transactions on Networking* 8, 5 (2000), 667–678.

S. Yang and B. Hajek. 2007. VCG-Kelly mechanisms for allocation of divisible goods: Adapting VCG mechanisms to one-dimensional signals. *IEEE Journal on Selected Areas in Communications* 25, 6 (2007), 1237–1243.