Pricing Granularity for Congestion-Sensitive Pricing *

Murat Yuksel and Shivkumar Kalyanaraman
Rensselaer Polytechnic Institute,
ECSE Department
110 8th Street, Troy, NY, 12180, USA.
yuksem@rpi.edu, shivkuma@ecse.rpi.edu

Abstract

One of the key issues for implementing congestion pricing is the pricing granularity (i.e. pricing interval or time-scale). The Internet traffic is highly variant and hard to control without a mechanism that operates on very low time-scales, i.e. on the order of round-trip-times (RTTs). However, pricing naturally operates on very large time-scales because of human involvement. Moreover, structure of wide-area networks does not allow frequent price updates for many reasons, such as RTTs are very large for some cases. In this paper, we investigate the issue of pricing granularity, identify problems, and propose solutions.

1 Introduction

One proposed method for controlling congestion in wide area networks is to apply *congestion-sensitive pricing* [3], which is a form of dynamic pricing. Many proposals have been made to implement dynamic pricing over wide area networks and the Internet [2, 6, 11, 10, 13]. 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 (or any other dynamic pricing protocol) makes it necessary to change the price after some time interval, what we call *pricing interval*.

Clark's Expected Capacity [2] scheme proposes long-term contracts as the pricing intervals. Kelly's packet marking scheme [6] 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 [9] proposes price updates at interior routers which cannot happen continuously and have to happen over some time interval. Wang and Schulzrinne's RNAP [11] framework proposes to update the price at each service level agreement which has to happen over some time interval. Hence, congestion pricing can only be implemented by updating prices over some time interval, i.e. pricing interval.

*This work is sponsored by NSF under contract number ANI9819112.

It has been realized that there are numerous implementation problems for dynamic or congestion-sensitive 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.
- *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 price of the service must be communicated to them before it is charged. This makes it necessary to inform the users of the network service 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-sensitive and adaptive pricing protocols. In this paper, 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 2, 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 an approximate analytical model for the correlation of prices and congestion measures in Section 3. In Section 4, we validate the model by simulation experiments and present the results. Finally, in Section 5 we discuss the implications of the work and possible future work.

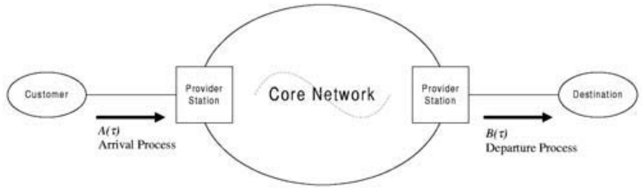


Figure 1. A sample customer-provider network.

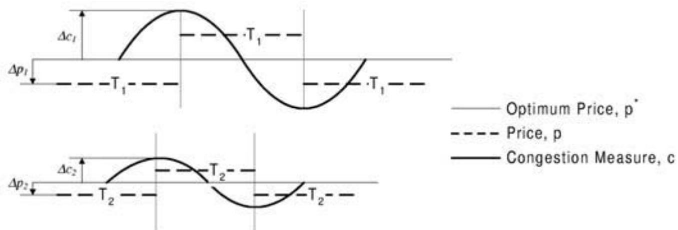


Figure 2. Congestion measure relative to congestion-sensitive prices in a steady-state network being priced.

2 Dynamics of Congestion-Sensitive Pricing

This section explains the behavior of congestion-sensitive 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 to send traffic to the destination through the provider's network. The provider observes the congestion level, c, in the network core and adjusts its advertised price, p, according to it. Note that c and p are in fact functions of time (i.e. $c(t)$ and $p(t)$ where t is time), but we use c and p throughout the paper for simplicity of notation. It is a realistic assumption to say that 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 relationship between c and p for two different pricing interval lengths, $T_1 > T_2$.

When the provider sees that the congestion level has been decreasing, it decreases the advertised price so that the network resources are not under-utilized. Then the customer starts sending more traffic in response to the decrease in price, and congestion level in the core starts increasing accordingly. The congestion level continues to increase until the price is increased by the provider at the beginning of the next pricing interval. When the provider increases price because of the increased congestion in the last pricing interval, the customer starts sending less traffic than before. Then congestion level starts decreasing. This behavior continues on in steady-state. This explains how congestion-sensitive prices can control the network congestion. The important difference is that with a larger pricing interval the congestion level oscillates larger as represented in Figure 2.

3 Analytical Model for Correlation of Prices and Congestion Measures

3.1 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. [8]

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

$$m_i = \sum_{s=1}^{r} m_{is} \tag{1}$$

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

$$k_i = \sum_{s=1}^{r} k_{is} \tag{2}$$

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

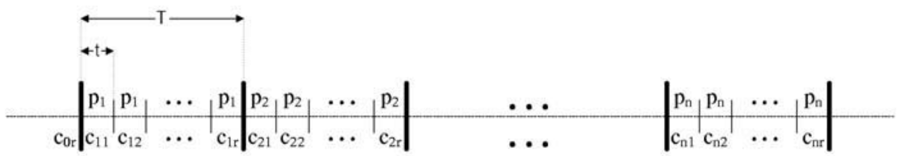


Figure 3. Prices and congestion measures for subsequent observation intervals.

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 following formula[1]

$$p_i = a(t,r)\ \ c_{(i-1)r} \tag{3}$$

where $a(t,r)$, *pricing factor*, is a function of pricing interval and observation interval defined by the congestion pricing algorithm. 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 price, but it splits the effect of congestion measures and interval lengths to price. We will use a instead of $a(t,r)$ for notation simplicity.

Within this context, the following equations hold:

$$c_{ij} = c_{0r} + \sum_{u=1}^{i-1}(m_u - k_u) + \sum_{s=1}^{j}(m_{is} - k_{is}) \tag{4}$$

$$c_{ir} = c_{0r} + \sum_{j=1}^{i}(m_j - k_j) \tag{5}$$

where $i \geq 1$. Reasoning behind (4) and (5) 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

$$\overline{p} = a\overline{c} \tag{6}$$

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.

$$c_{0r} = \overline{c} \tag{7}$$

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

$$Corr_n = \frac{E_n[(c-\overline{c})(p-\overline{p})|m,k]}{\sqrt{E_n[(c-\overline{c})^2|m,k]E_n[(p-\overline{p})^2|m,k]}} \tag{8}$$

[1] Note that this is a simplifying formula for tractability, and cannot capture all aspects of congestion pricing.

assuming that total of m packet arrivals and k packet departures happen during the n rounds.

We can calculate the numerator term in (8) as follows:

$$E_n[(c-\overline{c})(p-\overline{p})|m,k] = \frac{1}{rn}\sum_{i=1}^{n}\sum_{j=1}^{r}(p_i-\overline{p})(c_{ij}-\overline{c}) \tag{9}$$

By applying (3), (6) and (7) into (9) we can get

$$E_n[(c-\overline{c})(p-\overline{p})|m,k] = \frac{1}{rn}\sum_{i=1}^{n}\sum_{j=1}^{r}(ac_{(i-1)r}-ac_{0r})(c_{ij}-c_{0r}) \tag{10}$$

Then by applying (4) and (5) into (10), we get the following

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

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 obtain the variance of congestion measures and the variance of prices as follows:

$$E_n[(c-\overline{c})^2|m,k] = \frac{1}{rn}\sum_{i=1}^{n}\sum_{j=1}^{r} \left(H_1 + H_2 + 2\sum_{u=1}^{i-1}(m_u-k_u)\sum_{s=1}^{j}(m_{is}-k_{is})\right) \tag{12}$$

$$E_n[(p-\overline{p})^2|m,k] = \frac{a^2}{n}\sum_{i=2}^{n} H_1 \tag{13}$$

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 (11), (12), and (13). Specifically, we need to apply the operation

$$E_n[x] = \sum_{m_{ij}=0}^{\infty}\sum_{k_{ij}=0}^{\infty} E_n[x|m,k]P_{m_{ij};k_{ij}} \tag{14}$$

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.

3.1.1 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:

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

$$E_n[(c-\overline{c})^2] = \frac{t}{2}(\lambda_2+\mu_2)(rn+1)-t^2\lambda_1^2(1+r-r^2+r^2n) \quad (16)$$

$$E_n[(p-\overline{p})^2] = \frac{a^2tr}{2}(n-1)(\lambda_2+\mu_2-2tr\lambda_1^2) \quad (17)$$

Let σ_A^2 be the variance of the arrival process and σ_B^2 be the variance of the departure process. By substituting (15), (17), and (16) into (8) we get the correlation model for the first n rounds as follows:

$$Corr_n = \sqrt{\frac{r(n-1)(\sigma_A^2+\sigma_B^2+\lambda_1^2-2tr\lambda_1^2)}{(\sigma_A^2+\sigma_B^2+2\lambda_1^2)(rn+1)-2t\lambda_1^2(1+r-r^2+r^2n)}} \quad (18)$$

3.1.2 Model-II

As a more realistic model, we 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:

$$P_{m_{ij};k_{ij}} = P_{m_{ij}}P_{k_{ij}|m_{ij}} \quad (19)$$

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:

$$P_{k_{ij}|m_{ij}} = \sum_{c_{i(j-1)}=k_{ij}-mij}^{\infty} P_{k_{ij}|m_{ij};c_{i(j-1)}}P_{c_{i(j-1)}} \quad (20)$$

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 (20), $P_{c_{i(j-1)}}$ is known for a steady-state $M/M/1$ system. Let $\rho = \lambda_1/\mu$, then $P_{c_{i(j-1)}} = (1-\rho)\rho^{c_{i(j-1)}}$. [7] 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 happened 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:

$$P_{k>0\ in\ t} = \int_0^t P\{E_r(k,\mu) < x\}\left[1 - P\{E(\mu) < t-x\}\right]dx \quad (21)$$

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

$$P\{B(t) \leq k_{ij}|m_{ij};c_{i(j-1)}\} = P_{0\ in\ t} + \sum_{k=1}^{k_{ij}} P_{k>0\ in\ t} \quad (22)$$

Notice that $P_{0\ in\ t} = 1 - P[E(\mu) < t]$. So, we use $P_{0\ in\ t}$ to derive CDF in (22). Then, we derive pmf. Afterwards, we apply the operation in (20) and derive $P_{k_{ij}|m_{ij}}$ as:

$$P_{k_{ij}|m_{ij}} = \frac{1}{\mu}\left(\frac{\lambda_1}{\mu}\right)^{(k_{ij}-m_{ij})}\left[1 - e^{-\mu t}\sum_{i=0}^{k_{ij}}\frac{(\mu t)^i}{i!}\right] \quad (23)$$

Even though we have found a nice solution to $P_{k_{ij}|m_{ij}}$ in (23), it does not allow us to get a closed-form model for the correlation after the relaxation operation in (14). In order to get a closed-form correlation model, we approximated the term with summation in (23). Notice that the term with summation is equivalent to ratio of two Gamma [1] functions, i.e.:

$$e^{-\mu t}\sum_{i=0}^{k_{ij}}\frac{(\mu t)^i}{i!} = \frac{\Gamma(k_{ij}+1,\mu t)}{\Gamma(k_{ij}+1)}$$

We approximated the ratio $\Gamma(x,y)/\Gamma(x)$ by using geometric methods which cannot be explained within the length of this paper. After the approximation, we did get a closed-form correlation model. But, it is not possible to provide it in hardcopy format[2] because it is a very large expression. However, we will provide numerical results of the model later in Section 4.

[2]Available upon request.

3.2 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 (18) 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 (18) 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 (18), 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 (18), we can see that the correlation decreases while the mean of the incoming traffic increases.

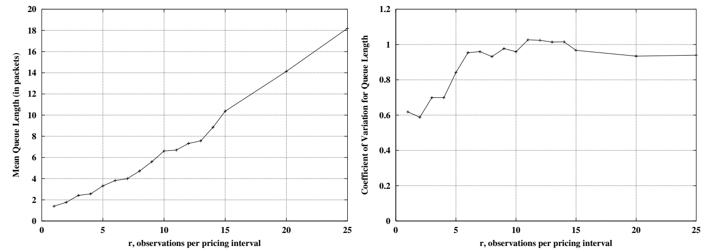
4 Experimental Results

4.1 Experimental Configuration

We use Dynamic Capacity Contracting (DCC) [12] as the congestion pricing protocol in our simulations. DCC provides a contracting framework over DiffServ [5] architecture. The provider places its stations at edges of the DiffServ domain. The customers can get network service through these stations by making *short-term contracts* with them. During the contracts, the station observes congestion in network core, and uses that congestion information to update the price at the beginning of each contract. The short-term contracts corresponds to the pricing intervals in our modeling.

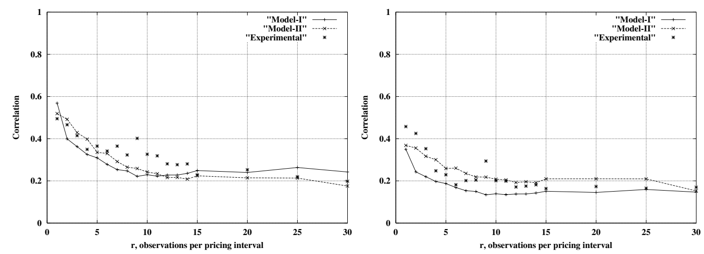
There are 5 customers trying to send traffic to the same destination over the same bottleneck with a capacity of 1Mbps. Customers have equal budgets and their total budget is 150 units. We observe the bottleneck queue length and use it as congestion measure. The observation interval is fixed at $t = 80ms$ 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 advertised prices and observed bottleneck queue lengths during simulations.

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 [3]. 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.



(a) Mean (in packets) (b) Coefficient of variation, $\frac{\sigma}{\mu}$

Figure 4. Statistics of bottleneck queue.



(a) For 15 pricing intervals (b) For 25 pricing intervals

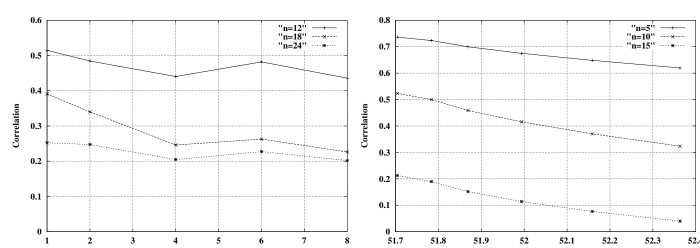
Figure 5. Fitting analytical model to experimental results.

4.2 Results

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

Figures 4-a shows mean of the bottleneck queue length. We observe steady increase in mean and variance of bottleneck queue as the pricing interval increases. Furthermore, Figure 4-b 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 observe that coefficient of variation increases as the pricing interval in-

[3] Note that $x = B/p$ maximizes surplus for a customer with utility $u(x) = B\ log(x)$.



(a) Effect of traffic mean (b) Effect of traffic variance

Figure 6. Effect of traffic patterns to the correlation (for $T = 800ms$ and $r = 10$).

creases until $10r$, and stays fixed there after. This is because the congestion 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 and 4-b 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 obtained from simulations. Figures 5-a and 5-b represent the correlations obtained by inserting appropriate parameter values to the model and corresponding experimental correlations, respectively for $n = 15$ and $n = 25$. We observe that Model-II fits better than Model-I, since Model-II considers the dependency 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 3.2. Figures 5-a and 5-b show that the correlation decreases slower than $1/r$ when r increases linearly. This validates the first result. Figure 6-b 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 6-b 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 6-a 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 6-a and 6-b.

5 Summary

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 mean and variance of the incoming traffic. This implies that congestion pricing schemes need to employ very small pricing intervals to maintain high level of congestion control for the Internet traffic with high variance [4].

From the model 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 not possible to deploy over low speed modems. This result itself means that deployment of congestion pricing over the Internet is highly challenging.

References

[1] D. G. Childers. *Probability and Random Processes*. McGraw Hill, Inc., 1997.

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

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

[4] 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*, 5(6):835–846, December 1997.

[5] S. B. et. al. An architecture for Differentiated Services. *IETF RFC 2475*, December 1998.

[6] 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*, 49:237–252, 1998.

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

[8] S. H. Low and D. E. Lapsley. Optimization flow control – I: Basic algorithm and convergence. *IEEE/ACM Transactions on Networking*, 7(6):861–875, 1999.

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

[10] 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*, 18(12):2499–2513, 2000.

[11] X. Wang and H. Schulzrinne. An integrated resource negotiation, pricing, and QoS adaptation framework for multimedia applications. *IEEE Journal on Selected Areas of Communications*, 18(12):2514–2529, 2000.

[12] M. Yuksel and S. Kalyanaraman. Distributed Dynamic Capacity Contracting: A congestion pricing framework for diff-serv. In *Proceedings of International Conference on Management of Multimedia Networks and Services (MMNS)*, pages 198–210, 2002.

[13] M. Yuksel and S. Kalyanaraman. A strategy for implementing Smart Market pricing scheme on diff-serv. In *Proceedings of Communication Quality and Reliability Symposium part of GLOBECOM*, 2002.