Effect of Pricing Intervals on the Congestion-Sensitivity of Network Service Prices

Murat Yuksel, Shivkumar Kalyanaraman, Biplab Sikdar
Rensselaer Polytechnic Institute, Troy, NY
yuksem@cs.rpi.edu, shivkuma@ecse.rpi.edu, sikdab@rpi.edu

***Abstract*— Several proposals have been made for congestion-sensitive pricing of the Internet. One key implementation obstacle for these dynamic pricing schemes is the necessity of frequent price updates whereas the structure of wide area networks does not allow frequent price updates for many reasons, such as round-trip-times are very large for some cases. As the networks allow infrequent price updates, more control is achieved by the pricing schemes with more frequent price updates. So an important issue to investigate is to find a maximum value for the interval (i.e. *pricing interval*) over which price updates occur, such that the level of congestion control can remain in a desired range. This paper presents our modeling and analysis work for the length of pricing intervals. To represent the level of control over congestion, we use correlation between prices and congestion measures. After developing approximate models for the correlation, we find and prove that the correlation degrades at most inversely proportional to an increase in the pricing interval. We also find that the correlation degrades with an increase in mean or variance of the incoming traffic.**

***Index Terms*—Network Pricing, Congestion Pricing, Quality-of-Service, Congestion Control**

I. Introduction

One proposed method for controlling congestion in wide area networks is to apply *congestion-sensitive pricing* [1], [2]. Many proposals have been made to implement dynamic pricing over wide area networks and the Internet [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14]. Most of these schemes aimed to employ congestion pricing. The main idea of congestion-sensitive pricing is to update price of the network service dynamically over time such that it increases during congestion epochs and causes users to reduce their demand. So, implementation of congestion-sensitive pricing protocols makes it necessary to change the price after some time interval, what we call *pricing interval*.

Clark's Expected Capacity [3] scheme proposes long-term contracts as the pricing intervals. Kelly's Proportional Fair Pricing (also called as packet marking) scheme [5] proposes shadow prices to be fed back from network routers which has to happen over some time interval. MacKie-Mason and Varian's Smart Market scheme [6] proposes price updates at interior routers which cannot happen continuously and have to happen over some time interval. Odlyzko's Paris Metro Pricing scheme [8] proposes fixed prices for different sub-classes of network service, but congestion-sensitivity of the prices can only be achieved by updating them over some time interval. Wang and Schulzrinne's Resource Negotiation and Pricing (RNAP) [9] framework proposes to locally update the prices at each router which has to happen over some time interval.

This work is sponsored by NSF under contract number ANI9819112 and co-sponsored by Intel Corporation.

There are numerous implementation problems for dynamic or congestion pricing schemes, which can be traced into pricing intervals. We can list some of the important ones as follows:

- *Users do not like price fluctuations:* Currently, most ISPs employ flat-rate pricing which makes individual users happy. Naturally, most users do not want to have a network service with a price changing dynamically. In [15], Edell and Varaiya proved that there is a certain level of desire for quality-of-service. However, in [16] and [17], Odlyzko provides evidence that most users want simple pricing plans and they easily get irritated by complex pricing plans with frequent price changes. In other words, users like a service with *larger* pricing intervals.
- *Control of congestion degrades with larger pricing intervals:* Congestion level of the network changes dynamically over time. So, the more frequent the price is updated, the better the congestion control. From the provider's side, it is easier to achieve better congestion control with *smaller* pricing intervals.
- *Users want prior pricing:* It is also desired by the users that the service price must be communicated to them before it is charged. This makes it necessary to inform the users before applying any price update. So, the provider has to handle the overhead of that price communication. The important thing is to keep this overhead as less as possible, which can be done with *larger* pricing intervals.

Hence, length of pricing intervals is a key issue for the implementation of congestion pricing protocols. In this particular work, we focus on modeling and analysis of pricing intervals to come up with a maximum value for it such that the level of congestion control remains in an acceptable range. Beyond this range, pricing could be used to regulate demand, but it becomes less useful as a tool for congestion management. The rest of the paper is organized as follows: In Section II, we first explore steady-state dynamics of congestion-sensitive pricing with a detailed look at the behavior of prices and congestion relative to each other. We then develop and discuss approximate analytical models for the correlation of prices and congestion measures in Section III. In Section IV, we validate the models by simulation experiments and present results. Finally, in Section V we discuss implications of the work and possible future directions.

II. Dynamics of Congestion-Sensitive Pricing

This section investigates behavior of congestion prices and congestion measures relative to each other in a steady-state system. A sample scenario is described in Figure 1. The provider employs a pricing interval of T to implement congestion-sensitive pricing for its service. The customer uses that service

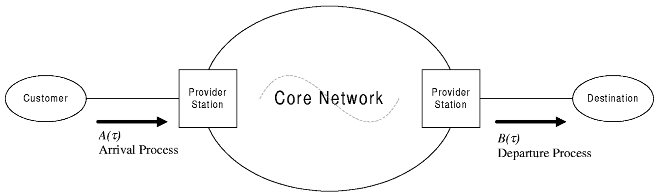


Fig. 1. A sample customer-provider network.

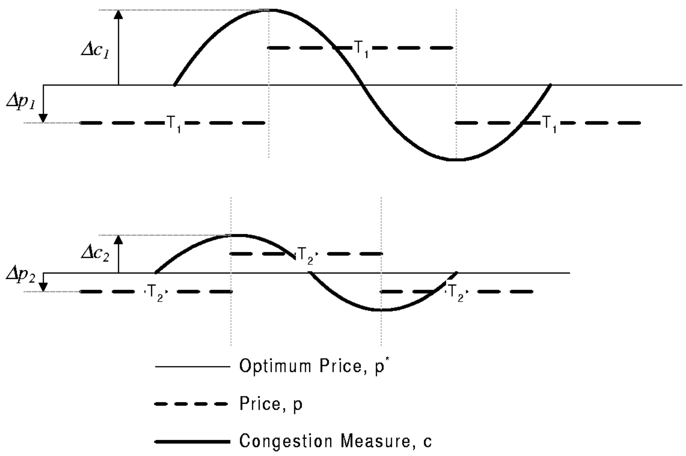


Fig. 2. Congestion measures relative to prices in a steady-state network.

to send traffic to the destination through the provider's network. The provider observes the congestion level, c, in the network core and adjusts price, p, according to it. Note that c and p are indeed functions of time (i.e. $c(t)$ and $p(t)$), but we use c and p throughout the paper for simplicity of notation. The provider can observe the network core over small time intervals, i.e. a few round-trip-times (RTTs). To understand effect of pricing interval to the dynamics of congestion-sensitive pricing, we look at the relationship between c and p over time.

Assuming that we have continuous knowledge of congestion level, c, we can represent the dynamics of congestion-sensitive pricing as in Figure 2. Figure 2 represents the steady-state relationship between c and p for two different pricing interval lengths, $T_1 > T_2$. The price, p, varies around an optimum value, p^*. The important issue to realize is that congestion control becomes better if the similarity between the price and congestion level is higher. Because of the implementation constraints explained in previouse section, the price cannot be updated continuously. This results in dissimilarity between the price and congestion level. Intuitively, if the correlation between the prices and the congestion measures is higher, fidelity of control over congestion becomes higher. Again by intuition, the correlation becomes smaller if the pricing interval is larger.

Another important issue is the *price oscillation* caused by the discontinuous price updates. As the pricing intervals get larger, the oscillation in price also gets larger. This, in effect, leads to oscillation in user demand (i.e. traffic). So, larger oscillations in price are expected to cause larger oscillation and *higher variance* in incoming traffic. Then, more oscillated traffic causes more oscillated congestion level. This behavior is represented in Figure 2 with the case that $\triangle c_1 > \triangle c_2$ and $\triangle p_1 > \triangle p_2$.

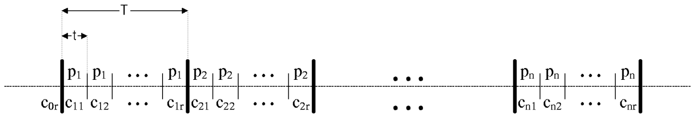


Fig. 3. Prices and congestion measures for subsequent observation intervals.

III. Analytical Model for Correlation of Prices and Congestion Measures

A. Assumptions and Model Development

Assume the length of pricing interval stays fixed at T over n intervals. Also assume the provider can observe the congestion level at a smaller time scale with fixed observation intervals, t. Assume that $T = rt$ holds, where r is the number of observations the provider makes in a single pricing interval. Assume that the queue backlog in the network core is an exact measure of congestion. [18]

We assume that the customer has a fixed budget for network service and he/she sends traffic according to a counting process, which is a continuous time stationary stochastic process $A(\tau)$,$\tau \geq 0$ with first and second moments of λ_1 and λ_2 respectively. In reality, λ_1 is not fixed, because the customer responds to price changes by changing its λ_1. However, since we assume steady-state and fixed budget for the customer, it is reasonable to say that the customer will send at a constant rate over a large number of pricing intervals. Let m_{ij} be the number of packet arrivals from the customer during the j th observation interval of ith pricing interval, where $i = 1..n$ and $j = 1..r$. So the total number of packet arrivals during the ith pricing interval is $m_i = \sum_{s=1}^{r} m_{is}$. Also assume that the packets leave after the network service according to a counting process, which is a continuous time stationary stochastic process $B(\tau)$,$\tau \geq 0$ with first and second moments of μ_1 and μ_2 respectively. Let k_{ij} be the number of packet departures during the jth observation interval of ith pricing interval, where $i = 1..n$ and $j = 1..r$. So the total number of packet departures during the ith pricing interval is $k_i = \sum_{s=1}^{r} k_{is}$. Assuming that no drop happens in the network core, the first moments of the two processes are equal in steady-state, i.e. $\lambda_1 = \mu_1$, but the second moments are not.

As represented in Figure 3, let p_i be the advertised price and c_{ij} is the congestion measure (queue backlog) at the end of the jth observation in the ith pricing interval. In our model we need a generic way of representing the relationship between prices and congestion. We assumed that the congestion-sensitive pricing algorithm calculates the price for the ith pricing interval according to the formula[1]

$$p_i = a(t,r) \ \ c_{(i-1)r} \tag{1}$$

where $a(t,r)$, *pricing factor*, is a function of pricing interval and observation interval defined by the congestion-sensitive pricing algorithm itself. In our modeling, we assume that $a(t,r)$ is only effected by the interval lengths, not by the congestion measures. Notice that this assumption does not rule out the effect of congestion measures on the price, rather it splits the effect of congestion measures and interval lengths to the price. We will use a instead of $a(t,r)$ for the rest of the paper just for

[1] Note that this is a simplifying formula for tractability, and does not fully express all aspects of congestion pricing.

notation simplicity. Within this context, the following equations hold:

$$c_{ij} = c_{0r} + \sum_{u=1}^{i-1}(m_u - k_u) + \sum_{s=1}^{j}(m_{is} - k_{is}) \tag{2}$$

$$c_{ir} = c_{0r} + \sum_{j=1}^{i}(m_j - k_j) \tag{3}$$

where $i \geq 1$. Reasoning behind Equations 2 and 3 is that the queue backlog (which is the congestion measure) at the end of an interval is equal to the number of packet arrivals minus the number of packet departures during that interval.

Let the average price be $\overline{p}$ and the average queue backlog be $\overline{c}$. By assuming that the system is in steady-state we can conclude that the following equation is satisfied

$$\overline{p} = a\overline{c} \tag{4}$$

Since the system is assumed to be in steady-state, we can assume the initial (right before the first pricing interval) congestion measure equals to the average queue backlog, i.e.

$$c_{0r} = \overline{c} \tag{5}$$

We want to approximate the model of correlation between p and c according to the above assumptions. We can write the formula for correlation between p and c over n pricing intervals as

$$Corr_n = \frac{E_n[(c-\overline{c})(p-\overline{p})|m,k]}{E_n[(c-\overline{c})^2|m,k]E_n[(p-\overline{p})^2|m,k]} \tag{6}$$

assuming that total of m packet arrivals and k packet departures happen during the n rounds. We can calculate the numerator term in Equation 6 as follows:

$$E_n[(c-\overline{c})(p-\overline{p})|m,k] = \frac{1}{rn}\sum_{i=1}^{n}\sum_{j=1}^{r}(p_i - \overline{p})(c_{ij} - \overline{c}) \tag{7}$$

By applying Equations 1, 4 and 5 into Equation 7 we can get

$$E_n[(c-\overline{c})(p-\overline{p})|m,k] = \sum_{i=1}^{n}\sum_{j=1}^{r}\frac{a(c_{(i-1)r} - c_{0r})(c_{ij} - c_{0r})}{rn} \tag{8}$$

Then by applying Equations 2 and 3 into Equation 8, we derive the following

$$E_n[(c-\overline{c})(p-\overline{p})|m,k] = \frac{a}{rn}\sum_{i=1}^{n}\sum_{j=1}^{r} \left[H_1 + \sum_{\theta=1}^{i-1}(m_\theta - k_\theta)\sum_{s=1}^{j}(m_{is} - k_{is})\right] \tag{9}$$

where $H_1 = \sum_u (m_u - k_u)^2 + \sum_u \sum_{v \neq u} 2(m_u - k_u)(m_v - k_v)$, $u = 1..i-1$ and $v = 1..i-1$.

Similarly, we calculate the terms in the denominator of Equation 6 and get the followings:

$$E_n[(p-\overline{p})^2|m,k] = \frac{a^2}{n}\sum_{i=2}^{n} H_1 \tag{10}$$

$$E_n[(c-\overline{c})^2|m,k] = \frac{1}{rn}\sum_{i=1}^{n}\sum_{j=1}^{r} \left[H_1 + H_2 + 2\sum_{u=1}^{i-1}(m_u - k_u)\sum_{s=1}^{j}(m_{is} - k_{is})\right] \tag{11}$$

where $H_2 = \sum_s (m_{is} - k_{is})^2 + \sum_s \sum_{z \neq s} 2(m_{is} - k_{is})(m_{iz} - k_{iz})$, $s = 1..j$, $z = 1..j$.

Now we can relax the condition on m and k by summing out conditional probabilities on Equations 9, 11, and 10. Specifically, we need to apply the operation

$$E_n[x] = \sum_{m_{ij}=0}^{\infty}\sum_{k_{ij}=0}^{\infty} E_n[x|m,k]P_{m_{ij};k_{ij}} \tag{12}$$

for all $i = 1..n$ and $j = 1..r$, where $P_{m_{ij};k_{ij}}$ is $P\{A(t) = m_{ij}; B(t) = k_{ij}\}$. This operation is non-trivial because of the dependency between the processes $A(\tau)$ and $B(\tau)$, and it is not possible to reach a closed-form solution without simplifying assumptions. After this point, we develop two *approximate* models by making simplifying assumptions.

1) Model-I: Although the arrival and departure processes are correlated, there might also be cases where the correlation is negligible. For example, if the distance between arrival and departure points is more, then the lag between the arrival and departure processes also becomes more which lowers the correlation between them. So, for simplicity, we assume *independence* between the arrival and departure processes and derive an *approximate* model. The independence assumption makes it very easy to relax the condition on m and k, since the joint probability of having $A(t) = m_{ij}$ and $B(t) = k_{ij}$ becomes product of probability of the two events. After the relaxation, we then substitute $\mu_1 = \lambda_1$ because of the steady-state condition, and get the followings:

$$E_n[(c-\overline{c})(p-\overline{p})] = \frac{atr}{2}(n-1)(\lambda_2 + \mu_2 - 2tr\lambda_1^2) \tag{13}$$

$$E_n[(c-\overline{c})^2] = \frac{t}{2}(\lambda_2 + \mu_2)(rn+1) - t^2\lambda_1^2(1 + r - r^2 + r^2 n) \tag{14}$$

$$E_n[(p-\overline{p})^2] = \frac{a^2 tr}{2}(n-1)(\lambda_2 + \mu_2 - 2tr\lambda_1^2) \tag{15}$$

Let σ_A^2 be the variance of the arrival process and σ_B^2 be the variance of the departure process. By substituting Equations 13, 15, and 14 into 6 we get the correlation model for the first n rounds as follows:

$$Corr_n = \frac{1}{at(\frac{\sigma_A^2+\sigma_B^2}{2} + \lambda_1^2)(rn+1) - a(t\lambda_1)^2(1 + r - r^2 + r^2 n)} \tag{16}$$

2) Model-II: To make a more realistic model, we try to develop a model where the arrival and departure processes are not considered independent. We consider the system as an $M/M/1$ queueing system with a service rate of μ. Notice that μ is different from the parameters μ_1 and μ_2 which are first and second moments of $B(\tau)$. We now try to derive the joint probability as follows:

$$P_{m_{ij};k_{ij}} = P_{m_{ij}} * P_{k_{ij}|m_{ij}} \quad (17)$$

where $P_{m_{ij}} = P\{A(t) = m_{ij}\}$ and $P_{k_{ij}|m_{ij}} = P\{B(t) = k_{ij}|A(t) = m_{ij}\}$. Notice that $P_{m_{ij}}$ is probability of having m_{ij} events for the Poisson distribution with mean $\lambda_1 t$. However, it is not that easy to calculate $P_{k_{ij}|m_{ij}}$, since probability of having k_{ij} departures depends not only on the number of arrivals m_{ij} but also the number already available in the system which is $c_{i(j-1)}$. Let N be the random variable that represents the number available in the system, then we can rewrite $P_{k_{ij}|m_{ij}}$ as follows:

$$P_{k_{ij}|m_{ij}} = \sum_{c_{i(j-1)}=k_{ij}-m_{ij}}^{\infty} P_{k_{ij}|m_{ij};c_{i(j-1)}} * P_{c_{i(j-1)}} \quad (18)$$

where $P_{c_{i(j-1)}} = P\{N = c_{i(j-1)}\}$. Observe that the minimum value of $c_{i(j-1)}$ can be $k_{ij} - mij$, because the condition $k_{ij} \leq m_{ij} + c_{i(j-1)}$ must be satisfied for all time intervals. In Equation 18, $P_{c_{i(j-1)}}$ is known for a steady-state $M/M/1$ system. Let $\rho = \lambda_1/\mu$, then $P_{c_{i(j-1)}} = (1-\rho)\rho^{c_{i(j-1)}}$. [19] However, calculation of $P_{k_{ij}|m_{ij};c_{i(j-1)}}$ is not simple, because the m_{ij} arrivals may arrive such that there is none waiting for the service. Fortunately, this is a very rare case for a loaded system. So, we can formulate $P_{k_{ij}|m_{ij};c_{i(j-1)}}$ for the usual case as if all the m_{ij} arrivals happenned at the beginning of the interval t. Within this context, we now derive $P_{k_{ij}|m_{ij};c_{i(j-1)}}$.

Let $E(\mu)$ be an Exponential random variable with mean $1/\mu$, and $E_r(k,\mu)$ be an Erlangian random variable with mean k/μ. Then, we can formulate the probability of having $k > 0$ departures in time t as follows:

$$P_{k>0 \ in \ t} = \int_0^t P\{E_r(k,\mu) < x\} \left[1 - P\{E(\mu) < t - x\}\right] dx \quad (19)$$

Now, we can formulate the CDF of $P_{k_{ij}|m_{ij};c_{i(j-1)}}$ as follows:

$$P\{B(t) \leq k_{ij}|m_{ij};c_{i(j-1)}\} = P_{0 \ in \ t} + \sum_{k=1}^{k_{ij}} P_{k>0 \ in \ t} \quad (20)$$

Notice that $P_{0 \ in \ t} = 1 - P[E(\mu) < t]$. By using the CDF formula in Equation 20, we then find pmf as $P_{k_{ij}|m_{ij};c_{i(j-1)}} = P\{B(t) \leq k_{ij}|m_{ij};c_{i(j-1)}\} - P\{B(t) \leq k_{ij}-1|m_{ij};c_{i(j-1)}\}$. Afterwards, we apply the operation in Equation 18. After going through all the explained derivations above [2], we finally derive $P_{k_{ij}|m_{ij}}$ as:

$$P_{k_{ij}|m_{ij}} = \frac{1}{\mu}\left(\frac{\lambda_1}{\mu}\right)^{(k_{ij}-m_{ij})} \left[1 - e^{-\mu t}\sum_{i=0}^{k_{ij}} \frac{(\mu t)^i}{i!}\right] \quad (21)$$

Even though we have found a nice solution to $P_{k_{ij}|m_{ij}}$ in Equation 21, it does not allow us to get a closed-form model for the correlation after the relaxation operation in Equation 12. In order to get a closed-form correlation model, we approximated [3] the summation term in Equation 21. We did get a closed-form correlation model after the approximation. But, it is not possible to provide it in hardcopy format [4] because it is a very large expression. However, we will provide numerical results of the model later in Section IV.

B. Model Discussion

Since Model-II is a very large expression, we only discuss Model-I. Assuming that the other factors stay fixed, the correlation model in Equation 16 implies three important results:

1) *The correlation degrades at most inversely proportional to an increase in pricing intervals (T):* For the smallest n value (i.e. 1), denominator of Equation 16 will have $r+1$ as a factor which implies linear decrease in the correlation value while the pricing interval, $T = rt$, increases linearly. Notice that its effect will be less when n is larger.
2) *Increase in traffic variances (σ_A^2 and σ_B^2) degrades the correlation:* From Equation 16, we can observe that the correlation decreases when the variance of the incoming or outgoing traffic increases.
3) *Increase in traffic mean (λ_1) degrades the correlation:* Again from Equation 16, we can see that the correlation decreases while the mean of the incoming traffic increases.

These above results imply that lower pricing intervals must be employed when variance and/or mean of the traffic starts increasing. We validate these three results in Section IV by experiments. Note that the model reveals non-intuitive effect of traffic mean on the correlation. Also, observe that the model incorporates not only the effect of pricing intervals on the correlation, but also the effects of statistical parameters (e.g. traffic mean and variance).

As previously mentioned, the correlation between prices and congestion measures is a representation of the achieved control over congestion. Congestion-sensitive pricing protocols can use such a model to maintain the control at a predefined level by solving the inequality $Corr_n \geq Corr_{min}$ for r, which defines the length of the pricing interval. If feedback from the other end (i.e. egress node in DiffServ [20] terminology) is provided, then such a model can be implemented in real-time. σ_B^2 can be calculated by using the feedbacks from the other end, and σ_A^2 and λ_1 can be calculated by observing the incoming traffic.

IV. Experimental Results and Model Validation

A. Experimental Configuration

We use Dynamic Capacity Contracting (DCC) [21] as the congestion-sensitive pricing protocol in our simulations. DCC provides a contracting framework over DiffServ architecture. The provider places its stations at the edge routers of the DiffServ domain. The customers can get network service through these stations by making *short-term contracts* with them. The provider station provides a variety of short-term contracts to the customer and the customer selects the contract which maximizes his/her utility. During the contracts, the station receives

[2] Derivations are available in the Maple file at *http://networks.ecse.rpi.edu/˜ yuksem/intervals/derivation.mws*

[3] The approximation is available in the Maple file at *http://networks.ecse.rpi.edu/˜ yuksem/intervals/approximations.mws*.

[4] It is available in the Maple file at *http://networks.ecse.rpi.edu/˜ yuksem/intervals/the_model.mws*.

congestion information about the network core at a time-scale smaller than contracts. The station uses that congestion information to update the service price at the beginning of each contract. The short-term contracts in DCC corresponds to the pricing intervals in our modeling.

In ns [22], we simulate DCC with varying pricing intervals (i.e. contract lengths). There are 5 customers trying to send traffic to the same destination over the same bottleneck with a capacity of 1Mbps. Customers have equal budgets and their total budget is 150 units. We observe the bottleneck queue length and use it as congestion measure. The observation interval is fixed at $t = 80ms$ and RTT for a customer is $20ms$. We increase the pricing interval by incrementing the number of observations (i.e. r) per contract. We run several simulations and calculate correlation between the advertised prices during the contracts and the observed bottleneck queue lengths.

Customers send their traffic with mean changing according to the advertised prices for the contracts. We assume that the customers have fixed budgets per contract with additional leftover from the previous contract. The customers adjust their sending rate according to the ratio B/p where B is the customer's budget and p is the advertised price for the contract [5]. Notice that since the customers' budget is fixed, the *average* sending rate of the customers is actually *fixed on long run*, which fits to the fixed average incoming traffic rate assumption in the model.

B. Results

In this section, we present several simulation results for validation of the model and the three results it implies.

Figures 4-a and 4-b show mean and variance of the bottleneck queue length. We can see significant increase (at least linear) in mean and variance of the bottleneck queue as the pricing interval increases linearly. Furthermore, Figure 4-c shows the change in the coefficient of variation for the bottleneck queue length as the pricing interval increases. Note that an increase in the coefficient of variation means a decrease in the level of control. We can observe that the coefficient of variation increases as the pricing interval increases until $10r$, and stays fixed there after. This is because the congestion-sensitive pricing protocol looses control over congestion after a certain length of pricing interval, which is $10r$ in this particular experiment. These results in Figures 4-a to 4-c validate our claim about the degradation of control when pricing interval increases. Furthermore, they also show that dynamic pricing does not help congestion control when the pricing interval is longer than a certain length.

To validate the model, we present the fit between our correlation models and experimental results we obtained from above mentioned simulation configuration. Figures 5-a and 5-b represent the correlations obtained by inserting appropriate parameter values to the model and corresponding experimental correlations, respectively for the cases $n = 15$ and $n = 25$. We observe that Model-II fits better than Model-I, which is mainly because of the dependency consideration between arrival and departure processes. Notice that the model is dependent on the experimental results because of the parameters for incoming and outgoing traffic variances (i.e. σ_A^2 and σ_B^2), pricing factor (i.e. a), and mean of the incoming traffic (i.e. λ_1). We first calculate the parameters σ_A^2, σ_B^2, a (ratio of average price by average bottleneck queue length) and λ_1 from the experimental results, and then use them in the model.

We now validate the three results implied in Section III-B. Figures 5-a and 5-b show that the correlation decreases slower than $1/r$ when the pricing interval (i.e. r) increases linearly. This validates the first result. Figure 5-d represents the effect of change in the variance of incoming and outgoing traffic (i.e. σ_A^2 and σ_B^2) on the correlation. The horizontal axis shows the increase in variances of both the incoming and outgoing traffic. The results in Figure 5-d for different values of n obviously show that an increase in traffic variances causes decrease in the correlation. This validates the second result. Finally for validation of the third result, Figure 5-c represents the effect of change in the mean of the incoming traffic (i.e. λ_1) on the correlation. We can see that increase in λ_1 causes decrease in the correlation. Another important realization is that the correlation is more sensitive to variance changes than mean changes as it can be seen by comparing Figures 5-c and 5-d.

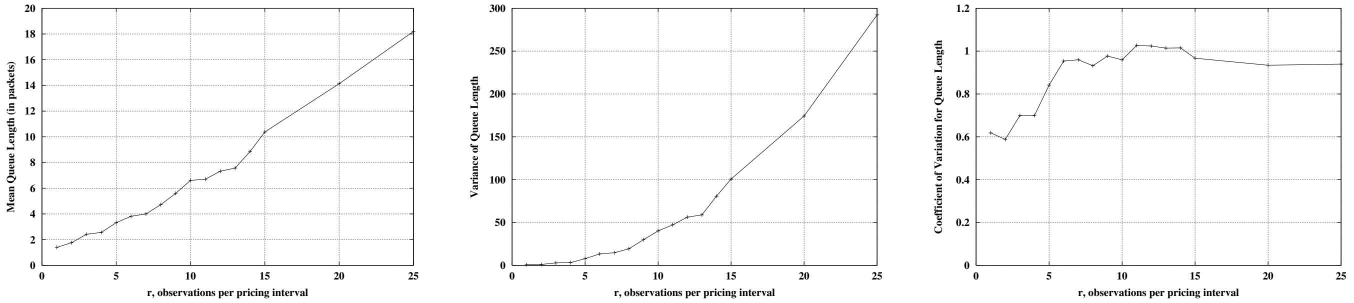
Before concluding this section, we would like to stress on the relationship between the correlation and the level of congestion control. As we previously stated, Figures 5-a and 5-b show the effect of increasing pricing intervals on the correlation for different values of n. We can see that the correlation value stays almost fixed after the pricing interval reaches to $10r$. Also, Figure 4-c shows the coefficient of variation for the bottleneck queue length in the experiments. Remember that coefficient of variation for the queue length represents the level of congestion control being achieved. We observe in Figure 4-c that it reaches to its maximum value (approximately 1) when the pricing interval reaches to $10r$, which is the same point where the correlation starts staying fixed in Figures 5-a and 5-b. So, by comparing Figure 4-c with Figures 5-a and 5-b, we can observe that the correlation decreases when the level of congestion control decreases, and also it stays fixed when the level of congestion control stays fixed. This shows that the correlation can be used as a metric to represent the level of congestion control.

V. Conclusions and Discussions

We investigated steady-state dynamics of congestion-sensitive pricing in a customer-provider network. With the idea that correlation between prices and congestion measures is a measurement for level of congestion control, we modeled the correlation. We found that the correlation decreases at most inversely proportional to an increase in pricing interval. We also found that the correlation is inversely effected by the mean and variance of the incoming traffic. This implies that congestion-sensitive pricing schemes need to employ very small pricing intervals to maintain high level of congestion control for current Internet traffic with high variance [23].

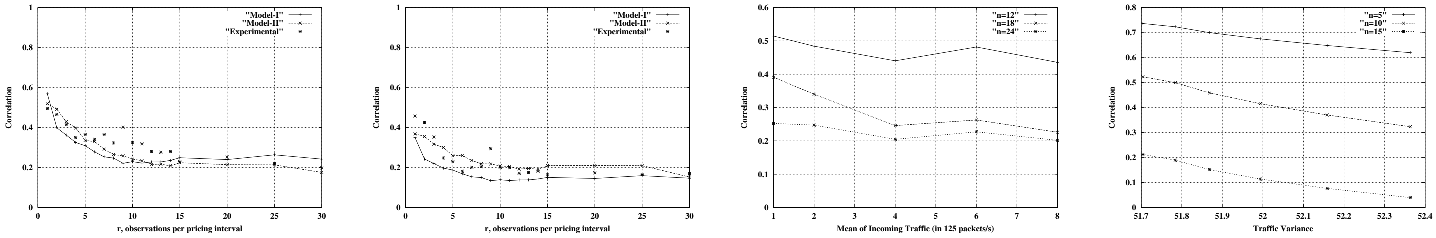
From the models and also from the simulation experiments we observed that the correlation between prices and congestion measures drops to very small values when pricing interval reaches to 40 RTTs even for a low variance incoming traffic. Currently, we usually have very small RTTs (measured by milliseconds) in the Internet. This shows that pricing intervals should be 2-3 seconds for most cases in the Internet, which is

[5]Note that the ratio $x = B/p$ maximizes customer's surplus given that his/her utility is $u(x) = B\ log(x)$.



(a) (b) (c)

Fig. 4. (a) Mean queue length (in packets), (b) Variance of queue length, (c) Coefficient of variation, $\frac{\sigma}{\mu}$, of queue length as the pricing interval (in number of observations) increases.



(a) (b) (c) (d)

Fig. 5. Fitting analytical model to experimental results for simulation time of (a) 15 pricing intervals, (b) 25 pricing intervals. Effect of change in the (c) mean (d) variance of the incoming traffic to the correlation for a pricing interval of $800ms$, i.e. $r = 10$.

not possible to deploy for many reasons such as low speed lines. This result itself means that deployment of congestion-sensitive pricing over the Internet is highly challenging.

The results obviously show that there will be need for intermediate middle-ware components (i.e. intermediaries) between individual users and ISPs, when ISPs deploy congestion-sensitive pricing for their service. These middle-ware components will be expected to lower price fluctuations such that price changes will be possible implement over low speed modems. This scenario suggests that congestion-sensitive prices can be implemented among ISPs to control congestion, but there has to be middle-ware components which can handle the transition of the congestion-sensitive prices to the individual customers in a smooth way. Alternatively, instead of using congestion-sensitive pricing directly for the purpose of congestion control, it can be used to improve fairness of an underlying congestion control mechanism. We believe that the second approach is more realistic way of implementing congestion-sensitive pricing over the Internet.

Another key implementation problem for congestion pricing is that current Internet access is point-to-anywhere. It is not possible to obtain information about the exit points of the traffic. However, it is not possible to determine congestion information and prices without coordinating entry and exit points of the traffic. So, this particular aspect implies that it is highly challenging to implement congestion pricing at individual user to ISP level. But, if an ISP has enough control over the entry and exit points, then it is possible. Alternatively, if ISPs of the current Internet collaborate on providing information about the entry and exit points to each other, then again it will be possible.

Future work will include complex modeling of the dynamics of congestion-sensitive pricing by relaxing some of the assumptions. For example, a model without fixed arrival rate assumption would represent the behavior of the system more appropriately.

Another important issue to explore is how much congestion control can be achieved with exactly what level of correlation between prices and congestion measures. In this paper we used the correlation value as a direct representation of the level of congestion control that was achieved. Although we supported this idea by providing the match between the correlation and the coefficient of variation in Section IV-B, this issue needs more investigation.

REFERENCES

[1] R. Cocchi, S. Shenker, D. Estrin, and L. Zhang, "Pricing in computer networks: Motivation, formulation and example," *IEEE/ACM Transactions on Networking*, vol. 1, December 1993.

[2] J. K. MacKie-Mason and H. R. Varian, "Pricing the congestible network resources," *IEEE Journal on Selected Areas of Communications*, vol. 13, pp. 1141–1149, 1995.

[3] D. Clark, *Internet cost allocation and pricing*, Eds McKnight and Bailey, MIT Press, 1997.

[4] A. Gupta, D. O. Stahl, and A. B. Whinston, *Priority pricing of Integrated Services networks*, Eds McKnight and Bailey, MIT Press, 1997.

[5] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control in communication networks: Shadow prices, proportional fairness and stability," *Journal of Operations Research Society*, vol. 49, pp. 237–252, 1998.

[6] J. K. MacKie-Mason and H. R. Varian, *Pricing the Internet*, Kahin, Brian and Keller, James, 1993.

[7] J. K. MacKie-Mason, L. Murphy, and J. Murphy, *Responsive pricing in the Internet*, Eds McKnight and Bailey, MIT Press, 1997.

[8] A. M. Odlyzko, "A modest proposal for preventing Internet congestion," Tech. Rep., AT & T Research Lab, 1997.

[9] X. Wang and H. Schulzrinne, "Pricing network resources for adaptive applications in a Differentiated Services network," in *Proceedings of Conference on Computer Communications (INFOCOM)*, 2001.

[10] X. Wang and H. Schulzrinne, "RNAP: A resource negotiation and pricing protocol," in *International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV)*, 1999, pp. 77–93.

[11] N. Semret, R. R.-F. Liao, A. T. Campbell, and A. A. Lazar, "Pricing, provisioning and peering: Dynamic markets for differentiated Internet services and implications for network interconnections," *IEEE Journal on Selected Areas of Communications – to be published*, 2001.

[12] N. Semret, R. R.-F. Liao, A. T. Campbell, and A. A. Lazar, "Market pricing of differentiated Internet services," in *Proceedings of IEEE/IFIP International Workshop on Qualityof Service (IWQoS)*, 1999, pp. 184–193.

[13] A. Orda and N. Shimkin, "Incentive pricing in multi-class communication networks," in *Proceedings of Conference on Computer Communications (INFOCOM)*, 1997.

[14] M. Yuksel and S. Kalyanaraman, "Simulating the Smart Market pricing scheme on Differentiated Services architecture," in *Proceedings of Communication Networks and Distributed Systems Modeling and Simulation Conference (CNDS) part of Western Multi-Conference (WMC)*, 2001.

[15] R. J. Edell and P. P. Varaiya, "Providing Internet access: What we learnt from the INDEX trial," Tech. Rep. 99-010W, University of California, Berkeley, 1999.

[16] A. M. Odlyzko, "The economics of the Internet: Utility, utilization, pricing, and quality of service," Tech. Rep., AT & T Research Lab, 1998.

[17] A. M. Odlyzko, "Internet pricing and history of communications," Tech. Rep., AT & T Research Lab, 2000.

[18] S. H. Low and D. E. Lapsley, "Optimization flow control – I: Basic algorithm and convergence," *IEEE/ACM Transactions on Networking*, vol. 7, no. 6, pp. 861–875, 1999.

[19] L. Kleinrock, *Queueing Systems, Volume I: Theory*, John Wiley and Sons, 1975.

[20] S. Blake et. al, "An architecture for Differentiated Services," *IETF RFC 2475*, December 1998.

[21] R. Singh, M. Yuksel, S. Kalyanaraman, and T. Ravichandran, "A comparative evaluation of Internet pricing models: Smart market and dynamic capacity contracting," in *Proceedings of Workshop on Information Technologies and Systems (WITS)*, 2000.

[22] "UCB/LBLN/VINT network simulator - ns (version 2)," http://www-mash.cs.berkeley.edu/ns, 1997.

[23] M. E. Crovella and A. Bestavros, "Self-similarity in World Wide Web traffic: Evidence and possible causes formulation and example," *IEEE/ACM Transactions on Networking*, vol. 5, no. 6, pp. 835–846, December 1997.