Pricing Experiments for a Computer-Telephony-Service Usage Allocation

Jimmy S. Shih, Randy H. Katz, Anthony D. Joseph
Department of Electrical Engineering and Computer Science
University of California at Berkeley
Berkeley, CA 94720-1770

Abstract-Charging a higher price during periods of congestion should more efficiently allocate scarce resources by encouraging users to conserve. We study its benefits by conducting several pricing experiments over two semesters with students in the dormitories using a computer-telephony-service. Users can use the service to make and receive phone calls from their computers or phones. While we do not charge users real money, we limit each user to a certain number of tokens a week. With this experimental setup, we conducted a different pricing experiment each week to better understand how prices can be used to entice users to talk less, talk at another time, or use a lower quality connection.

With our token scheme as a budget constraint, we can use static pricing policies to influence users' behaviors, but cannot use a simple congestion pricing scheme to encourage users to talk less. For example, we can use time-of-day pricing to encourage users to shift 30% of their usages from the peak to the off-peak hours. We can also use call-duration pricing, a higher rate as a call lasts longer, to encourage 3 times as many calls (18% instead of 6%) to terminate after a price increase. However, when using a simple congestion pricing scheme that charges a rate depending on the number of people calling, we find that we cannot get users to terminate their calls earlier. We believe that users do not change their behaviors because they do not know how long the price increases or decreases will last. Thus to make a congestion pricing scheme more effective, we believe that the price changes need to be more permanent to entice users to change their behaviors[1].

I. INTRODUCTION

Congestion pricing, charging a higher price when a resource usage is approaching its capacity, should be able to more efficiently allocate scarce resources by encouraging some users to conserve more. Thus, it can benefit the service providers by better utilizing the available resources and reducing the need for over-provisioning. It can also benefit the users by reducing congestion and avoiding degraded services.

The state-of-the-art on congestion pricing consists mostly of theoretical work. Several seminal papers [1,2,3] state the advantages and disadvantages of congestion pricing. Other papers [4,5,6,7] prove and simulate the benefits of congestion pricing using mathematical models for predicting user behaviors. However, there is no reported experience of using congestion pricing in a real system. There are also no user studies that measure the actual benefits of congestion pricing when compared with other pricing policies.

We conducted a user study with congestion pricing using a computer-telephony-service. Users can use the service to make and receive phone calls from their computers or phones. The service is similar to Internet telephony services like Dialpad and Net2Phone. We limit each user to a certain number of tokens a week and charge him/her a certain token rate for using the service. We deployed the service to students in the dormitories for two semesters. During the Fall of 2000, we had 41 students signed up and during the Spring of 2001, we had 91 students. With this setup, we conducted several pricing experiments to better understand congestion pricing.

With our token scheme as a budget constraint, we can use static pricing policies - prices are known to users beforehand - to change user behaviors. However, we cannot use a simple congestion pricing scheme to encourage users to talk less during congestion. For example, we can use time-of-day pricing - a static policy that charges a higher rate during peak hours – to successfully encourage users to shift 30% of their usages from the peak to the off-peak hours. We also can use call-duration based pricing - another static pricing policy that increases the rate as a call lasts longer – to encourage 3 times as many calls (18% instead of 6%) to terminate in the next minute after a price increase. However, when we use a simple congestion pricing scheme, we cannot get more people to terminate their calls after a price increase. We believe the reason is that they do not know how long the price increases or decreases will last. Thus to make congestion pricing more effective, we believe that the price changes need to be more permanent, or that the durations of the congested periods need to be more predictable, so that users can be more enticed to change their behaviors[1].

For the rest of this paper, we first discuss the related work on congestion pricing. We then describe the computer-telephony-service and the pricing experiments. Afterwards, we discuss how we deploy the service to conduct the experiments. Finally, we describe the experimental results before concluding with our findings.

II. RELATED WORK

Many papers [1,2] favor using congestion pricing for allocating scarce resources. They argue that congestion pricing can reduce congestion, better utilize available resources, send correct signals about expansion, and not place artificial limits on usages. However they also point out that congestion pricing requires a more complicated accounting mechanism, offers less predictable prices, and has users in the control loop. Fishburn and Odlyzko [3] also argue that flat-rate pricing is better than congestion pricing in the context of

---

[1] We confirmed this hypothesis in later congestion pricing experiments.

Internet bandwidth because it only requires a little bit more resource to gradually improve quality for everyone.

Several papers [4,5,6] use proofs and simulations to analyze congestion pricing. They show that congestion pricing can better utilize network resources and increase users' utilizations. However, Paschalidis and Tsitsiklis [7] show through their simulations that a static pricing policy like time-of-day pricing is almost as good as congestion pricing. These simulations depend a lot on how one models user reactions to changing prices.

We are unaware of reported experience with using congestion pricing in a real system. However, with regard to prototyping efforts, Stiller, et al. [8] use emulated routers to demonstrate that an IP telephony system can use congestion pricing for allocating bandwidth at IP routers. They simulate holding periodic auctions at each router for deciding whose voice-over-IP calls can transmit IP packets through it. Nevertheless, without real experience, we cannot know the appropriate schemes for setting congestion prices or presenting the prices to users.

To our knowledge, there are no reported user studies with using congestion pricing in a real system. However, Klausz, Croson, and Croson [9] have done a study using simulated games where users participate in auctions of scarce resources. They find that the prices set in these simulated auctions can increase utilization and reduce congestion. The INDEX project [10] performed a user study to measure user demand for Internet access using usage-based pricing. Usage-based pricing informs users the prices beforehand and charges according to how much resources they use. In contrast, congestion pricing modifies prices in the middle of a usage to reflect the current congestion. They find that the different pricing policies, like by volume or by connect time, can significantly impact user demand. For each pricing policy, they find that the demand is very sensitive to prices and that the difference among users is persistent and large. They also find that users can quickly adapt to a new pricing policy. These findings are very encouraging for our work on using congestion pricing for allocating resources.

III.   COMPUTER-TELEPHONY-SERVICE

We use a computer-telephony-service to conduct our pricing experiments because it has many desirable properties. First, phone calls are long enough that users can react to changing prices. Thus we can modify prices to encourage users to talk less, talk at another time, or talk using a lower quality connection. Second, calls go through gateways between the Internet and the PSTN. These are obvious bottlenecks because of the limited phone lines to the PSTN. Thus we can easily perform monitoring, admission control, accounting, and price settings at these places.

Users can use our service to make and receive phone calls from their computers or phones. From a computer, users can go to a web page, log in, and then enter the numbers they wish to call. See Fig. 1 for the web interface. When making calls from a computer, users will see on the web page real-time pricing information, like the current rate, the next minute rate, the call duration, the call charge, and the number of tokens they have left. From a phone, users can call a phone number, enter their user ids and pins, and then enter the numbers they want to call. The phone interface is very similar to using a calling card. When using a phone, users will hear the current price at the beginning of a call and whenever the price changes. They can also use the service to receive incoming calls from any computer or phone and then redirect them to their computers or phones as they choose.
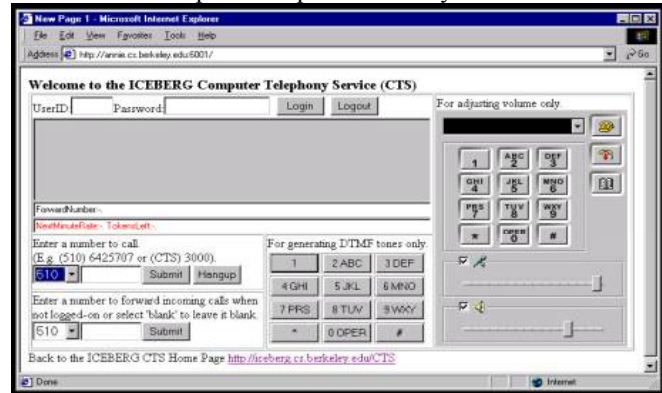

Fig 1. Web interface for using the service from a computer.

We implement the service using the H.323 protocol [11]. See Fig. 2 for a picture of the service's architecture. We use a Motorola Vanguard 6560 as a H.323 gateway to connect the Internet to the PSTN. We use a Primary Rate ISDN line to support 23 simultaneous calls between the gateway and the PSTN. We then implement a H.323 proxy for the gateway to perform functions like monitoring, admission control, accounting, and price setting. Users on their computers can use a web browser in conjuncture with a H.323 client like the Microsoft NetMeeting to make calls through the gateway via the proxy.
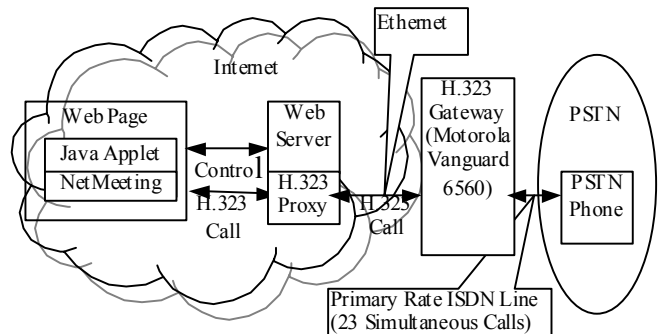

Fig. 2. Architecture for the computer-telephony-service.

There are only two modifications we needed to add to the service to support congestion pricing instead of flat-rate pricing. First we need to modify the accounting mechanisms on the H.323 proxy to generate an accounting record every minute instead of every call to better support dynamic prices. Second, we need to modify the proxy to provide real-time pricing information to users on their computers and phones either as updates on the web page or as recorded messages in the middle of a call.

IV. PRICING POLICIES

We initially use the computer-telephony-service to experiment with the following five pricing policies:

- Flat-rate pricing – same rate all the time.
- Congestion pricing – a rate that rises and falls with the total number of active calls.
- Time-of-day pricing – a higher rate during the peak hours.
- Call-duration pricing – a higher rate as a call lasts longer.
- Access-device pricing – a higher rate when using a phone and a lower rate when using a computer.

We use a flat-rate pricing as a worst case for comparison with congestion pricing. We select time-of-day pricing and call-duration pricing because they are static pricing policies that have the benefits of congestion pricing. Time-of-day pricing can encourage users to talk at another time while call-duration pricing can encourage users to talk shorter. We decide to experiment with access-device pricing because it can help reduce the load on the bottleneck resource, the phone lines to the PSTN, when users use their computers instead of phones to access the service.

V. DEPLOYMENT

We deployed the service to students in the dormitories during the Fall of 2000 and Spring of 2001 semesters. We target the students in the dormitories because they have good Internet connections, Ethernet access, and regular PSTN phones in their rooms. Students are familiar with using computers and like to talk on the phones. We had 41 users signed up during the Fall and 91 users during the Spring. Our users are mostly freshmen and sophomores, and come from a wide variety of departments.

We use a token system as a budget constraint that limits each user to 1000 free tokens a week and charges her/him a certain token rate for using the service. The unused tokens disappear at the end of the week so that we can perform a different pricing experiment each week. We do not charge users real money because it would have complicated the approval process for performing our user study in the dormitories.

Users can use our service to make and receive phone calls from their computers or phones. However, we limited the outgoing calls to only the San Francisco Bay Area during the Fall, but extended the call area to all of California during the Spring. Admittedly, our users have many options for making and receiving phone calls. On their computers, they can use free Internet telephony services like Dialpad and Net2Phone to make free long distance calls. However, our service has better voice quality because it is on the same local area network as their computers. On their phones, they can make free local calls or pay for long distance calls. Half of our users also have cell phones. Thus these outside options make it more challenging for us to use our token scheme to affect our users' calling behaviors.

Table 1 shows the pricing experiments we conducted during the two semesters. All the users face the same pricing policy each week. We started with access-device pricing because it is the easiest pricing policy for users to understand, a higher price for using a phone and a lower price for using a computer. We then experimented with static pricing policies, simple policies like time-of-day pricing and call-duration pricing, before experimenting with a more complicated policy like congestion pricing. In between these experiments, we used flat-rate pricing as a basis for comparison. The experiments are announced several weeks beforehand and users are reminded of the experiment in the beginning of the week. At the end of the week, users receive email statements describing their usages.

TABLE 1
PRICING POLICIES FOR THE TWO SEMESTERS.

Week Ending	Pricing Policy	Active Users	Total Calls	Total Minutes
10/16/00	flat-rate: 10 tokens/min	12	102	612
10/23/00	access-device: computer: 10 tokens/min phone: 20 tokens/min	15	91	406
10/30/00	flat-rate: 10 tokens/min	12	92	702
11/6/00	access-device: computer: 10 tokens/min phone:30 tokens/min	12	61	367
11/13/00	time-of-day: 11pm-7pm: 10 tokens/min 7pm-11pm: 30 tokens/min	13	117	657
11/20/00	call-duration: 1st-3rd min: 10 tokens/min 3rd min on: 30 tokens/min	12	59	349
11/27/00 (Thanks-giving)	congestion: 10X tokens/min X: number of simultaneous calls	10	63	530
12/4/00	congestion: 10X tokens/min X: number of active calls	14	85	837
12/11/00	flat-rate: 5 tokens/min	12	110	1238
2/12/01	flat-rate: 10 tokens/min	15	74	553
2/19/01	time-of-day: 11pm-7pm: 10 tokens/min 7pm-11pm: 30 tokens/min	21	133	927
2/26/01	call-duration: 1st-3rd min: 5 tokens/min 3rd–10th min: 10 tokens/min 11th-20th min: 20 tokens/min 21th min on: 30 tokens/min	31	128	961
3/5/01	flat-rate: 10 tokens/min	41	196	1925
3/12/01	congestion: 10X tokens/min X: number of active calls	42	251	2078
3/19/01	congestion: 5X tokens/min X: number of active calls	42	282	2445
3/26/01	flat-rate: 5 tokens/min	40	296	2441
4/2/01 (Spring break)	call-duration: 1st-5rd min: 5 tokens/min 6rd–15th min: 10 tokens/min 16th-25th min: 20 tokens/min 26th min on: 30 tokens/min	28	134	660

During the Fall, we had about 15 users using the service each week and during Spring, we had about 42 users a week. Table 1 shows that number of active users using the service each week and the total number of calls and minutes they made. Each week, about half of the users used up all their tokens.

VI. EXPERIMENTS AND RESULTS

Flat-Rate Pricing

We find that the results under flat-rate pricing are as expected. As we decrease the price from 10 tokens a minute to 5 tokens a minute, we find that each user uses the service more. This suggests that we do place a budget constraint on users by constraining each user to 1000 tokens a week and charging about 10 tokens a minute. In Fig. 3, we graph the calling pattern containing all the calls under flat-rate pricing. We observe that are obvious peak hours between 7pm-11pm where 54% of the usages occur during that time period.
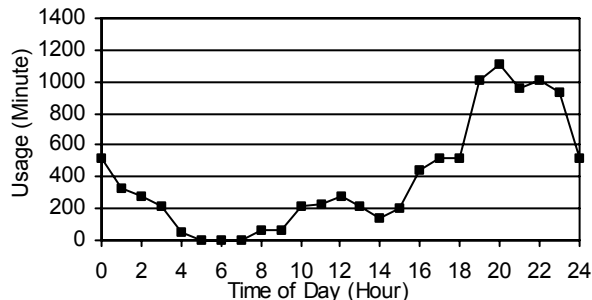
Fig. 3. Calling pattern of all the calls under flat-rate pricing.

In Fig. 4, we graph the percentage of calls that are longer than a certain duration. We find that about 30% of the calls are longer than 5 minutes, 20% longer than 11 minutes, and 10% longer than 23 minutes. Thus for call-duration pricing, we can predict the percentage of calls that will experience a price increase when we increase the price after a certain duration. In Fig. 5, we also graph the probability that a user will drop a call after a certain duration. We find that after the 3rd minute, the probability that a call will get dropped in the next minute is about a constant, 5.8% with a standard error of 0.4%.
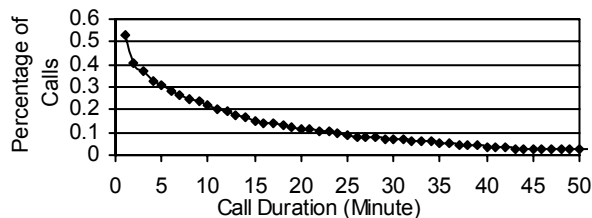
Fig. 4. Under flat-rate pricing, the percentage of calls longer than a certain duration.
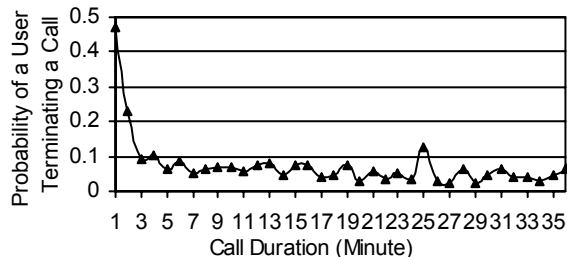
Fig. 5. Under flat-rate pricing, the probability that a user will terminate a call after a certain duration.

Time-Of-Day Pricing

For time-of-day pricing, we decided to first experiment with charging 30 tokens during the peak hours, 7pm-11pm, and 10 tokens during the off-peak hours, 11pm-7pm. We selected 30 tokens because we want a big price difference to see how much we can entice users to shift their usages. We performed the time-of-day experiment twice, once in the Fall and once in the Spring. When comparing their results with the flat-rate pricing the week before and the week afterwards, we find that we can encourage each user to shift about 30% of his usages from the peak to the off-peak hours. In Fig. 6, we aggregate the calls from the two time-of-day experiments and scale it to overlay with the flat-rate calling pattern in Fig. 3. We observe that with the 20 tokens price difference, the peak usage is over-shifted to before and after the 7-11pm period. We also observe a small peak around 11am from calls made during the weekends. Thus the results from the time-of-day experiments demonstrate that we can use our token scheme to entice users to call at another time.
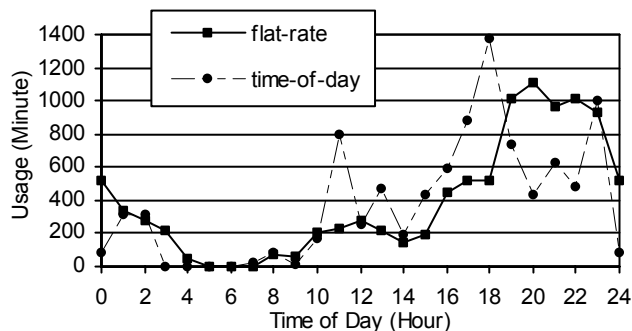
Fig. 6. Flat-rate pricing versus time-of-day pricing.

Call-Duration Pricing

We conducted three experiments with call-duration pricing and find that it can encourage users to talk less. For all the experiments, we inform users in the beginning of the week when a price increase will occur. In the first experiment, we increase the price after the 3rd minute and cannot cause more calls to terminate. In the second experiment, we increase the price after the 3rd, the 10th, and the 20th minute. We find that the price increase after the 3rd minute again has no effect, but the price increase after the 10th and the 20th minute cause about 3 times as many calls, 18% instead of 5.6%, to terminate in the next minute. See Fig. 7 for a comparison of the second call-duration experiment with the flat-rate behavior in Fig. 5. In the third experiment, we increase the price after the 5th, the 15th, and the 25th minute and are again able to cause about 3 times as many calls to terminate after each price increase. Thus our results from the call-duration experiments show that we can use our token system to encourage users to talk less after a price increase.
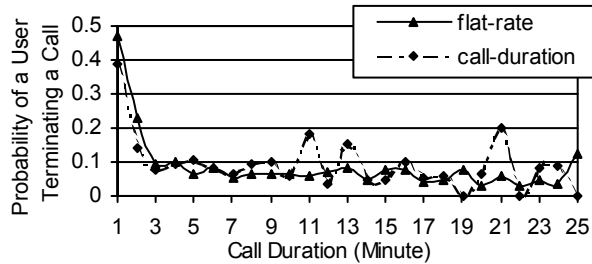
Fig. 7. Flat-rate pricing versus call-duration pricing with the price increase after the 3rd, the 10th, and the 20th minute of a call.

Congestion Pricing

For congestion pricing, we decided to set the current price as a function of the number of simultaneous calls on the service. Thus we increase the price when more people are calling and decrease the price when less people are calling. We conducted two congestion pricing experiments during the Spring semester. In the first experiment, we set the price equal to ten times the number of simultaneous calls. In the second experiment, we set it to five times. During the experiments, we have up to five people calling at the same time. During the first experiment, a price increase was announced 57 times and a price decrease was announced 56 times. During the second experiment, a price increase was announced 80 times and a decrease was announced 79 times. There are usually several minutes between price changes. However, in both experiments, we find that after a price increase or decrease, the percentage of calls that terminates in the next few minutes remain unchanged at around 6%. See Figure 8 for the percentage of calls dropped after a price increase.
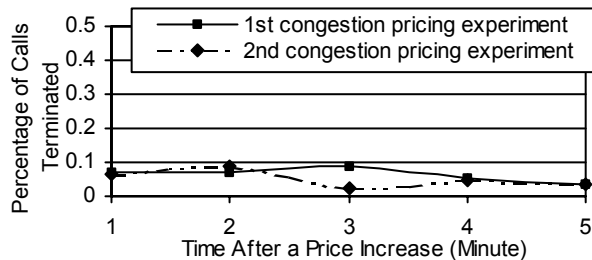

Figure 8: The percentage of calls terminating after a price increase.

We expected to see a higher percentage of calls get dropped after a price increase. From talking to some users, they do not react to price changes because they do not know how long the price increases or decreases will last. Whenever they notice a price increase, they hope that the increase is only temporary. Thus we think that for congestion pricing to be effective, the price changes need to be more permanent to entice users to change their behaviors[2].

Access-Device Pricing

For access-device pricing, we find that we cannot entice users to switch their access devices. Based on our surveys, even though using our service from a computer received high mean opinion scores, users felt that they do not need to switch because they can use free Internet telephony services like Dialpad and Net2Phone on their computers.

VII. CONCLUSION

Congestion pricing has the potential to efficiently allocate scarce resources. To study its benefits, we use a computer-telephony-service to conduct different pricing experiments with students in the dormitories. We deployed the service for two semesters to about 40 users. We do not charge users real money, but limit each user to a certain number of tokens a week for using the service. With our token scheme as the budget constraint, we can use static pricing policies like time-of-day pricing and call-duration pricing can entice users to talk at another time or talk shorter. However, we cannot use a simple congestion pricing scheme to encourage users to talk less. The users do not change their behavior under our congestion pricing scheme because they do not know how long the price increases or decreases will last. Thus to make congestion pricing more effective, we believe that the price changes need to be more permanent to entice users to modify their behaviors[2].

REFERENCE
1. J.K. MacKie-Mason, L. Murphy, and J. Murphy, "The role of responsive pricing in the Internet," *Tech. Report, University of Michigan*, June 1995.
2. J.K MacKie-Mason, and H.R. Varian, "Pricing the Internet," *JFK School for Government*, May 1993.
3. P.C. Fishburn, and A.M. Odlyzko, "Dynamic behavior of differential pricing and quality of service options for the Internet," *Decision Support Systems*, vol. 28, pp. 123-36, March 2000.
4. L. Murphy, J. Murphy, and J.K. MacKie-Mason, "Feedback and efficiency in ATM networks," *1995 IEEE International Conference on Communications (ICC '96)*, vol. 2, New York, NY, USA: IEEE, 1996, pp. 1045-9.
5. J.M. Peha, "Dynamic pricing as congestion control in ATM networks," *GLOBECOM 97. IEEE Global Telecommunications Conference*, vol. 3, New York, NY, USA: IEEE, 1997, pp. 1367-72.
6. X. Wang, and H. Schulzrinne, "An integrated resource negotiation, pricing, and QoS adaptation framework for multimedia applications," *IEEE Journal on Selected Areas in Communications*, vol.18, (no.12), IEEE, Dec. 2000, pp.2514-29.
7. I.Ch. Paschalidis, and J.N. Tsitsiklis, "Congestion-dependent pricing of network services," *IEEE/ACM Transactions on Networking*, vol. 8, (no. 2), pp. 171-84, April 2000.
8. B. Stiller, G. Fankhauser, G. Joller, P. Reichl, and N. Weiler, "Open charging and QoS interfaces for IP telephony," *INET '99: The Internet Summit*, San Jose, California, U.S.A., June 22-25, 1999.
9. F.J. Klausz, D.C. Croson, and R.T.A Croson, "An experimental auction to allocate congested IT resources: the case of the University of Pennsylvania modem pool," *Proceedings of the Thirty-First Hawaii International Conference on System Sciences*, Los Alamitos, CA, USA: IEEE Comput. Soc, 1998, pp.363-73, vol.6.
10. R. Edell, and P. Varaiya, "Providing Internet access: what we learn from INDEX," *IEEE Network*, vol. 13, (no. 5), pp. 18-25, September-October 1999.
11. ITU-T Recommendation H.323 Version 3, "Packet based multimedia communication systems," 1998.

---

[2] We confirmed this hypothesis in later congestion pricing experiments where we informed users that each price change will last at least 3 minutes.