Telecommunication Systems 18:4, 315–329, 2001

State-Dependent Pricing and Its Economic Implications

QIONG WANG chiwang@lucent.com
Bell Laboratories, Lucent Technologies, Room 2B-308, 600 Mountain Avenue, Murray Hill, NJ 07974, USA

JON M. PEHA peha@ece.cmu.edu
Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Received March 1998; Revised October 1999

Abstract. In a packet-switched network, the service provider can charge users a state-dependent price, which depends on the extent to which the network is congested. Alternatively, one can charge users a long-term average price, which is set based on *expected* demand and capacity availability, but independent of *instantaneous* network conditions. In this paper, we compare the benefits of different pricing schemes from the service provider, society, and consumer perspectives. Our results suggest that adopting state-dependent pricing improves both profit and total benefits to the society, but may be detrimental to consumer benefits.

1. Introduction

Best-effort service comprises a dominant portion of all services offered by today's packet-switched networks. The service does not have to meet any specified quality of service, is interruptable, so it can be offered without capacity reservation. Because of the randomness of packet arrivals, congestion is bound to happen unless the carrier deploys excessive large amount of capacity. It has long been argued that the benefit from providing best-effort service can be enhanced if price is allowed to vary over very short time periods such as microseconds [3,6]. In periods when the network has more capacity than demand, the price should be lowered to encourage use, and in periods when the network is congested, the price should be raised to allocate limited capacity to the most valuable packets. The approach of varying price according to network congestion status is defined as *state-dependent pricing*. Several state-dependent pricing schemes have been proposed [1,2,5].

To accommodate the needs of emerging applications, guaranteed services are now being introduced to packet-switched networks. A guaranteed service is non-interruptable, meets specified quality of service, and thus requires some form of advance capacity reservation. We refer the packet-switched networks that offer both guaranteed and best-effort services as *integrated-services networks*. In an integrated-services network, packet arrival rates of guaranteed services fluctuate as calls come and go, and the network guarantees that these packets will be transmitted in a timely manner. Therefore, capacity left for best-effort service varies from time to time. In this situation, state-dependent prices are not only affected by variations in the arrival of best-effort packets, but also by fluctuations in the capacity available to carry those packets [9].

An alternative to state-dependent pricing is *long-term average pricing* under which price is set according to expected packet arrival rate and capacity availability. A long-term average price can be changed based on time-of-day to reflect anticipated variation, but does not vary with random fluctuations of network congestion status. At times when the network is congested, instead of raising packet price to reduce the number of packets to be admitted, the network drops packets independent of senders' willingness to pay.

To decide whether state-dependent pricing or long-term average pricing should be adopted, one needs to examine benefits of each approach from the service provider, consumer, and society perspectives. Benefits to the service provider are measured by *profit*. For consumers, a user can derive some positive utility if a packet is sent, and zero or some negative utility if a packet is dropped. The difference between these two utility values equals the user's willing to pay to get the packet sent, and the difference between the willingness to pay and the price paid is the net benefit that a user derives. Benefits to consumers as a whole can be measured by *consumer surplus*, which is the sum of net benefits of all users. Benefits to society can be measured by the sum of profit and consumer surplus, which is defined as *social welfare*. In this paper, we will use profit, consumer surplus, and social welfare as criteria to evaluate different pricing schemes.

There are many factors that can affect evaluations of different pricing schemes. For example, state-dependent pricing requires more complicated billing and accounting systems in order to charge users based on the network congestion status at the time of sending their packets. As a result, the service provider's profit of adopting state-dependent pricing is affected by the cost of building those complicated systems. Moreover, networks are usually equipped with some kind of congestion control mechanisms, some of which are more effective than others. Therefore, even if the packet arrival rate and capacity availability are exactly the same, the congestion status can be very different in networks that adopt different congestion control mechanisms. Consequently, the relative advantage of state-dependent pricing, which dynamically adjusts prices to follow the network congestion status, depends on what congestion control mechanisms are used in the network. In this paper, we will limit our analysis to the case in which neither billing and accounting costs nor congestion control mechanisms are considered. Those limitations favor state-dependent pricing in the evaluation of pricing mechanisms. If the comparison shows that even in this case, adopting state-dependent pricing does not improve profit, consumer surplus, or social welfare, then the mechanism does not deserve further consideration. Otherwise, more studies are needed to examine if the advantage is significant enough to justify the cost of building complicated billing and accounting systems, and whether state-dependent pricing can keep its superiority in cases where the network has an effective congestion control mechanism.

In comparing different pricing schemes, we assume that a carrier sets prices to maximize its profit without considering competition from other carriers' pricing strategies. We justify this assumption by the following reasons. First, the pricing problem considered in this paper is that of the short term: state-dependent prices are varied in seconds or even microseconds, profit is counted on hourly basis. Within such a short time interval, consumers are not likely to switch back and forth between competing carriers based on price differences. Second, in many places of US, there is only one carrier that controls the last mile connection to users. Furthermore, even in a competitive situation, the carrier can still set short-term prices to maximize profit from the use of network capacity, and leave competition for subscribers to longer-term marketing strategies, such as setting a competitive subscription fee, and developing promotional programs.

The major finding of the paper is that adopting state-dependent pricing always improves profit and social welfare, but can reduce consumer surplus in some situations. Therefore, there exist potential conflicts between a carrier and its users, and between maximizing social efficiency and protecting consumer interests in making network pricing decisions.

The paper is organized as follows: in section 2, we will explain why adopting state-dependent pricing can reduce consumer surplus based on a generic service model. In section 3, we will consider a model of integrated-services networks, and compare different pricing schemes by simulation. The simulation results prove by existence that consumer surplus can be reduced under a network pricing scheme that increases profit and social welfare. The results also show that this conflict is more likely to happen when demand is less elastic or fewer buffer spaces are allocated to best-effort service. The conclusion of the paper is given in section 4.

2. Pricing of a generic service facility

Consider a facility that can serve C customers in each period. There are only two types of periods: peak and off-peak, differentiated by the number of customers requesting the service. Denote p as price for the service, and in each period, only customers willing to pay at least p are eligible for service. Given p, denote the numbers of customers in peak and off-peak periods as $D_\mathrm{h}(p)$ and $D_\mathrm{l}(p)$, respectively. For every p, $D_\mathrm{h}(p) > D_\mathrm{l}(p)$. In any period, if the number of eligible customers exceeds capacity C, then only C of them can be served, and the rest are rejected.

Assume that in both peak and off-peak periods, demand elasticity is greater than 1, i.e.,

$$\left|\frac{\partial D_\mathrm{h}/D_\mathrm{h}}{\partial p/p}\right| > 1 \quad \text{and} \quad \left|\frac{\partial D_\mathrm{l}/D_\mathrm{l}}{\partial p/p}\right| > 1,$$

meaning that decreasing price by any percentage will lead to a larger percentage increase in number of customers. This implies that the service provider can always improve profit by reducing price and increasing volume, so the carrier maximizes profit by using all its capacity. Thus, with state-dependent pricing, the peak-period price P_h and off-peak period price P_l are set such that $D_\mathrm{h}(P_\mathrm{h}) = C$ and $D_\mathrm{l}(P_\mathrm{l}) = C$. Under long-term average pricing, the service provider can only set one price P_s for both peak and off-peak periods. Depending on $D_\mathrm{l}(p)$, as well as the frequency that a peak or off-peak period occurs, the

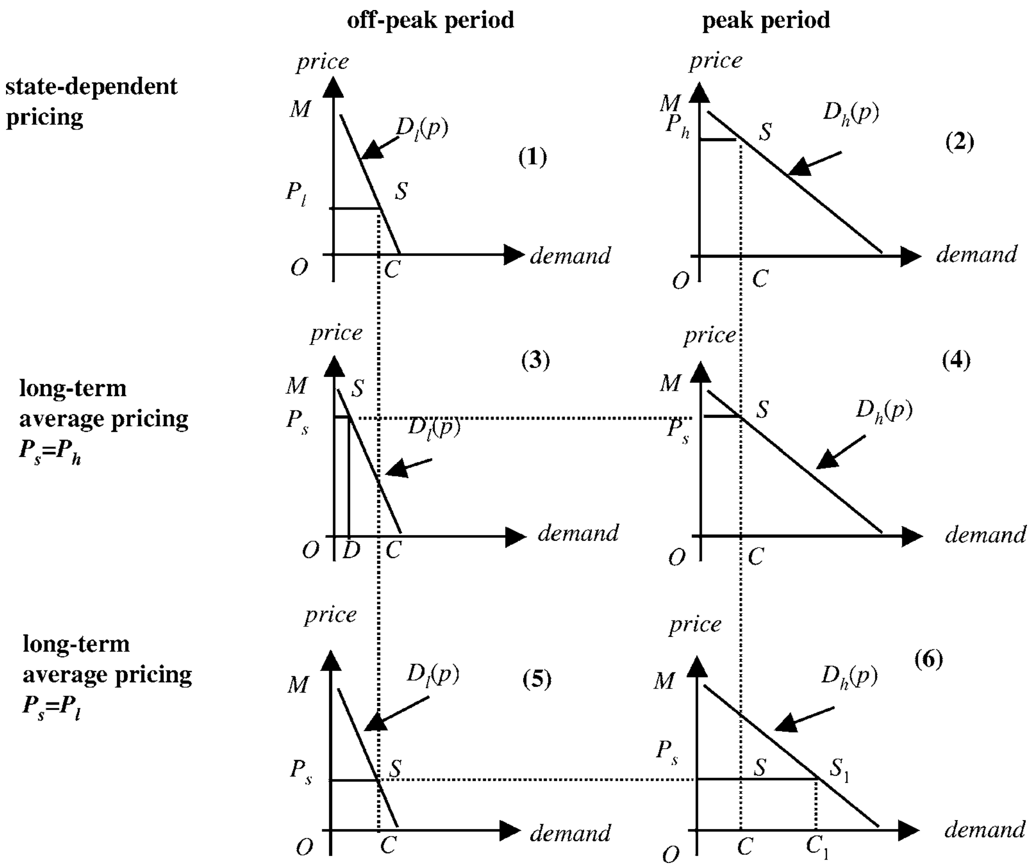

Figure 1. A comparison of long-term average pricing and state-dependent pricing.

value of P_s that maximizes profit ranges from P_l to P_h, and capacity will be fully utilized in peak periods. To make our argument, we only need to consider two special cases: (1) $P_s = P_l$ and (2) $P_s = P_h$, and compare state-dependent pricing and long-term average pricing in both situations.

Profit, social welfare, and consumer surplus realized under different pricing schemes can be represented by different areas in figure 1. In that figure, plots (1) and (2) correspond to state-dependent pricing during peak and off-peak periods. Plots (3) and (4) correspond to long-term average pricing when $P_s = P_h$, and plots (5) and (6) correspond to long-term average pricing when $P_s = P_l$. For example, with state-dependent pricing, profit obtained in an off-peak period equals price P_l times the number of customers served, C, which is the area of rectangle P_lSCO in plot 1, denoted as ΔP_lSCO. Similarly, in the same plot, consumer surplus is represented by ΔMSP_l, which is the area of triangle MSP_l, and social welfare is represented by $\Delta MSCO$.

Notice that in plot 6, which corresponds to a peak period when long-term average price $P_s = P_l$, the number of customers with willingness to pay no less than P_s is C_1. If there were no capacity constraint, all those customers would have been served, and the

Table 1
Comparison of profit, social welfare and consumer surplus based on a generic service model.

		Long-term average pricing ($P_s = P_h$)		State-dependent pricing		Long-term average pricing ($P_s = P_l$)
Profit	off-peak	$\Delta P_s SDO$ (plot 3)	$<^{\#}$	$\Delta P_l SCO$ (plot 1)	=	$\Delta P_s SCO$ (plot 5)
	peak	$\Delta P_s SCO$ (plot 4)	=	$\Delta P_h SCO$ (plot 2)	$>^{*}$	$\frac{\lvert OC\rvert}{\lvert OC_1\rvert} \cdot \Delta P_s S_1 C_1 O$ (plot 6)
Social welfare	off-peak	$\Delta MSDO$ (plot 3)	<	$\Delta MSCO$ (plot 1)	=	$\Delta MSCO$ (plot 5)
	peak	$\Delta MSCO$ (plot 4)	=	$\Delta MSCO$ (plot 2)	$>^{**}$	$\frac{\lvert OC\rvert}{\lvert OC_1\rvert} \cdot \Delta MS_1 C_1 O$ (plot 6)
Consumer surplus	off-peak	ΔMSP_s (plot 3)	<	ΔMSP_l (plot 1)	=	ΔMSP_s (plot 5)
	peak	ΔMSP_s (plot 4)	=	ΔMSP_h (plot 2)	$<^{\star}$	$\frac{\lvert OC\rvert}{\lvert OC_1\rvert} \cdot \Delta MS_1 P_s$ (plot 6)

led by the assumption that $\left|\frac{\partial D_l / D_l}{\partial p / p}\right| > 1$.

* $\frac{|OC|}{|OC_1|} \Delta P_s S_1 C_1 O = \frac{|OC|}{|OC_1|} \cdot |OC_1| \cdot |P_s O| < |OC| \cdot |P_h O| = \Delta P_h SCO.$

** $\frac{|OC|}{|OC_1|} \Delta MS_1 C_1 O = \frac{|OC|}{|OC_1|} \cdot |OC_1| \cdot \frac{1}{2}(|MO| + |S_1 C_1|) < \frac{1}{2}|OC|(|SC| + |MO|) = \Delta MSCO.$

⋆ $\frac{|OC|}{|OC_1|} \Delta MS_1 P_s = \frac{|OC|}{|OC_1|} \cdot \frac{1}{2}|OC_1| \cdot |MP_s| > \frac{1}{2}|OC| \cdot |MP_h| = \Delta MSP_h.$

resulting profit, social welfare, and consumer surplus would be $\Delta P_s S_1 C_1 O$, $\Delta MS_1 C_1 O$, and $\Delta MS_1 P_s$. However, given the capacity constraint, only C out of C_1 customers can be served, so those quantities need to be multiplied by a service ratio $|OC|/|OC_1|$, where $|OC|$ and $|OC_1|$ are the lengths of the straight lines OC and OC_1, respectively.

The comparison of profit, social welfare and consumer surplus are shown in table 1. Areas representing those quantities under state-dependent pricing are listed in the 5th column, and those under long-term average pricing are listed in the 3rd column (when $P_s = P_h$) and the 7th column (when $P_s = P_l$). Comparisons are made in the 4th and 6th columns. It is shown that if $P_s = P_h$, then adopting state-dependent pricing improves on all three criteria. If $P_s = P_l$, then adopting state-dependent pricing only improves profit and social welfare, and reduces consumer surplus.

This result can be explained from figure 1. Note in the first case ($P_s = P_h$), profit is higher under state-dependent pricing because the service provider has the flexibility to reduce the price during off-peak periods to serve more customers. The surpluses of those additional customers contribute to the increase of total consumer surplus. In the second case ($P_s = P_l$), adopting state-dependent pricing does not increase the number of customers served. The scheme only enables the carrier to raise price in peak-periods to reduce the number of customers eligible for the service, instead of randomly rejecting eligible customers. As limited capacity is used to serve more valuable customers, profit and social welfare rise. However, customers also pay more, which results in a decrease in consumer surplus.

To conclude, from a profit and social welfare perspective, state-dependent pricing is better than long-term average pricing because it enables the carrier (1) to increase the output, and/or (2) to select more valuable customers to serve when there is not enough capacity. While the first factor also helps to increase consumer surplus, the second one can reduce it. Consequently, for the sake of consumer surplus, state-dependent pricing may or may not be preferable, depending on which factor dominates. While the above conclusion is reached based on a generic service model, we will demonstrate that it also applies to a more concrete integrated-services network model in the next section.

3. Pricing of integrated-services networks

We show a model of integrated-services networks in section 3.1, and discuss different pricing schemes that apply to it in section 3.2. In section 3.3, we present the design of a simulation procedure that compares different pricing schemes, and the results are analyzed in section 3.4.

3.1. The network model

Assume one guaranteed service and one best-effort service are offered on a network link, of which the transmission capacity is C_{T} fixed-length packets per second. A price per packet is charged to best-effort service. Consumer willingness to pay for a packet equals the user's utility of sending the packet minus the disutility of not sending it. Let $\lambda_{\mathrm{b}}(p)$ be the arrival rate of best-effort packets, for which the willingness to pay is above a given price p.

Define a call of guaranteed service as a packet stream, the transmission of which meets a specified Quality of Service (QoS) objective. Assume call arrival is a random process with a mean of λ_{c}, and calls are first-come-first-served. Assume all calls have a common distribution of call duration, and let r_{c} be the expected value of the distribution. Denote λ_{g} as the packet arrival rate of each call.

Assume guaranteed service and best-effort service share both transmission capacity and buffer space. The QoS objective of guaranteed service is achieved by guaranteeing that the packet drop rate will be below a given threshold. Two mechanisms are employed to achieve this. First, the admission control algorithm is employed to prevent the number of guaranteed calls in progress from exceeding a pre-specified threshold M_{c}. Second, to limit the interference from packets of best-effort service, the Partial Buffer Sharing (PBS) mechanism [8] is adopted to give packets of guaranteed service the priority of using the buffer. While packets of guaranteed service are always admitted into the buffer as long as there are empty spaces, packets of best-effort service are rejected once the buffer occupancy reaches a given limit B. B is defined as the buffer threshold for best-effort service, which is less than the buffer size C_{b}. The value of B is set at the maximum level beyond which the packet drop rate of guaranteed service will exceed its guaranteed objective, assuming the maximum number of guaranteed calls (M_{c}) are in progress.

3.2. Pricing schemes

Given the above network model, we discuss how to set the long-term average and state-dependent prices to maximize profit.

3.2.1. Long-term average pricing

Let $\lambda_b(p)$ be the packet arrival rate of best-effort service, which is a function of price p. The expected profit can be written as:

$$\Phi(p) = \sum_{i=0}^{M_c} \pi_i \big[1 - d_i(\lambda_b)\big] p \lambda_b(p), \tag{1}$$

where i is the number of calls of guaranteed service in progress, which ranges from 0 to M_c, the maximum number of calls that can be accommodated. π_i is the probability that there are i calls of guaranteed service in progress, which is determined by the call arrival and departure processes. d_i is the packet drop rate of best-effort service, which can be determined by the packet arrival processes of both guaranteed and best-effort services, transmission capacity C_T, and the buffer threshold for the best-effort service B.

Under long-term average pricing, a single price p_l is set to maximize the profit function defined in equation (1).

3.2.2. State-dependent pricing

State-dependent pricing differs from long-term average pricing in that prices are set according to current network status instead of expected values. Notice that equation (1) can be rewritten as:

$$\Phi(p) = \sum_{i=0}^{M_c} \pi_i \Phi_i(p) = E\big[\Phi_i(p)\big], \tag{2}$$

where $\Phi_i(p) = [1 - d_i(\lambda_b)] p \lambda_b(p)$ is the profit from best-effort service given that there are i calls of guaranteed service in progress. Instead of choosing a single p_l to maximize $E[\Phi_i(p)]$, under state-dependent pricing, a vector of prices $(p_r^0, p_r^1, \ldots, p_r^{M_c})$ is specified, where each p_r^i maximizes an individual $\Phi_i(p)$. p_r^i is set as the current price if the current number of calls of guaranteed service is i. We define this kind of state-dependent pricing as response pricing as price only changes in response to a change in the number of calls in progress.

Under a more sophisticated state-dependent pricing scheme called "spot pricing", prices are set based on the current buffer occupancy level, as well as the number of guaranteed calls underway. It works as follows: Time is divided into fixed length segments with duration Δt. The spot price p_s^j is set at the beginning of the jth time segment. The objective is to maximize profit in segment j and subsequent segments where profit in equation (2) is written in the following recursive form:

$$\Phi_i^{n_j}\big(p_s^j\big) = \big[1 - d_i^{n_j}(\lambda_b)\big] p_s^j(i, n_j) \lambda_b\big(p_s^j\big) \Delta t + \sum_{n_{j+1}=0}^{B} P_{j+1}(n_{j+1} \mid p_s^j, n_j, i) \Phi_i^{n_{j+1}}\big(p_s^{j+1}\big), \tag{3}$$

where i is the number of calls of guaranteed service underway, and n_j is the buffer queue length at the beginning of the segment j. The resulting packet arrival and drop rates are denoted as $\lambda_b(p_s^j)$ and $d_i^{n_j}(\lambda_b)$, respectively. $[1 - d_i^{n_j}(\lambda_b)]p_s^j(i, n_j)\lambda_b(p_s^j)\Delta t$ can be viewed as the profit obtained within the segment j. $P_{j+1}(n_{j+1} \mid p_s^j, n_j, i)$ is the conditional probability that buffer occupancy is n_{j+1} at the beginning of segment $j+1$, given the buffer occupancy, spot price, and number of calls of guaranteed service in the previous segment. $\sum_{n_{j+1}=0}^{B} P_{j+1}(n_{j+1} \mid p_s^j, n_j, i)\Phi_i^{n_{j+1}}(p_s^{j+1})$ is the future profit from segment $j+1$ on.

Equation (3) is a standard form of dynamic programming, and the value of p_s^j that maximizes $\Phi_i^{n_j}(p_s^j)$ can be derived from a backward recursion algorithm [4].

3.3. *The design of the simulation*

We design a simulation to compare response pricing and spot pricing with long-term average pricing based on the network model formulated in section 3.1. For the purpose of running the simulation, we made additional specifications about the random processes and parameters involved. Those specifications are shown in table 2.

Special attention should be paid to the second row, which shows that packet arrival rate is a function of price, which is set through either long-term average or state-dependent pricing schemes. The function specifies the demand for best-effort service, and the parameters of the function are interpreted as follows: $\lambda_{\max}$ is the maximum packet arrival rate (arrival rate when the price is 0), $p_{\max}$ is the maximum willingness to pay per packet of all consumers, and α is a parameter that specifies both the strength of demand and demand elasticity. A smaller α implies a smaller arrival rate at a given price per packet. Moreover, when α is small, a slight increase in price will result in a large decrease in the packet arrival rate.

Based on those inputs, we determine long-term average prices, response prices and spot prices to maximize the profit functions defined in equations (1), (2), and (3), respectively. As we assume the arrival of guaranteed calls is Poisson and call duration is

Table 2
Specifications of processes and parameters as simulation inputs.

Processes	Specifications
Packet arrival of best-effort service	Poisson with arrival rate λ_b, a function of price p $\lambda_b(p) = \lambda_{\max}\left[1 - \left(\frac{p}{p_{\max}}\right)^{\alpha}\right]$, where $\lambda_{\max} = 10000$ (packets/s) $p_{\max} = 5 \cdot 10^{-7}$ (\$/packets) and $\alpha = 0.5$
Call arrival of guaranteed service	Poisson with arrival rate $\lambda_g = 0.2$ (calls/s)
Call duration distribution	Exponential with mean $r_c = 75$ (s)
Packet arrival per call of guaranteed service	Poisson with arrival rate $\lambda_c = 100$ (packets/s/call)
Parameters	**Specifications**
Total capacity	$C_T = 3640$ packets/s
Buffer threshold for best-effort service	$B = 20$
Maximum number of calls of guaranteed service	$M_c = 18$

exponentially distributed, π_i, the probability that there are i guaranteed calls in progress, is determined by

$$\pi_i = \frac{(\lambda_c r_c)^i / i!}{\sum_{m=0}^{M_c} (\lambda_c r_c)^m / m!}.$$

Long-term average and response pricing does not consider instantaneous changes of buffer occupancy. Therefore, in deciding these two prices d_i the packet drop rate of best-effort service when there are i guaranteed calls in process, is derived from the steady-state analysis of the packet queue in the buffer. In deciding spot prices, we apply transient analysis to determine d_i. In both analyses, the packet arrival rate is $\lambda_b + i\lambda_g$ if queue length is less than B, and $i\lambda_g$ otherwise. d_i is the sum of probabilities that queue length is larger than or equal to B.

We then run a simulation program to compare profit, consumer surplus, and social welfare under different pricing schemes. Each simulation is run multiple times, using different seeds to generate random numbers. Results from those runs are considered as independent samples of output variables. With probability 95%, the mean values of profit and social welfare are within $\pm 3\%$ of the values shown, and the mean consumer surplus is within $\pm 8\%$ of the values shown. A T-test is used to determine whether the observed differences in profit, social welfare, and consumer surplus from one pricing scheme to another are statistically significant.

3.4. *Discussion of simulation results*

We first simulate a base case using the inputs specified in table 2, and then study the effects of varying the demand parameter α, and the buffer threshold for best-effort service, B.

3.4.1. Base case

In this case, the long-term average price is \$3.028 $\cdot 10^{-7}$ per packet, and the response price as a function of the number of guaranteed calls in progress is shown in figure 2. In figure 3, we show sample values of spot prices as a function of buffer occupancy when the number of guaranteed calls in progress is 0, 9 and 18. The figures show that when the network is less congested, i.e., when the number of guaranteed calls is small and/or buffer occupancy is low, both response price and spot price stay constant because they are based only on demand and not constrained by capacity availability. As the network gets more congested, those prices increase, which means only packets with higher willingness to pay will be transmitted.

Distributions of buffer occupancy under different pricing schemes are shown in figure 4. Notice that zero buffer occupancy means that capacity lays idle and full buffer occupancy means that packets have to be dropped. Therefore, it is desirable to keep the buffer partially full. Based on this criteria, the distribution of buffer occupancy is best with spot pricing. This is because the price is a direct function of buffer occupancy under spot pricing, so packet arrival rate can be increased when the buffer is empty, and

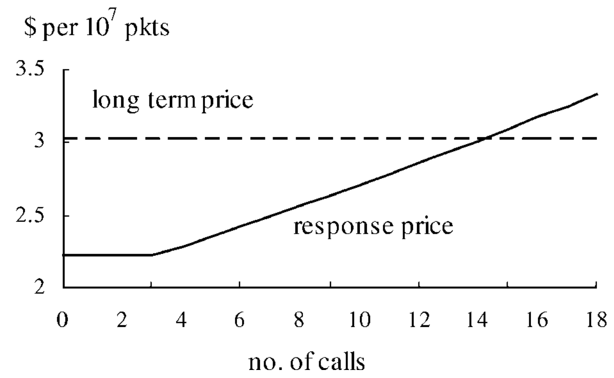

Figure 2. Long-term average and response prices.

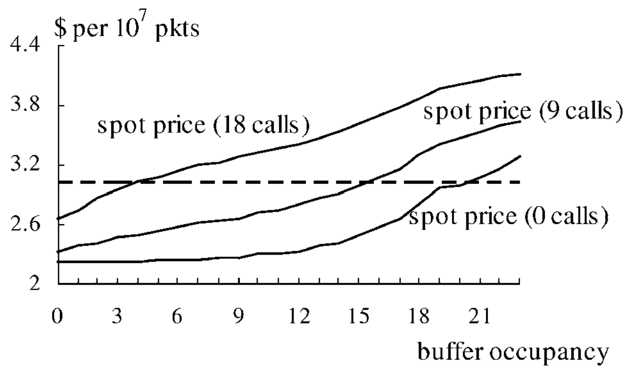

Figure 3. Sample of spot prices.

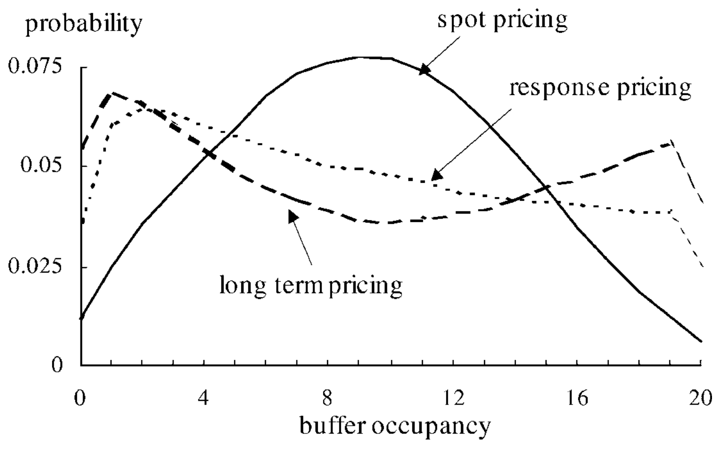

Figure 4. Distributions of buffer occupancy under different pricing schemes.

Table 3
Mean profit, social welfare, and consumer surplus ($/60 min).

	Long-term average pricing	Response pricing	Spot pricing
Profit	2.28	2.33	2.37
Social welfare	2.99	3.06	3.15
Consumer surplus	0.71	0.73	0.78

decreased when the buffer is full. The distribution of buffer occupancy is also better with response pricing than that with long-term average pricing because the former has the flexibility of varying prices while the latter does not.

The fact that the buffer has a higher probability of being partially full under spot pricing leads to superior throughput under that scheme. The simulation shows that in a 60-minute segment, mean throughput is $2.22k$ packets/s under spot pricing, $2.14k$ packets/s under response pricing, and $2.08k$ packets/s under long-term average pricing (accuracy $\pm 4\%$ with probability 95%). T-tests show with 99.5% statistical significance that throughput is better with response pricing than that with long-term average pricing, and better with spot pricing than that with the other two.

The comparison of profit, social welfare, and consumer surplus in a 60-minute segment is shown in table 3. Based on T-tests with 99.5% significance level, we conclude that these values are greatest with spot pricing and smallest under long-term average pricing.

3.4.2. Other cases

We will now vary input parameters to test the generality of the results obtained in the base case. This section will show that profit and social welfare are consistently higher with state-dependent pricing, but consumer surplus can be lower in some cases. The decrease in consumer surplus can be explained by the insights discussed in section 2.

Figures 5–7 show profit, social welfare, and consumer surplus with different pricing schemes as functions of the demand parameter α. As α increases, profit under each pricing scheme increases because at any given price, packet arrival rate $\lambda_b(p)$ increases. Moreover, profit is always greatest with spot pricing and smallest with long-term average pricing, which is consistent with the base case scenario. Figure 6 shows that the same is true for social welfare. However, the trend for consumer surplus is different, as shown in figure 7. First, instead of increasing, consumer surplus falls with α because a larger α means stronger but less elastic demand, so the carrier can charge a higher price and extract more wealth from consumers. Furthermore, the more complicated the pricing scheme, the faster the rate of decrease. As a result, consumer surplus under state-dependent pricing is higher when α is small and lower when α is large. For example, in comparison with long-term average pricing, consumer surplus is 9.1% higher under spot pricing and 2.6% higher under response pricing when $\alpha = 0.5$ (both differences are at 99.5% significance level). However, when $\alpha = 1.0$, consumer surplus is 5.1% lower under spot pricing than long-term average pricing, and the difference between response pricing and long-term average pricing is not statistically significant.

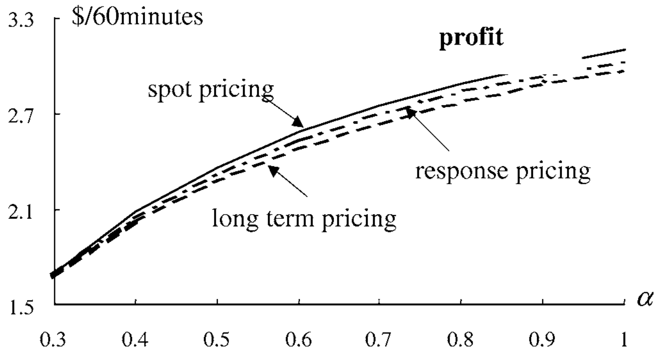

Figure 5. Changes of profit with α.

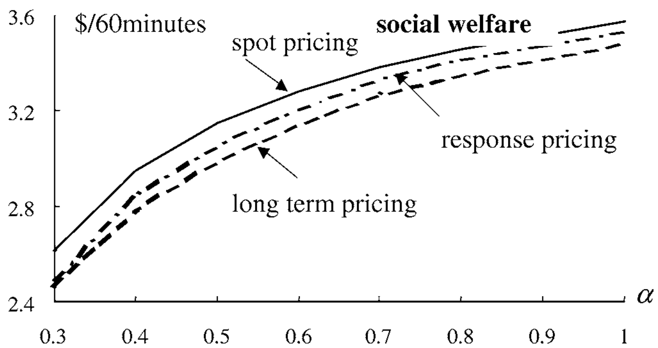

Figure 6. Changes of social welfare with α.

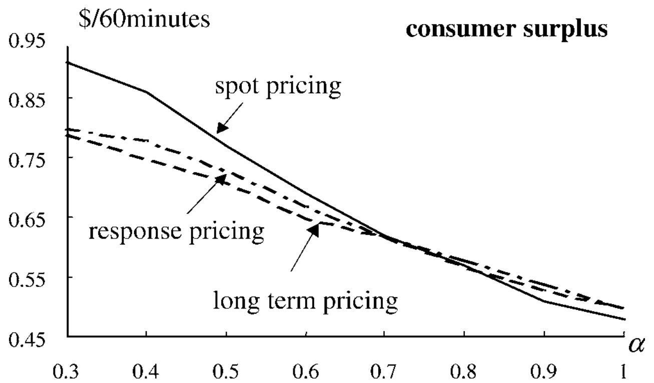

Figure 7. Changes of consumer surplus with α.

The fact that consumer surplus decreases more quickly as α increases is consistent with results derived from the generic service model in section 2. We have concluded that state-dependent pricing increases consumer surplus by improving throughput, and decreases consumer surplus by allowing the carrier to charge higher prices during congested periods. With small α, demand is more elastic, so a slight change in price can

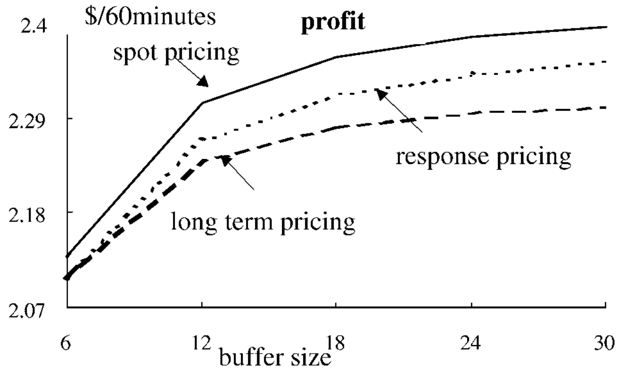

Figure 8. Impact of buffer threshold B on profit.

cause large variations in packet arrivals. In this situation, packet arrival rate can be effectively controlled to improve throughput with relatively small change in price. Consequently, consumer surplus grows with the adoption of state-dependent pricing. With large α, demand is less elastic, i.e., the packet arrival rate changes less drastically in response to changes in price. Therefore, varying price to improve throughput leads to much higher prices during periods of congestion, and the impact of these high prices on consumers is greater than the impact of throughput improvement.

Figures 8–10 show profit, social welfare, and consumer surplus under each scheme as a function of B, the buffer threshold for best-effort service. With smaller values of B, profit and social welfare are higher with state-dependent pricing than those with long-term average pricing, but there is no difference in consumer surplus between response pricing and long-term average pricing. Consumer surplus with spot pricing is significantly lower than with either response pricing or long-term average pricing. This phenomenon can also be explained by insights we derived from the generic service model: with smaller buffer spaces allocated to packets of best-effort service, the network becomes congested more frequently. As a result, state-dependent pricing is more of a scheme of raising price during congestion than a scheme of improving throughput by reducing price during noncongested periods. As we pointed out in section 2, the benefits of adopting state-dependent pricing in this case will mainly be captured by the service provider, so the scheme does not benefit consumers.

4. Conclusions on policy implications

In this paper, we discuss implementing state-dependent pricing in integrated-services networks, and compare two such schemes, spot pricing and response pricing, with long-term average pricing. We show that state-dependent pricing can achieve a higher profit and social welfare than long-term average pricing. This is largely due to two effects:

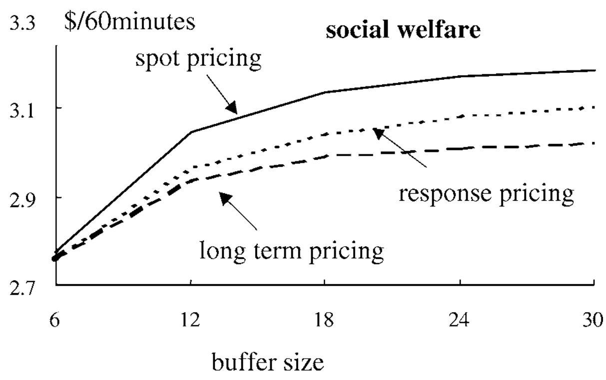

Figure 9. Impact of buffer threshold (B) on social welfare.

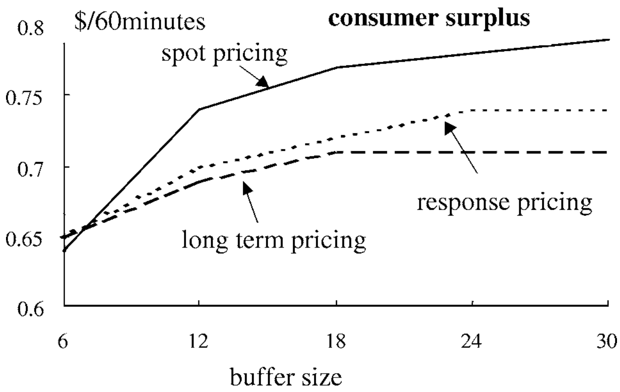

Figure 10. Impact of buffer threshold (B) on consumer surplus.

(1) state-dependent pricing serves as a traffic management mechanism that results in a higher throughput of packets; and

(2) state-dependent pricing allows more valuable traffic to be carried during congested periods.

While improving throughput benefits both carriers and consumers, raising price during congested periods enables the service provider to extract more wealth from consumers. Therefore, in some cases, while profit and social welfare will be enhanced by adopting state-dependent pricing, consumer surplus will be reduced.

This analysis is carried out without considering the impact of congestion control mechanisms. As demonstrated in [7], some congestion control mechanisms are more efficient than state-dependent pricing in maximizing throughput. Therefore, in networks equipped with those congestion control mechanisms, the ability of state-dependent pricing to improve throughput and social welfare would be diminished. This makes it more likely that the cost to consumers of state-dependent pricing will outweigh the benefits.

Acknowledgement

The authors gratefully acknowledge Brendan Dunham for his hard work in coding simulation programs, and Professor Marvin Sirbu for his helpful comments. Responsibilities for errors are solely retained by the authors.

References

[1] K. Danielsen and M. Weiss, User control modes and IP allocation, in: *Internet Economics*, eds. L. McKnight and J. Bailey (MIT Press, Cambridge, MA, 1997) pp. 305–322.

[2] A. Gupta, D.O. Stahl and A.B. Whinston, Priority pricing of integrated services networks, in: *Internet Economics*, eds. L. McKnight and J. Bailey (MIT Press, Cambridge, MA, 1997) pp. 323–352.

[3] S. Jordan and H. Jiang, Connection establishment in high-speed networks, IEEE Journal on Selected Areas in Communications 13(7) (1995) 1150–1161.

[4] R.E. Larson and J.L. Casti, *Principles of Dynamic Programming* (Marcel Dekker, New York, 1978).

[5] J.F. Mackie-Mason, L. Murphy and J. Murphy, Responsive pricing in the Internet, in: *Internet Economics*, eds. L. McKnight and J. Bailey (MIT Press, Cambridge, MA, 1997) pp. 279–304.

[6] J.F. Mackie-Mason and H. Varian, Pricing congested network resources, IEEE Journal on Selected Areas in Communications 13(7) (1995) 1141–1149.

[7] J.M. Peha, Dynamic pricing as congestion control in ATM networks, in: *Proc. of IEEE Globecom-97*, Phoenix, AZ, 1997, pp. 1367–1372.

[8] S. Ramesh, C. Rosenberg and C. Kumar, Revenue maximization in ATM networks using the CLP capability and buffer priority management, IEEE/ACM Transactions on Networking 4(6) (1996) 941–950.

[9] Q. Wang, J.M. Peha and M.A. Sirbu, Optimal pricing for integrated-services networks, in: *Internet Economics*, eds. L. McKnight and J. Bailey (MIT Press, Cambridge, MA, 1997) pp. 353–376.