

On-line Tuning of Prices for Network Services*

Technical Report SIE-020004

Enrique Campos-Náñez and Stephen D. Patek
Department of Systems and Information Engineering
University of Virginia
{ec3z, patek}@virginia.edu

November 6, 2002

Abstract

Recent investigations into the pricing of multiclass loss networks have shown that static prices are optimal in the asymptotic regime of many small sources. These results suggest that nearly optimal prices for highly aggregated systems can be computed from the solution to a limiting deterministic optimization model. When the assumption of many small sources does not hold, static prices are still preferable (for practical reasons), but we are left with the difficult issue of computing an optimal solution when the stochastic nature of the process cannot be ignored. In this paper, we develop a computational procedure for optimizing static prices that operates by adjusting prices in response to actual customer arrivals and departures and is robust to parametric uncertainty about the underlying system. We provide initial arguments for the convergence properties of our optimization algorithm, and we illustrate its application in several numerical examples.

1 Introduction

The explosive growth of Internet-based industries in the last decade and the recent proliferation of corporate networks raises the question of how to accommodate the ever growing demand for network services. Often, especially in the core of the Internet, the answer is to “over engineer” the system by installing more capacity when the need arises. [Odl98] However, in many contexts, such as in broadband

*This work is supported in part by a grant from the National Science Foundation (ECS-9875688 (CAREER)) and by a scholarship from Consejo Nacional de Ciencia y Tecnología (CONACYT-68428).

access networks, the option to add extra capacity is not so clear, potentially making it necessary to introduce arbitration mechanisms that efficiently allocate resources and regulate network traffic.

Pricing has been studied as one such mechanism. In [GK99], prices are defined on a per-unit-flow basis and are interchanged between a set of distributed regulating entities to achieve fairness. In [CSEZ93], prices per unit flow are calculated to optimize a network objective for differentiated services. Similar approaches can be found in [CDR01, CKSW00, CKW00, CS97, CSS96], where usage-based charging is achieved by communicating resource prices (which can be interpreted as Lagrangian multipliers for network capacity constraints) between the network regulating entities. In [LV93, SKR00], a traditional economic pricing mechanism is used by the (central) owner of the network resources to establish the fees to be charged per unit of each resource involved in a service request. A similar approach is used in [PS99] where the length of the queue at a given network link is used to set prices for network resources, and to achieve congestion control. Game theoretic approaches, that use market mechanisms, such as bidding, to implement pricing mechanisms that work in a distributed manner to achieve a fair use of the resources have been studied in [CSEZ93, FYN88, KM99, SLCL99]. A client-based approach is developed in [Mar01a, Mar01b, Mar02], where prices for a given priority of service are fixed beforehand, but the user (flow) has to decide how much of its flow to send using each priority class. The effect of best-effort traffic has been studied in [CO98, FKSS96, FKSS99, Kel97, LA00, LS01], and more recently in [BS02].

Per-session pricing models in multiclass loss networks have been considered in [FD95, KA99, PT00, PL02], where the resource requirements for accepted calls in each class are reserved by some mechanism, such as RSVP, and the issue is to characterize pricing policies that maximize overall revenue or social welfare. In [PT00] (for the single link case) and [PL02] (for the network case), the near-optimality of static pricing policies has been established. Similar results appear in [LS01] under more general service time distributions.

In this paper, we focus on the single link case and address computational issues associated with the per-session pricing model in [PT00]. Our goal is to develop “on-line” computational algorithms that apply while the network is in operation. So far, the literature on optimal per-session pricing has focused on off-line computation, where the parameters for the multiclass loss network are given in advance and the main issue is to determine an optimal set of prices assuming that the given parameter values are correct. Specifically, in [PT00], Paschalidis and Tsitsiklis compute dynamic pricing policies for a single link model by means of both exact and approximate dynamic programming, and they compare these results to the optimal static prices obtained from the solution to a deterministic optimization model that applies in the many small sources asymptotic regime. Later, in [PL02], Pascha-

lidis and Liu adapt the simulation-based algorithm of [MT01] to compute optimal static prices for the network case (where dynamic programming can be ruled out due to the so-called “curse of dimensionality”).

The main drawback to the computational procedures from [PT00, PL02] is that they do not naturally extend to an on-line setting, where (1) instead of having access to a simulation of the process we have only the process itself and (2) robustness to parametric uncertainty (e.g. arrival rates and holding times) is critical. This paper explores the possibility of adapting the simulation-based procedure of [CNP02] (based on [MT01]) to allow for efficient and robust on-line pricing of network services. Our first contribution is to spell out a version of the simulation-based procedure from [CNP02] that applies to the continuous-time operation of the system and, in the case where arrival rate functions and expected holding times are known, is guaranteed to converge with probability one to an extremum of the average reward process. Next, to address the case where the parameters of the model are unknown (and possibly slowly time varying), we develop an adaptive price-setting algorithm by allowing for the simultaneous estimation of model parameters. Our numerical results show that the adaptive algorithm has the ability to identify nearly-optimal prices, even with significant uncertainty about arrival rates and holding times.

The remainder of this paper is organized as follows. In Section 2, we review the optimization model from [PT00] for pricing services in multiclass loss systems with a single resource. We focus on the problem of revenue maximization with respect to static prices. In Section 3, toward the goal of developing an on-line pricing mechanism, we discuss a model-based optimization procedure (Algorithm 1) adapted from the simulation-based algorithm of [MT01]. Even though this algorithm operates by adjusting prices in response to arrivals and departures in real-time, we refer to it as “model-based” because it requires explicit knowledge of the parameters of the underlying system. (The main difference between Algorithm 1 and the earlier approach from [MT01] is that Algorithm 1 incorporates a device known as i^* -adaptation [CNP02] which resolves an important technical issue associated with the original simulation-based procedure.) In Section 4, we propose an on-line algorithm (Algorithm 2) which is adapted from the model-based procedure of Section 3 and allows for the simultaneous estimation of model parameters. In Section 5, we present a numerical investigation of Algorithms 1 and 2, focusing particularly on the ability of Algorithm 2 to track changes in the parameters of the underlying model. In Section 6, we summarize the paper and outline directions for future research.

2 Problem Formulation

We consider a multiclass loss system with a single resource with capacity C shared by users of K classes of calls or services.¹ Each service class $k \in \{1, 2, \dots, K\}$ can be described by an arrival rate function $\alpha_k(u_k)$, which we assume to be a non-increasing twice differentiable function of a price $u_k \in [0, \bar{u}_k]$, with bounded first and second derivatives. We assume that requests for calls constitute a Poisson arrival process with rate $\alpha(u_k)$ whenever price u_k is used. To disallow the possibility that we “turn off” the arrival process associated with any class, we assume that $\alpha_k(\bar{u}_k) > 0$ for all $k = \{1, 2, \dots, K\}$. Let $\mathcal{U} = \prod_{k=1}^K [0, \bar{u}_k]$ denote the set of all feasible price combinations. The duration of a call of class k is assumed to be exponentially distributed with mean $1/\beta_k$, during which the network reserves m_k out of the total capacity C . At a time t , the state of the system can be described as $i(t) = (i_1(t), i_2(t), \dots, i_K(t))$, a vector that represents the number of ongoing calls of class k present in the system. We use I to denote the set of all feasible usage profiles (i_1, i_2, \dots, i_K) . (The assumption of Poisson arrivals and exponential service is common in the literature on network pricing. For us, the fact that the system is memoryless is critical, at least for motivating the computational techniques that we develop.)

Finally, we assume that the system implements a strict admission control rule. Suppose that at time t a request for a call of type k is received. The call will be admitted if $m \cdot i(t) + m_k \leq C$, where the vector m is defined by $m = (m_1, m_2, \dots, m_K)$. To simplify notation, define $A(i) = \{k \mid m \cdot i + m_k \leq C\}$ to be the set of calls that satisfy the admission control rule when the system is in state $i \in I$.

2.0.1 Revenue Maximization

In this paper, we focus on the problem of setting prices to shape the arrival process so that we maximize the expected rate at which revenue accrues to the service provider. Define Π to be the set of mappings $\pi : I \rightarrow \mathcal{U}$ that assign a set of prices to each state $i \in I$. The revenue maximization problem is then defined as

$$\max_{\pi \in \Pi} \lim_{T \rightarrow \infty} \frac{1}{T} E \left[\int_{t=0}^T \sum_{k \in A(i(t))} \alpha_k(\pi_k(i(t))) \pi_k(i(t)) dt \right].$$

As discussed in [PT00], there is no dependence of the average reward on the initial state $i(0)$, and we use λ^* to denote the optimal value of the problem.

¹As an example, consider a broadband access network with a single uplink to Internet.

2.0.2 Maximizing Social-Welfare

The modeling framework above also allows us to capture the problem of maximizing social welfare, as follows. Assume that the utility of a call of type k is a random variable U_k and that the user of the system will initiate a call request whenever the current price is lower than the utility perceived, i.e. $U_k \geq u_k$. The network, as a control entity might be interested in maximizing the expected utility of calls admitted into the system. In this case the optimization problem is to maximize the following objective with respect to $\pi \in \Pi$:

$$\lim_{T \rightarrow \infty} \frac{1}{T} E \left[\int_{t=0}^T \sum_{k \in A(i(t))} \alpha_k(\pi_k(i(t))) \phi_k(\pi_k(i(t))) dt \right],$$

where $\phi_k(\pi_k(i(t))) = E[U_k | U_k \geq \pi_k(i(t))]$ is the expected utility of an accepted class- k call. Again, we do not indicate any dependence on the initial condition $i(0)$ because of the fact that the limit exists and is the same for all initial states. Both the revenue maximization and social welfare objectives are of the average reward form, and can be solved using traditional dynamic programming techniques (see, for example, [Ber95]).

2.1 Static Pricing Policies

For simplicity, we focus on the revenue maximization problem throughout the rest of this paper. Also, we restrict attention a subset of Π , namely the so-called static pricing policies, where, for a given solution, the same class-dependent price u_k always applies regardless of the state of the process. The revenue maximization problem then can be expressed as $\max_{u \in \mathcal{U}} \lambda(u)$, where

$$\lambda(u) = \lim_{t \rightarrow \infty} \frac{1}{T} E \left[\int_0^T \sum_{k \in A(i(t))} \alpha_k(u_k) u_k dt \right]. \quad (1)$$

2.2 Discussion of Previous Results

A theoretical justification for restricting attention to static pricing policies is given in [PT00], where static prices are shown to be optimal in the asymptotic regime of “many small sources.” More precisely, consider a scaled version of the system which is obtained by scaling the original capacity of the link as $C^{(c)} = cC$, where $c > 1$, and by scaling the original arrival rate functions as $\alpha_k^{(c)}(u_k) = c\alpha_k(u_k)$. Let λ_c^* be the average reward obtained by an optimal dynamic policy, and let $\lambda(u_c^*)$ be the average reward associated with an optimal static policy. A key result

from [PT00] is that $\lim_{c \rightarrow \infty} \lambda_c^* = \lim_{c \rightarrow \infty} \lambda(u_c^*)$, i.e. that static pricing policies are optimal in the asymptotic regime. Consequently, nearly optimal static prices exist for cases where the bandwidth requirements m are small compared to the capacity C . This result has been extended in [LS01] to general service distributions.

While the near-optimality of static pricing policies is reassuring, if the assumption of many small users does not apply (or is questionable), the solution to the deterministic optimization model that gives rise to the upper bound in [PT00, PL02, LS01] may be far from optimal, and we must deal with the practical question of how to compute optimal static prices.

2.3 On-line Pricing Algorithms

Our goal in this paper is to make progress toward “on-line” pricing algorithms, where static prices for services are adjusted in response to observations of the system in operation. We seek to achieve the following characteristics in our computational approach.

- *Real-Time Operation:* Our price optimization mechanism should operate with respect to observations of the actual system in operation. That is, rather than assuming access to a simulation of the process (which could be used for off-line calculations) we only require access to the real-time stream of arrivals and departures of calls of various classes.
- *Efficient Operation:* Our computational algorithm should have no need to store per-call information and should maintain only $O(K)$ state variables, where K is the number of classes of calls.
- *Robust Operation:* Our optimization procedure should have the ability to discover and respond to uncertain and possibly time-varying parameters of the underlying system.

In Section 3 below, we develop a model-based on-line price-setting algorithm which achieves the first two characteristics above. We refer to this algorithm as “model-based” because it cannot be implemented without knowledge of the parameters of the model, particularly the arrival rate functions $\alpha_k(u_k)$ and the expected holding times $1/\beta_k$. In Section 4, we extend the model-based algorithm so that it can adapt to unknown and slowly time-varying parameters, thereby addressing the third characteristic above.

3 Model-Based Pricing

The pricing model of Section 2 can be viewed as a continuous-time Markov reward process, where (in the case of revenue maximization) the objective is to maximize the expected average rate at which reward accrues to the service provider. As is done in [PT00, PL02], it is possible to uniformize the process (see, for example, Chapter 5 of [Ber95]) to obtain an equivalent discrete-time description of the process. Conceptually, the uniformization involves Poisson sampling of the process at a rate ν^* equal to (or higher than) the fastest possible transition rate of the continuous-time system. For the pricing model of Section 2, the rate of transitions out of state $i \in I$ is

$$\nu_i(u) = \sum_{k \in A(i)} \alpha_k(u_k) + \sum_{k=1}^K i_k \beta_k, \quad (2)$$

from which we obtain the following upper bound on the rate of transitions

$$\nu^* = \sum_{k=1}^K \lfloor C/m_k \rfloor \beta_k + \alpha_k(0). \quad (3)$$

With ν^* in hand, we may infer discrete-time state transition probabilities, as well as expected rewards associated with accepted call arrivals, and, in principle, we may apply any algorithmic technique available to us to optimize the average-reward of discrete-time Markov processes. In this section, we adapt the discrete-time simulation-based methodology of [MT01] (as extended in [CNP02, CNP01]) for on-line use in a continuous-time setting.

3.1 An Algorithm to Compute Optimal Prices

The rationale behind the discrete-time algorithm of [MT01] is (1) to exploit the regenerative structure of the underlying Markov process to build asymptotically unbiased estimates of the gradient of the objective function and (2) to use the estimates of the gradient in adjusting the control parameters (i.e. prices) toward an extremum of the average reward function. An important feature of the methodology in [MT01] is that updates to the controllable parameters occur only when the system transitions into a specially marked recurrent state $i^* \in I$. Unfortunately, for a given set of control parameters, transitions to i^* can be extremely infrequent, resulting in high-variance estimates of the gradient and ultimately in unacceptably slow convergence. This issue has been addressed in [CNP02, CNP01] through the introduction of an i^* -adaptation procedure, where the marked state i^* can be reset

as needed to account for changes in the steady-state distribution of the process as the control parameters are adjusted.

In Algorithm 1 below, we present a price-setting algorithm, which can be interpreted as a continuous-time implementation of the discrete-time algorithm in [CNP02, CNP01]. Since the algorithm proceeds in response to arrival and departure events of the actual process, it applies in an on-line setting.

Algorithm 1 - Model-Based Price Optimization

Given scale parameter $\eta > 0$, stepsize rule $\{\gamma_m : m = 1, 2, \dots\}$ such that $\sum_m \gamma_m = \infty$, and $\sum_m \gamma_m^2 < \infty$, initial observation period threshold $\tau > 0$, initial set of prices $(u_{0,1}, u_{0,2}, \dots, u_{0,K})$ for service classes $1, \dots, K$, and an estimate $\tilde{\lambda}$ of the average reward associated with the initial set of prices, compute recursively $\{(u_m, \tilde{\lambda}_m, \tau_m, i_m^*, s_m) : m = 0, 1, \dots\}$ as follows.

(Initialization)

Set $m = 0$, $u_0 = (u_{0,1}, u_{0,2}, \dots, u_{0,K})$, $\tilde{\lambda}_0 = \tilde{\lambda}$, and $\tau_0 = \tau$. Set t_0 to be the time at which the algorithm is initialized, and set i_0^* to be the state of the system at time t_0 . Set the initial timeout threshold $s_0 = t_0 + \tau_0$. Finally, set local variables $F = \mathbf{0} \in \mathfrak{R}^K$, $L = \mathbf{0} \in \mathfrak{R}^K$, and $G = 0$.

(n -th State Transition)

Let t_n denote the time of the n -th state transition, and let $i_n = i(t_n^+)$ be the new state of the system at that time. Proceed as follows.

1. (Update Local Variables) First, generate a Poisson random variable T as

$$T \sim \text{Poisson}((t_n - t_{n-1})(\nu^* - \nu_{i(t_{n-1})}(u_m))). \quad (4)$$

Next, update F , L , and G as follows.

$$F = F + T \cdot L \left(g_{i_{n-1}}(u_m) - \tilde{\lambda}_m \right) \quad (5)$$

$$- \frac{T^2 + T}{2} (g_{i_{n-1}}(u_m) - \tilde{\lambda}_m) l_{i_{n-1}}(u_m) \quad (6)$$

$$+ T \nabla g_{i_{n-1}}(u_m), \quad (7)$$

$$L = L + T l_{i_{n-1}}(u_m), \text{ and} \quad (8)$$

$$G = G + T (g_{i_{n-1}}(u_m) - \tilde{\lambda}_m). \quad (9)$$

In addition, if the n -th state transition corresponded to a class- k arrival, then update L again as

$$L = L + \sum_{k \in A(i_{n-1})} \frac{\alpha'_k(u_{m,k})}{\alpha_k(u_{m,k})} e_k,$$

and use L to revise F and G as

$$\begin{aligned} F &= F + T \cdot L \left(g_{i_n}(u_m) - \tilde{\lambda}_m \right) + \nabla g_{i_n}(u_m), \\ G &= G + T(g_{i_n}(u_m) - \tilde{\lambda}_m). \end{aligned}$$

2. (Regenerative Cycle Complete) *If the n -th state transition resulted in a return to the state i_m^* , then update the price vector and the average reward estimate as*

$$u_{m+1} = u_m + \gamma_m F, \tag{10}$$

$$\tilde{\lambda}_{m+1} = \tilde{\lambda}_m + \eta \gamma_m G. \tag{11}$$

Leave the observation period threshold and marked state unchanged as $\tau_{m+1} = \tau_m$ and $i_{m+1}^ = i_m^*$. Set a new timeout threshold $s_{m+1} = t_n + \tau_{m+1}$. Finally, reset the local variables $F = \mathbf{0}$, $L = \mathbf{0}$, and $G = 0$, and set $m = m + 1$.*

(Timeout)

If real time t ever reaches the current timeout threshold s_m , then proceed as follows.

Increase the observation period threshold as $\tau_{m+1} = \tau_m + 1/\nu^$. Leave the price vector and average reward estimate unchanged as*

$$\begin{aligned} u_{m+1} &= u_m, \\ \tilde{\lambda}_{m+1} &= \tilde{\lambda}_m. \end{aligned}$$

Set $i_{m+1}^ = i(t)$ and $s_{m+1} = t + \tau_{m+1}$. Finally, reset the local variables as $F = \mathbf{0}$, $L = \mathbf{0}$, and $G = 0$, and set $m = m + 1$.*

3.2 Discussion

Algorithm 1 works by collecting reward and likelihood information during regenerative cycles of the process to obtain asymptotically unbiased estimates of the gradient of the objective function. The “local” variables, F , L , and G , all have interpretations similar to those in [MT01, CNP02]. Specifically, at the end of a regenerative cycle, the vector F contains an estimate of the gradient of the objective function, and this estimate is used in Eqn. (10) to update the set of prices for services. The vector L contains likelihood information about the observed system trajectory, and the scalar G contains information about the objective function and is used in Eqn. (11) to improve the estimate $\tilde{\lambda}_m$. Note the small memory footprint

of this algorithm: we need only store and manipulate $(3K + 5)$ real variables as the optimization procedure evolves.

As mentioned earlier, Algorithm 1 can be interpreted as a continuous-time version of the optimization procedure of [MT01, CNP02] (which was originally developed for discrete-time Markov chains). The conversion from discrete-time to continuous time here is mostly straightforward, although there is one peculiar feature of the algorithm that should be addressed, namely the “Augmented Markov Process” where we insert auxiliary “no change” state transition events into the actual process of customer arrivals and departures. Recall that the constants ν^* and $\nu_i(u)$ are, respectively, an upper bound on the transition rate of the Markov chain and the transition rate out of state i when the price vector u is used. Recall also that the equivalent discrete-time model of the continuous-time process is constructed from Poisson sampling at rate ν^* . Thus, $[\nu^* - \nu_{i(t_{n-1})}(u_m)]$ corresponds to the rate at which samples are drawn without actual changes to the state of the underlying continuous-time system. These samples correspond to “self-transitions” in the discrete-time model, which figure significantly in the evolution of the optimization algorithm. In mapping back from discrete-time, it is necessary to introduce “no change” events into the actual stream of arrivals and departures to account for the effect of self-transitions in the equivalent discrete-time model.

The subroutine “(Timeout)” provides a continuous-time implementation of the i^* -adaptation procedure from [CNP02, CNP01], which serves to prevent the optimization algorithm from getting stuck in lengthy regenerative cycles. To recognize the need for such a feature recall that updates to the price vector occur only at renewal points (i.e. upon transitions to the specially marked recurrent state $i^* \in I$). As the optimization algorithm progresses and an optimal set of prices emerges, the set of states that are likely to be visited frequently can vary significantly, and it can be difficult to choose a priori a fixed state to serve as i^* throughout the entire evolution of the process. The i^* -adaptation procedure avoids this difficulty by setting an observation period threshold τ that is enforced in the sense that only regenerative cycles shorter than τ are allowed to impact the evolution of prices (i.e. regenerative cycles longer than τ are ignored). As discussed in [CNP02], to avoid biasing the results, it is critical for the threshold τ to be incremented over time so that, asymptotically, no regenerative cycles get dropped. In present context, the i^* -adaptation procedure is implemented as a timeout process, where the current observation period threshold τ_m is used to set a timeout epoch s_m at which time the marked state will be reset to a better value (specifically, the current state i_n), and the data collected up until the timeout period is ignored.

Theorem 1 (Convergence of Algorithm 1) *Given the stepsize rule $\gamma_m = a/(b + m)$, with $a, b > 0$, and assuming that the model parameters $\alpha_k(\cdot)$ and β_k are*

known for all service classes $k = 1, \dots, K$, Algorithm 1 converges with probability one to critical point u^* where $\nabla\lambda(u^*) = 0$.

This theorem follows directly from the main result of [CNP02]. (Note that, under the modeling assumptions Section 2, all states of the underlying process are recurrent. Thus, the hypotheses of [CNP02] hold.) A small difference introduced by replacing the discrete-time threshold, which corresponds to a fixed number of steps, with a fixed time interval for the continuous-time is shown in the Appendix to have no effect on the convergence of the algorithm. We illustrate Algorithm 1 in action in Section 5 after first discussing an adaptive version of the algorithm designed to be robust to parametric uncertainty about the process.

4 Adaptive Pricing

One of the shortcomings of the algorithm of the preceding section is the assumption of knowledge of all the parameters involved in the model. In this section, we extend the model-based algorithm to allow for the simultaneous estimation of model parameters and provide initial arguments for the convergence properties of the resulting adaptive price-setting algorithm.

To set the stage for this discussion, we assume that arrival rate functions $\alpha_k(\cdot)$ and service rates β_k are parameterized by a set of fixed but unknown parameters $\theta^* \in \mathbb{R}^q$, which is assumed to belong to a compact set $\Theta \subset \mathbb{R}^q$. As before, we assume that requests for service arrive according to independent Poisson processes with rates $\alpha_1(u_1, \theta), \dots, \alpha_K(u_K, \theta)$ and that service times are exponentially distributed with rates $\beta_1(\theta), \dots, \beta_K(\theta)$. We assume that all of the parameters α_k and β_k are twice differentiable in (u, θ) , with bounded first and second derivatives. Analogously to the preceding section, we define

$$\nu_i(u, \tilde{\theta}) = \sum_{k \in A(i)} \alpha_k(u_k, \tilde{\theta}) + i_k \beta(\tilde{\theta}),$$

and we assume knowledge of a uniform bound ν^* on the transition rate of the underlying process, i.e.

$$\nu^* \geq \nu_i(u, \theta), \quad \forall u \in \mathcal{U}, \theta \in \Theta, i \in I.$$

Our objective from here is to create an efficient and robust on-line procedure for optimizing static prices $u \in \mathcal{U}$ while at the same time estimating the parameters $\theta^* \in \Theta$.

Algorithm 2 - Adaptive Price Optimization

Given scale parameters $\eta > 0$ and $\mu > 0$, stepsize rule $\{\gamma_m : m = 1, 2, \dots\}$ such that $\sum_m \gamma_m = \infty$, and $\sum_m \gamma_m^2 < \infty$, initial observation period threshold $\tau > 0$, initial set of prices $(u_{0,1}, u_{0,2}, \dots, u_{0,K})$ for service classes $1, \dots, K$, an estimate $\tilde{\lambda}$ of the average reward associated with the initial set of prices, and an initial estimate $\tilde{\theta}_0$ of the unknown parameters θ^* , compute recursively $\{(u_m, \tilde{\lambda}_m, \tilde{\theta}_m, \tau_m, i_m^*, s_m) : m = 0, 1, \dots\}$ as follows.

(Initialization)

Set $m = 0$, $u_0 = (u_{0,1}, u_{0,2}, \dots, u_{0,K})$, $\tilde{\lambda}_0 = \tilde{\lambda}$, $\tilde{\theta}_0 = \tilde{\theta}_0$, and $\tau_0 = \tau$. Set t_0 to be the time at which the algorithm is initialized, and set i_0^* to be the state of the system at time t_0 . Set the initial timeout threshold $s_0 = t_0 + \tau_0$. Finally, set local variables $F = \mathbf{0} \in \Re^K$, $L = \mathbf{0} \in \Re^K$, $\mathcal{L} = \mathbf{0} \in \Re^K$, and $G = 0$.

(n -th State Transition)

Let t_n denote the time of the n -th state transition, and let $i_n = i(t_n^+)$ be the new state of the system at that time. Proceed as follows.

1. (Update Local Variables) First, generate a Poisson random variable T with mean $(t_n - t_{n-1})\nu^*$,

$$T \sim \text{Poisson}((t_n - t_{n-1})(\nu^* - \nu_{i_n}(u_m, \tilde{\theta}_m))).$$

Next, update F, L, \mathcal{L} and G as follows.

$$\begin{aligned} F &= F + T \cdot L \left(g_{i_{n-1}}(u_m, \tilde{\theta}_m) - \tilde{\lambda}_m \right) \\ &\quad - \left[\frac{T(T+1)}{2} (g_{i_{n-1}}(u_m, \tilde{\theta}_m) - \tilde{\lambda}_m) \right. \\ &\quad \left. \cdot l_{i_{n-1}}(u_m, \tilde{\theta}_m) \right] \\ &\quad + T \nabla g_{i_{n-1}}(u_m, \tilde{\theta}_m), \\ L &= L + T \cdot l_{i_{n-1}}(u_m, \tilde{\theta}_m), \\ \mathcal{L} &= \mathcal{L} - T \frac{\nabla_{\tilde{\theta}_m} \nu_{i_{n-1}}(u_m, \tilde{\theta}_m)}{\nu^* - \nu_{i_{n-1}}(u_m, \tilde{\theta}_m)}, \\ G &= G + T (g_{i_{n-1}}(u_m, \tilde{\theta}_m) - \tilde{\lambda}_m). \end{aligned}$$

In addition, if the n -th state transition corresponded to a class- k arrival,

then update L and \mathcal{L} again as

$$\begin{aligned} L &= L + \sum_{k \in A(i_{n-1})} e_k \frac{\alpha'_k(u_{m,k}, \tilde{\theta}_m)}{\alpha_k(u_{m,k}, \tilde{\theta}_m)}, \\ \mathcal{L} &= \mathcal{L} + \frac{\nabla_{\tilde{\theta}_m} \alpha_k(u_m, \tilde{\theta}_m)}{\alpha_k(u_m, \tilde{\theta}_m)}. \end{aligned}$$

Otherwise, if the n -th state transition corresponded to a class- k departure, then update \mathcal{L} again as

$$\mathcal{L} = \mathcal{L} + \frac{\nabla_{\tilde{\theta}_m} \beta_k(\tilde{\theta}_m)}{\beta_k(\tilde{\theta}_m)}$$

(Here, the derivative is taken with respect to the estimate of the parameters $\tilde{\theta}$). Finally, use result of the event-specific calculations above to make final adjustments to F and G as follows.

$$\begin{aligned} F &= F + T \cdot L \left(g_{i_n}(u_m, \tilde{\theta}_m) - \tilde{\lambda}_m \right) \\ &\quad + \nabla g_{i_n}(u_m, \tilde{\theta}_m), \\ G &= G + T(g_{i_n}(u_m, \tilde{\theta}_m) - \tilde{\lambda}_m). \end{aligned}$$

2. (Regenerative Cycle Complete) If the n -th state transition resulted in a return to the state i_m^* , then update the price vector, the average reward estimate, and the parameter vector estimate as

$$\begin{aligned} u_{m+1} &= u_m + \gamma_m F, \\ \tilde{\lambda}_{m+1} &= \tilde{\lambda}_m + \eta \gamma_m G, \\ \tilde{\theta}_{m+1} &= \tilde{\theta}_m + \mu \gamma_m \mathcal{L}. \end{aligned} \tag{12}$$

Leave the observation period threshold and marked state unchanged as $\tau_{m+1} = \tau_m$ and $i_{m+1}^* = i_m^*$. Set a new timeout threshold $s_{m+1} = t_n + \tau_{m+1}$. Finally, reset the local variables $F = \mathbf{0}$, $L = \mathbf{0}$, $\mathcal{L} = \mathbf{0}$, and $G = 0$, and set $m = m + 1$.

(Timeout)

If real time t ever reaches the current timeout threshold s_m , then proceed as follows.

Increase the observation period threshold as $\tau_{m+1} = \tau_m + 1/\nu^*$. Leave the price vector, the average reward estimate, and the parameter vector estimate unchanged as

$$\begin{aligned} u_{m+1} &= u_m, \\ \tilde{\lambda}_{m+1} &= \tilde{\lambda}_m, \\ \tilde{\theta}_{m+1} &= \tilde{\theta}_m. \end{aligned} \tag{13}$$

Set $i_{m+1}^* = i(t)$ and $s_{m+1} = t + \tau_{m+1}$. Finally, reset the local variables as $F = \mathbf{0}$, $L = \mathbf{0}$, $\mathcal{L} = \mathbf{0}$, and $G = 0$, and set $m = m + 1$.

4.1 Discussion

As an extension of the model-based algorithm of the preceding section, Algorithm 2 works by collecting reward and likelihood information during regenerative cycles of the process to obtain estimates of the gradient of the objective function. All of the local variables from the model-based algorithm show up here, with the addition of \mathcal{L} which collects likelihood information used in Eqn. (12) to improve estimates of the unknown parameters θ^* . Note that, while Algorithm 2 uses the estimates $\tilde{\theta}_m$ to perform the same optimization procedure described in Algorithm 1, it simultaneously uses observed system behavior to further improve the estimates $\tilde{\theta}_m$. In this sense, Algorithm 2 is similar in spirit to earlier algorithms developed for the adaptive control of Markov chains (see, for example, [BV79]). Again, note the small memory footprint of the algorithm: we need only store and manipulate $(3K + 2q + 5)$ real variables as the optimization procedure evolves, where q is the dimension of the parameter vector $\tilde{\theta}$ and K is the number of the service classes (as usual).

The process by which Algorithm 2 estimates the parameter vector θ^* [cf. Eqns. (12-13)] has the following rationale. Suppose a Markov chain evolves according to $P_{ij}(\theta^*) = P(i_{n+1} = j \mid i_n = i, \theta^*)$, with θ^* unknown. Upon observing the state trajectory $\{i_0, i_1, \dots\}$, we could estimate θ^* by identifying the parameter vector $\tilde{\theta}^*$ that maximizes the likelihood of observing $\{i_0, i_1, \dots\}$. Algorithm 2 instantiates this idea by updating $\tilde{\theta}$ in the direction of the gradient of the likelihood associated with each observed regenerative cycle. Specifically, given an observed regenerative cycle $\{i_0, i_1, \dots, i_T\}$, the likelihood of observing this particular trajectory under θ_m and the prevailing set of prices u_m is given by

$$\mathcal{L}(u_m, \tilde{\theta}_m) = \prod_{n=1}^T P_{i_{n-1}i_n}(u_m, \tilde{\theta}_m),$$

where $P_{ij}(u_m, \tilde{\theta}_m)$ denotes the state transition probability of the uniformized (discrete-time) model that is associated with u_m and $\tilde{\theta}_m$. To improve the estimate $\tilde{\theta}_m$, we make a small stochastic approximation-type adjustment in the direction of $\nabla_{\tilde{\theta}} \mathcal{L}(u_m, \tilde{\theta}_m)$. Algorithm 2 accomplishes this in Eqns. (13-12) by making an update in the direction

$$\sum_{n=1}^T \frac{\nabla_{\tilde{\theta}} P_{i_{n-1}i_n}(u_m, \tilde{\theta}_m)}{P_{i_{n-1}i_n}(u_m, \tilde{\theta}_m)}. \quad (14)$$

The rationale for this is that

$$\nabla_{\tilde{\theta}} \mathcal{L}(u_m, \tilde{\theta}_m) = \mathcal{L}(u_m, \tilde{\theta}_m) \sum_{n=1}^T \frac{\nabla_{\tilde{\theta}} P_{i_{n-1}i_n}(u_m, \tilde{\theta}_m)}{P_{i_{n-1}i_n}(u_m, \tilde{\theta}_m)}.$$

Thus, update in the direction of Eqn. (14) differs from $\nabla_{\tilde{\theta}} \mathcal{L}(u_m, \tilde{\theta}_m)$ only by a scaling factor $\mathcal{L}(u_m, \tilde{\theta}_m)$. Our numerical results in Section 5 illustrate the effectiveness of this procedure.

4.2 Analytical Issues with Algorithm 2

The fact that Algorithm 1 converges with probability one to a critical point stems from the observation in [MT01] that the local variable F (at the end of a regenerative cycle) is asymptotically consistent with the direction of the gradient of the objective function. The analyses of [MT01] and [CNP02] (relating to the i^* -adaptation procedure) involve a careful accounting of the errors associated with imperfect estimates of average reward of the process $\lambda(u_m)$. The analysis of Algorithm 2 is complicated by the fact that new errors are introduced from imperfect estimates of the parameter vector θ^* . In order to prove that Algorithm 2 offers any beneficial convergence properties, we must resolve at least two issues:

1. *Does the optimization procedure allow enough observation of the process in order to guarantee the convergence?* - From the theory of adaptive control of Markov chains (see, for example, [Bor91], pages 121-122), conditions are known under which the optimization procedure will converge when the maximum likelihood estimator is used. In Algorithm 2, however, we use a gradient procedure to estimate θ^* , and convergence of the estimate remains to be shown. Moreover, if $\{\tilde{\theta}_m\}_{m=0}^{\infty}$ converges in Algorithm 2, it may converge to a locally optimal estimate.
2. *Does the error $(\theta^* - \tilde{\theta}_m)$ introduce enough noise to prevent the recovery of the true direction of the gradient?* - If the functions $\alpha_k(u_k, \tilde{\theta})$, and $\beta_k(\tilde{\theta})$ are such that the assumptions in [Bor91] are met, then it is possible to show

that the direction of the price vector update deviates from the gradient by a term that can be bounded by $M\|\theta^* - \tilde{\theta}_m\|$, for some constant M . This may be enough to guarantee that the direction of the gradient will be eventually recovered (of course as long as $\|\theta^* - \tilde{\theta}_m\| \rightarrow 0$.)

5 Numerical Evaluation

Here, we present numerical examples that illustrate the application of Algorithms 1 and 2. Examples 1 and 2 illustrate (1) the convergence of the model-based algorithm (Algorithm 1) assuming perfect knowledge of the arrival rate functions and holding times and (2) the ability of the adaptive algorithm (Algorithm 2) to identify the same set of prices in the presence of uncertainty about some of the model parameters. Examples 3 and 4 illustrate the ability of the adaptive algorithm to track time-varying parameters of the underlying system. All scenarios tested here correspond to situations where the “many small sources” assumption does not apply.

5.1 Example 1: Two Classes of Traffic

This example involves two classes of traffic ($K = 2$) in a system with small capacity. The parameters for the example are compiled in Table 1. Note that the capacity of the system is $C = 10.0$ (*Mbps*) and, since $m_1 = 1.0$ (*Mbps*) and $m_2 = 5.0$ (*Mbps*), there can be at most 10 active users in the system. Both service classes have a maximum (price-free) request arrival rate of $\bar{\alpha}_1 = \bar{\alpha}_2 = 10.0$ (*requests/s*). On the other hand, class-1 demand for services drops (linearly) to zero at a price of $\check{u}_1 = 1.0$ *priceunits(p.u.)*, whereas class-2 demand for service cuts off at the much higher value $\check{u}_2 = 10.0$ (*p.u.*). Even though class-2 users consume 5 times the bandwidth of class-1 users and stay in the system for the same length of time ($1/\beta_1 = 1/\beta_2 = 1.0$ (*s*) on average), the demand for class-2 service is so much higher than class-1 service for a given price that class 2 traffic offers better earning potential to the service provider. In applying Algorithms 1 and 2, we restrict the range of prices for class-1 service to the interval $u_1 \in [0, .9]$ (*p.u.*), and, similarly, we restrict the range of prices for class-2 service to the interval $u_2 \in [0, 9.0]$ (*p.u.*). Thus, it is impossible for the optimization algorithm to accidentally “turn off” the process of request arrivals in either class. For the adaptive case, we assume that service rate distribution parameters β_1 and β_2 are unknown.

The results of this example appear in Figs. 1 and 2. Fig. 1 shows the evolution of prices and estimated average reward over 10^6 state transitions in the Augmented Markov Process for both algorithms. Note that both the model-based and the adap-

Table 1: Parameters for Example 1. ($C = 10$)

| Parameter | Class 1 | Class 2 |
|------------------|---------------------------------------|---------------------------------------|
| $\alpha_k(u_k)$ | $\bar{\alpha}_1(1 - u_1/\bar{u}_1)_+$ | $\bar{\alpha}_2(1 - u_2/\bar{u}_2)_+$ |
| $\bar{\alpha}_k$ | 10.0 | 10.0 |
| \bar{u}_k | 1.0 | 10.0 |
| \bar{u}_k | .9 | 9.0 |
| m_k | 1.0 | 5.0 |
| β_k | 1.0 | 1.0 |

tive algorithms converge to nearly the same set of prices, $u_1 \approx .9$ and $u_2 \approx 7.0$. (This price vector has the effect of shutting off class-1 traffic as much as possible, leaving as much room as possible for the other, more profitable, service class.) Fig. 2 shows the evolution of the estimates of β_1 and β_2 in Algorithm 2. Note that the estimates eventually converge to the true value $\beta_1 = \beta_2 = 1.0$ (s^{-1}). We point out the relatively slow rate of convergence of both algorithms, especially with regard to price. Note that the results are plotted over 10 hours of the real-time operation of the system. On the other hand, even though it takes a long time for prices to settle, the average reward of the process seems to lock onto its final value relatively quickly.

5.2 Example 2: Three Classes of Traffic

This example involves three classes of traffic, with guaranteed rates $m_1 = 1$, $m_2 = 10$, $m_3 = 5.0$ (*Mbps*), in a system with capacity $C = 20$ (*Mbps*). The full set of parameters for the example are compiled in Table 2. As in Example 1, prices are optimized both for the case where all the parameters are known (Algorithm 1) and where the service rate parameters β_1, β_2 , and β_3 are unknown (Algorithm 2).

Table 2: Parameters for Example 2. ($C = 20$)

| Parameter | Class 1 | Class 2 | Class 3 |
|------------------|---------------------------------------|---------------------------------------|---------------------------------------|
| $\alpha_k(u_k)$ | $\bar{\alpha}_1(1 - u_1/\bar{u}_1)_+$ | $\bar{\alpha}_2(1 - u_2/\bar{u}_2)_+$ | $\bar{\alpha}_3(1 - u_3/\bar{u}_3)_+$ |
| $\bar{\alpha}_k$ | 10.0 | 10.0 | 10.0 |
| \bar{u}_k | 1.0 | 10.0 | 5.0 |
| \bar{u}_k | .9 | 9.0 | 4.8 |
| m_k | 1.0 | 10.0 | 5.0 |
| β_k | 10.0 | 1.0 | 5.0 |

The results of this example appear in Figs. 3 and 4. Fig. 3 shows the evolution of prices and estimated average reward over 10 hours of operation. Note that, as before, both the model-based and the adaptive algorithms converge to nearly the same set of prices. Fig. 2 shows the evolution of the estimates of β_1, β_2 , and β_3 in Algorithm 2, with all three parameters eventually tracking their actual values.

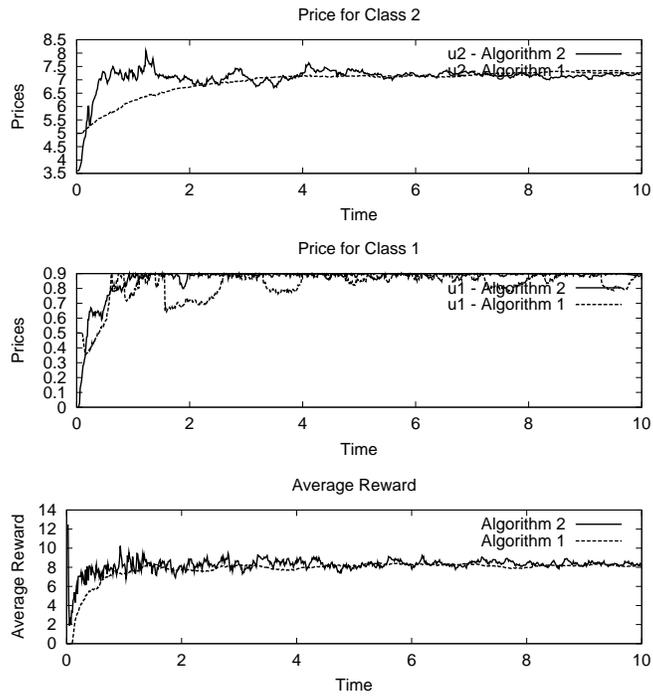


Figure 1: Evolution of prices and estimated average reward in Example 1.

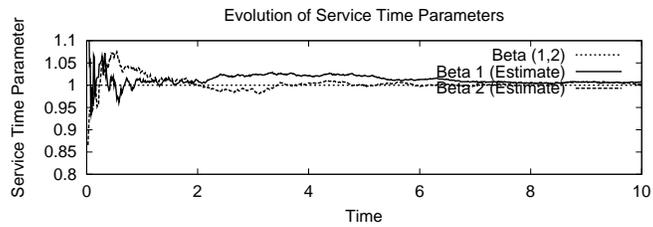


Figure 2: Evolution of the estimates of β in Example 1.

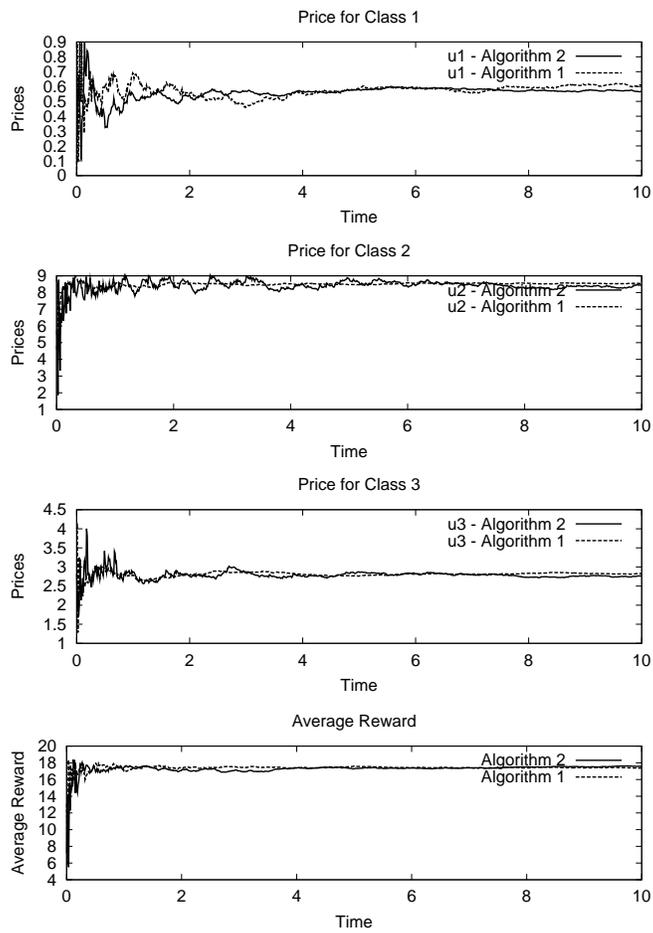


Figure 3: Evolution of prices and estimated average reward in Example 2.

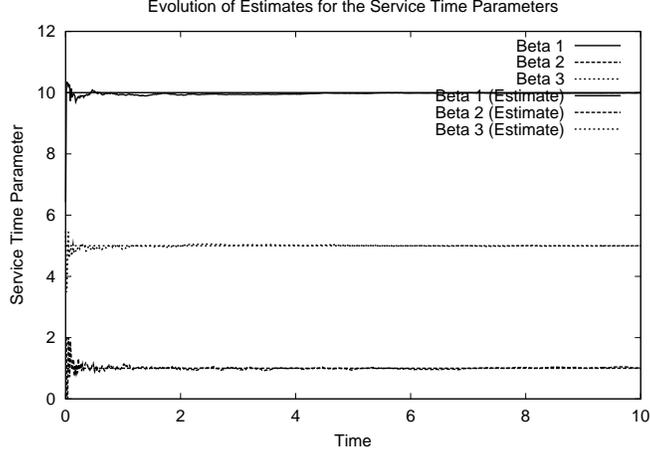


Figure 4: Evolution of the estimates of $\beta_1, \beta_2, \beta_3$ in Example 2.

5.3 Example 3: Tracking Abrupt Changes

This example seeks to explore the ability of Algorithm 2 to track abrupt changes in model parameters. As in Example 1, we consider a system with two classes of service and limited service capacity $C = 10$ (*Mbps*). The parameters of the model are expressed in Table 3. Note that the service rate parameters β_1 and β_2 are given as functions of time, i.e. for the first third of the experiment $\beta_1 = .1$ and $\beta_2 = 1.0$, for the second third of the experiment $\beta_1 = 1.0$ and $\beta_2 = .1$, and for the final third of the experiment we return to $\beta_1 = .1$ and $\beta_2 = 1.0$. Note that, throughout the experiment, both service classes have identical demand characteristics (i.e. their arrival rate functions $\alpha_k(\cdot)$ are the same). Finally, to accommodate the time-varying nature of the problem we modify Algorithm 2 slightly by imposing a constant stepsize rule, $\gamma_k = \gamma_0$ for some γ_0 small.

Table 3: Parameters for Example 3. ($C = 10$)

| Parameter | Class 1 | Class 2 |
|-----------------------|---------------------------------------|---------------------------------------|
| $\alpha_k(u_k)$ | $\bar{\alpha}_1(1 - u_1/\bar{u}_1)_+$ | $\bar{\alpha}_2(1 - u_2/\bar{u}_2)_+$ |
| $\bar{\alpha}_k$ | 10.0 | 10.0 |
| \bar{u}_k | 1.0 | 1.0 |
| \bar{u}_k | .9 | .9 |
| m_k | 1.0 | 1.0 |
| β_k | | |
| $t \in [0, 6)$ hrs. | .1 | 1.0 |
| $t \in [6, 12)$ hrs. | 1.0 | .1 |
| $t \in [12, 20)$ hrs. | .1 | 1.0 |

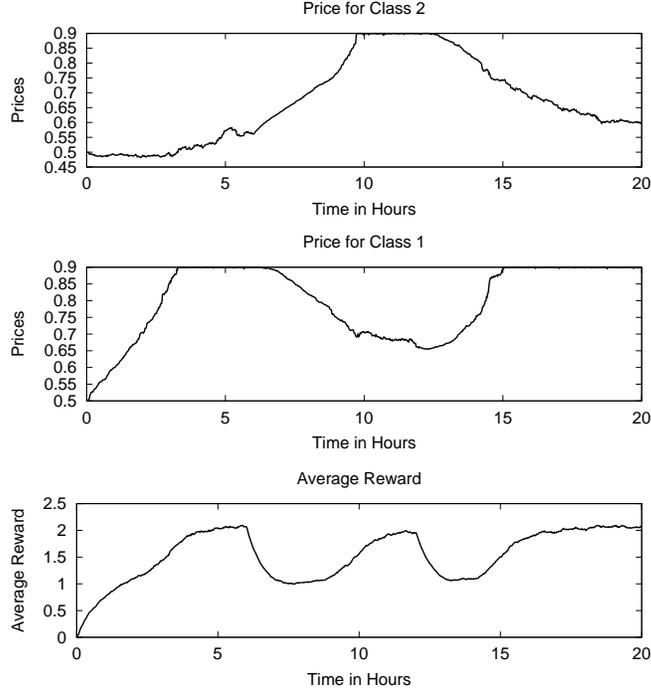


Figure 5: Evolution of prices and estimated average reward in Example 3.

The results of this example appear in Figs. 5 and 6. Fig. 5 shows the evolution of prices and estimated average reward over the entire 9×10^9 augmented state transitions. Note that the prices for classes 1 and 2 appear to alternate after the switch in β_1 and β_2 at $t = 6hrs.$ and $t = 12hrs.$. Thus, Algorithm 2 seems to be making the correct response to the instantaneous jump in the service rate parameters. (Note that the estimated average reward dips during the transient after each switch.) Fig. 6 shows the evolution of the estimates of β_1 and β_2 . Note that with the constant step size rule the estimates of β_1 and β_2 quickly track the instantaneous jumps in these parameters, although the estimates remain somewhat noisy throughout.

5.4 Example 4: Tracking Smooth Changes

In this example we explore the ability of Algorithm 2 to track smooth changes in model parameters. As in Examples 1 and 3, we consider a system with two service classes and limited capacity $C = 10 (Mbps)$. The parameters of the model are expressed in Table 4. As in Example 3, the service rate parameters β_1 and β_2

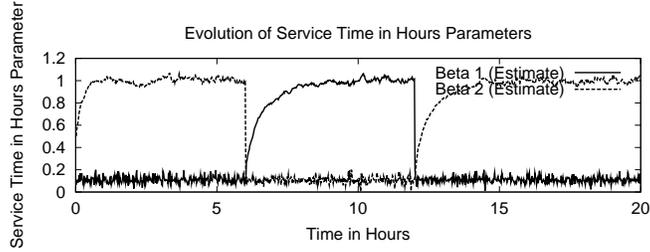


Figure 6: Evolution of the estimates of β_1 and β_2 in Example 3.

are given as functions of time. i.e. for the first third of the experiment $\beta_1 = .1$ and $\beta_2 = 1.0$, for the second third of the experiment β_1 increases linearly in time (on average) to $\beta_1 = 1.0$ and β_2 decreases linearly in time (on average) to $\beta_2 = .1$, and for the final third of the experiment β_1 and β_2 hold at 1.0 and $.1 (s^{-1})$, respectively. The demand characteristics for both service classes in this example are time-invariant and are the same as in Example 3.

Table 4: Parameters for Example 4. ($C = 10$).

| Parameter | Class 1 | Class 2 |
|-----------------------|---------------------------------------|---------------------------------------|
| $\alpha_k(u_k)$ | $\bar{\alpha}_1(1 - u_1/\bar{u}_1)_+$ | $\bar{\alpha}_2(1 - u_2/\bar{u}_2)_+$ |
| $\bar{\alpha}_k$ | 10.0 | 10.0 |
| \bar{u}_k | 1.0 | 1.0 |
| \bar{u}_k | .9 | .9 |
| m_k | 1.0 | 1.0 |
| β_k | | |
| $t \in [0, 6)hrs.$ | .1 | 1.0 |
| $t \in [6, 12)hrs.$ | (linear increase) | (linear decrease) |
| $t \in [12, 20) hrs.$ | 1.0 | .1 |

The results of this example appear in Figs. 7 and 8. Fig. 7 shows the evolution of prices and estimated average reward over the entire 9×10^9 augmented state transitions. Interestingly, from the plot of the average reward estimate, Algorithm 2 appears to be able to extract a higher average reward during the transient, especially where both service classes approach the same (relatively small) expected holding time $1/\beta_1 = 1/\beta_2 \approx 2 (s)$. Fig. 8 shows the evolution of the estimates of β_1 and β_2 . Again, with the constant step size rule, the estimates of β_1 and β_2 are able to track the smooth changes in these parameters between $t = 6hrs.$ and $t = 12hrs.$. As before, however, the constant stepsize rule gives rise to somewhat noisy estimates of the parameters.

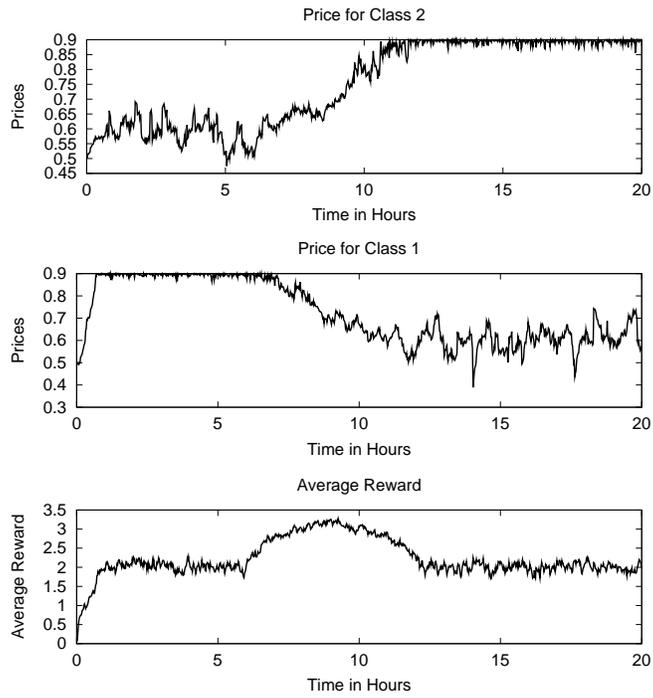


Figure 7: Evolution of prices and estimated average reward in Example 4.

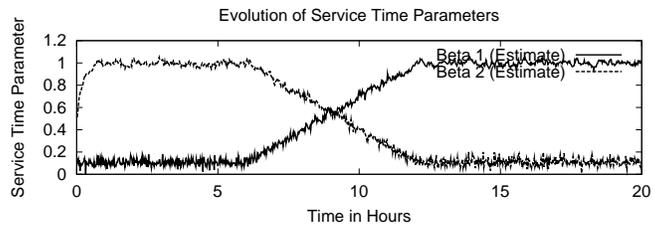


Figure 8: Evolution of the estimates of β_1 and β_2 for Example 4.

6 Conclusions

Seeking to develop efficient and robust pricing mechanisms for network services, we have adapted the simulation-based optimization procedure of [MT01, CNP02] for on-line use in continuous time. The most natural implementation of this scheme results in a model-based procedure (Algorithm 1) which is guaranteed to converge to a critical point assuming that all of the arrival and service rates of the underlying Markov process are known. Toward the goal of being robust to parametric uncertainty, we have also developed an adaptive procedure (Algorithm 2) which allows for the simultaneous estimation of model parameters. Both algorithms are efficient in the sense that they do not require that the entire history of the process be stored in order to make adjustments toward an optimal set of prices. In Algorithm 1 the memory requirements are $O(K)$, where K is the number of service classes, and in Algorithm 2 the memory requirements are $O(K + q)$, where q is the number of uncertain parameters in the model. Our numerical examples illustrate that our on-line algorithms are able to adjust prices toward optimal levels, even in the presence of modeling uncertainty, although the convergence appears to be quite slow.

While the results of this study are encouraging, there are many outstanding questions and issues to resolve. We first acknowledge that our numerical examples only provide a very rough proof-of-concept. It remains to be seen how our algorithms (particularly Algorithm 2) perform under more realistic traffic conditions. We have yet to experiment with general arrival and service time distributions, and we also have not experimented with the social welfare objective discussed in Section 2. Another primary objective for future work is to establish analytically the convergence properties of Algorithm 2. Our analysis, which is on-going, will proceed by first developing an adaptive version of the discrete-time results in [MT01, CNP02] and then by extending these results to the continuous-time implementation proposed in this paper.

References

- [Ber95] D. Bertsekas. *Dynamic Programming and Optimal Control*, volume II. Athena Scientific, 1995.
- [Bor91] V. S. Borkar. *Topics in Controlled Markov Chains*. Longman Scientific & Technical, 1991.
- [BS02] T. Basar and R. Srikant. Revenue-maximizing pricing and capacity expansion in a many-users regime. *Proceedings of IEEE INFOCOM*, 2002.

- [BV79] V. S. Borkar and P. Varaiya. Adaptive Control of Markov Chains, I: Finite Parameter Set. *IEEE Transactions on Automatic Control*, AC(24):953–957, 1979.
- [CDR01] C. Courcoubetis, A. Dimakis, and M. I. Reiman. Providing Bandwidth Guarantees over a Best-Effort Network: Call-Admission and Pricing. *Proceedings of INFOCOM*, 2001.
- [CKSW00] C. Courcoubetis, F. P. Kelly, V. A. Siris, and R. Weber. A Study of Simple Usage-based Charging Schemes for Broadband Networks. *Telecommunication Systems*, 15(3-4):323–343, 2000.
- [CKW00] C. Courcoubetis, F. P. Kelly, and R. Weber. Measurement-Based Usage Charges in Communications Networks. *Operations Research*, 48(4):535–548, July/August 2000.
- [CNP01] E. Campos-Nanez and S. D. Patek. On Improving the Performance of Simulation-Based Algorithms for Average Reward Processes with Application to Network Pricing. *Proceedings of the 2001 Winter Simulation Conference*, 2001.
- [CNP02] E. Campos-Nanez and S. D. Patek. Dynamically Identifying Regenerative Cycles in Simulation-Based Optimization Algorithms for Markov Chains. *Submitted to IEEE Transactions on Automatic Control*, 2002.
- [CO98] J. Crowcroft and P. Oechslin. Differentiated End-to-End Internet Services Using a Weighted Proportional Fair Sharing TCP. *CCR*, 28(3), July 1998.
- [CS97] C. Courcoubetis and V. A. Siris. An Approach to Pricing and Resource Sharing for Available Bit Rate (ABR) Services. Technical Report 212, Institute of Computer Science (ICS), 1997.
- [CSEZ93] R. Cocchi, S. Shenker, D. Estrin, and L. Zhang. Pricing in Computer Networks: Motivation, Formulation, and Example. *IEEE/ACM Transactions on Networking*, 1(6):614–627, 1993.
- [CSS96] C. Courcoubetis, V. A. Siris, and G. D. Stamoulis. Integration of Pricing and Flow Control for Available Bit Rate Services in ATM Networks. *Proceedings of IEEE Globecom*, 1996.
- [FD95] D. Ferrari and L. Delgrossi. Charging for QoS. *IEEE Transactions on Networking*, 1995.

- [FKSS96] W. Feng, D. D. Kandlur, D. Saha, and K. G. Shin. TCP Enhancements for an Integrated Services Internet. Technical Report RC 20618, I.B.M. T.J. Watson Research Center, 1996.
- [FKSS99] W. Feng, D. D. Kandlur, D. Saha, and K. G. Shin. Understanding TCP Dynamics in a Differentiated Services Internet. *IEEE/ACM Transactions on Networking*, 7(2):173–187, April 1999.
- [FYN88] D. Ferguson, Y. Yemini, and Ch. Nikolau. Microeconomic Algorithms for Load Balancing in Distributed Computer Systems. *Proceedings of DCS*, 1988.
- [GK99] R. J. Gibbens and F. P. Kelly. Resource pricing and the evolution of congestion control. *Automatica*, 35, 1999.
- [KA99] N. Keon and G. Anandalingam. Adaptive Flow Control for VoD Using Price Discounts. *Submitted*, 1999.
- [Kel97] F. Kelly. Charging and Rate Control for Elastic Traffic. *European Transactions on Telecommunications*, 8:33–37, 1997.
- [KM99] P. B. Key and Derek R. McAuley. Differential QoS and Pricing in Networks: where flow-control meets game theory. *IEE Proceedings Software*, March 1999.
- [LA00] R. J. La and V. Anantharam. Charge-Sensitive TCP and Rate Control in the Internet. *Proceedings of INFOCOM '2000*, 2000.
- [LS01] X. Lin and N. B. Shroff. Simplification of Network Dynamics in Large Systems. Technical report, Purdue University, 2001.
- [LV93] S. H. Low and P. P. Varaiya. A New Approach to Service Provisioning in ATM Networks. *IEEE/ACM Transactions on Networking*, 1(5):547–553, October 1993.
- [Mar01a] P. Marbach. Differentiated Services Networks: Pricing and Software Agents. *International Journal of Parallel and Distributed Systems and Networks; Special Issue on the Application of Intelligent Control and Optimization to Communications Systems and Networks*, 2001.
- [Mar01b] P. Marbach. Pricing Differentiated Services Networks: Bursty Traffic. *Proceedings of INFOCOM*, 2001.
- [Mar02] P. Marbach. The Role of Pricing in Differentiated Services Networks. *submitted*, 2002.

- [MT01] P. Marbach and J. N. Tsitsiklis. Simulation-Based Optimization of Markov Reward Processes. *ACM Transaction on Automatic Control*, 2001.
- [Od198] A. Odlyzko. The economics of the Internet: Utility, utilization, pricing, and Quality of Service. Technical report, AT&T Research Lab, 1998.
- [PL02] I. Ch. Paschalidis and Y. Liu. Pricing in Multiservice Loss Networks: Static Pricing, Asymptotic Optimality, and Demand Substitution Effects. *IEEE/ACM Transactions on Networking*, 10(3):425–438, 2002.
- [PS99] J. R. Perkins and R. Srikant. The Role of Queue Length Information in Congestion Control and Resource Pricing. *Proceedings of the 38th. Conference on Decision and Control*, 1999.
- [PT00] I. Ch. Paschalidis and J. N. Tsitsiklis. Congestion-Dependent Pricing of Network Services. *IEEE/ACM Transactions on Networking*, 8(2):171–184, 2000.
- [SKR00] R. Sinha, S. Kalyanaraman, and T. Ravichandran. Dynamic Capacity Contraction: A Framework for Congestion Sensitive Pricing. *Submitted to ICC*, 2000.
- [SLCL99] N. Semret, R. R. F. Liao, A. T. Campbell, and A. A. Lazar. Market Pricing of Differentiated Internet Services. Technical report, Columbia University, 1999.

The following arguments complement a proof included in [CNP02], which presents a similar mechanism for dynamic selection of i^* . To easily implement the mechanism in a continuous-time framework, a time threshold replaces a threshold based on number of steps in the discrete-time Markov chain. The only difference is that the number of steps is now random. We refer the reader to [CNP02] for the additional notation used in this section.

Let $\bar{\tau}_m$ be a time threshold for the duration of the cycle, and let τ_m be the random number of steps that can be observed by the uniformized chain during that same time interval. In other words, $\tau_m \sim \text{Poisson}(\bar{\tau}_m \nu^*)$. This threshold will be increased by one unit each time a cycle is abandoned, i.e. $\bar{\tau}_{m+1} = \bar{\tau}_m + \mathbf{1}_{\{i_{t_m} \neq i_m^*\}}$.

We have to redefine

$$t_{m+1} = \min\{t_m + \tau_m, \min\{n | n > t_m, i_n = i_m^*\}\}.$$

We can define $\varepsilon'_m = \mathbf{1}_{\{i_{t_m} \neq i_m^*\}} \gamma_m H(r_m)$, and show the following.

Proposition 1 *If $\gamma_m = \frac{a}{b+m}$ for $a, b > 0$ then*

$$\sum_m \|\varepsilon'_m\| \leq \infty.$$

To show this, we recall that, as presented in [CNP02], that if $\gamma_m = \frac{a}{b+m}$ for $a, b > 0$, and by defining $\{m_n\}_{n=0}^\infty$ to be the subsequence of abandoned cycles, then

$$\begin{aligned} E[\|\gamma_{m_n} H(r_{m_n})\| \mid t_{m_{n+1}} - t_{m_{n+1}-1} > \tau_{m_n}] \\ \leq \\ K \tau_{m_n} \frac{\rho^{\tau_{m_n}+1}}{(1-\rho^{\tau_{m_n}+1})^{n+b}} \left[\frac{\tau_{m_n}+1}{1-\rho} + \frac{1}{(1-\rho)^2} \right]. \end{aligned}$$

Here $K = -\log(1 - \bar{\rho})2aCC_2$. Using this result, we can show the following lemma.

Lemma 1

$$\begin{aligned} E[\|\varepsilon'_{m_n}\|] \leq \frac{K\rho}{(1-\bar{\rho})^b} e^{-\psi\bar{\tau}_n} \psi\bar{\tau}_n \\ \left[\frac{2}{1-\rho} + \frac{1}{(1-\rho)^2} + \frac{\psi\bar{\tau}_n}{1-\rho} \right], \end{aligned}$$

where $\psi = \nu^* \frac{1-(\rho+\bar{\rho})}{1-\bar{\rho}}$.

Proof: By conditioning on the number of events τ_n that can be observed during an interval of length $\bar{\tau}_n$, we can obtain

$$\begin{aligned} E[\|\varepsilon'_{m_n}\|] &= E[E[\|\gamma_{m_n} H(r_{m_n})\| \mid t_{m_n} - t_{m_n-1} > \tau_n]] \\ &= \sum_{t=0}^{\infty} P(\tau_n = t) E[\|\gamma_{m_n} H(r_{m_n})\| \mid t_{m_n} - t_{m_n} > t]. \end{aligned}$$

By the result in [CNP02], we obtain

$$\begin{aligned} E[\|\varepsilon'_{m_n}\|] &\leq \sum_t e^{-\nu^*\bar{\tau}_n} \frac{(\nu^*\bar{\tau}_n)^t}{t!} K \frac{\rho^{t+1}}{(1-\bar{\rho}^{t+1})^{t+b}} \\ &\quad \left[\frac{t^2+t}{1-\rho} + \frac{1}{(1-\rho)^2} \right] \\ &\leq \frac{K\rho e^{-\nu^*\bar{\tau}_n}}{(1-\bar{\rho})^b} \sum_t \left(\frac{\nu^*\bar{\tau}_n\rho}{1-\bar{\rho}} \right)^t \frac{1}{t!} \\ &\quad \left[\frac{t+t^2}{1-\rho} + \frac{t}{(1-\rho)^2} \right]. \end{aligned}$$

In the last inequality, the fact that $1 - \bar{\rho}^b > 1 - \bar{\rho}$ for all $b > 1$ was used. Finally, multiplying and dividing by $e^{-\psi\bar{\tau}_n}$ we obtain

$$E[\|\varepsilon'_{m_n}\|] \leq \frac{K\rho e^{-\psi\bar{\tau}_n}}{(1-\bar{\rho})^b} \sum_t e^{-\psi\bar{\tau}_n} (\psi\bar{\tau}_n)^t \frac{1}{t!} \left[\frac{t+t^2}{1-\rho} + \frac{t}{(1-\rho)^2} \right].$$

By using the first and second moments of a Poisson distribution with parameter $\psi * \bar{\tau}_n$, we obtain the desired result. ■

Using this result, we can show that the sum of bias introduced in this fashion is summable as long as $\rho + \bar{\rho} < 1$.

Proof of Proposition 1 Since $S_n = \sum_{k=0}^n \|\varepsilon'_{m_n}\|$ is a supermartingale, and

$$E[S_k] \leq \frac{K\rho}{(1-\bar{\rho})^b} \sum_{n=1}^{\infty} e^{-\psi\bar{\tau}_n} \left[\frac{2\psi\bar{\tau}_n}{1-\rho} + \frac{1\psi\bar{\tau}_n}{(1-\rho)^2} + \frac{(\psi\bar{\tau}_n)^2}{1-\rho} \right]$$

Notice that the right hand side of this expression is convergent almost surely if $\bar{\tau}_n = \bar{\tau}_0 + n$. Therefore, $\sum_n \|\varepsilon'_{m_n}\| < \infty$. ■