Dynamic Pricing with "Alternatives" for Mobile Networks

Saravut Yaipairoj Prof. Fotios C. Harmantzis
Telecommunications Management
Stevens Institute of Technology
Hoboken 07030, NJ USA
{syaipair, fharmant}@stevens.edu

Abstract

Dynamic pricing has been used as an effective mean to alleviate network congestion in mobile networks. It is also potentially capable of gaining higher revenue over fixed pricing scheme. However, dynamic pricing scheme fails to provide user satisfaction since there is absolutely no guarantee that, during the time of congestion, users would get the service with their expected price. In this paper, we propose a new pricing scheme that gives choices to mobile users between dynamic prices with superior quality of service or fixed low price with acceptable performance degradation. The performance results verify that the new pricing scheme improves the overall system utilization and yet guarantees users' satisfaction.

1. Introduction

In recent years, demand for mobile services has been rising exponentially, however, the bandwidth and frequency spectrum for mobile services is critically limited. Network planners have put a lot of efforts on the channel allocation scheme and spectrum efficiency improvement in order to allocate resources to users more efficiently [6]. Such well-known scheme is Call Admission Control (CAC). It is a strategy used to admit only limited number of users to the network in order to alleviate network congestion and provide desired Quality of Service (QoS). However, during mass-calling events such as at the end of concerts or football games, car accidents in downtown or blackouts and other rare catastrophic events, the network is overwhelmed by high volume of customers' calls [16]. Blocking out customers during time of congestion would help alleviate network congestion. However, it would result in unsatisfied customers and lost of revenue. In term of frequency spectrum usage, frequency reuse has been an effective technique to spread and reuse limited spectrum in cellular network. The frequency spectrum is extremely expensive. With the rising demand for mobile services, network planners try to increase available capacity through cell splitting and frequency reuse. However, even with the high level of frequency reuse, (the introduction of micro and pico cells) network planners still find network capacity to be limited. Furthermore, high level of frequency reuse requires costly implementation and potentially yields more interference in the system. Therefore, there are absolutely no efficient ways to solve the limited resource problems and guarantee QoS if the call arrival rate is temporarily high during peak hours [5].

Mobile users act independently and, sometimes, "selfishly" without considering the current network traffic conditions. As a matter of fact, when one user is admitted into the network, it will cause QoS degradation to other users. If users maximize their individual level of satisfaction, system overload situations will not be avoided [2],[14]. Therefore, we need a mechanism that provides incentives for users to behave in ways that improve overall utilization and performance. In commercial networks, pricing had been proposed as an effective mean to resolve the allocation of these scarce resources to users.

Network users are inherently price sensitive. Using prices, the network can send signals to the users, providing incentives, which influence their behavior [1]. In [3], Norris and Khanifar have introduced the concept of *dynamic pricing* used in mobile networks. In this particular scheme, the price of calls changes as demand fluctuates. It rises in accord with demand, deterring additional users from accessing the network or holding network resources for long periods during the peak hours. It has been proved that this pricing scheme can alleviate network congestion and potentially yields more revenue to the system, see [7,8,9,12,13].

Despite the beneficiary of dynamic pricing, it is hard to come up with a tariff that mobile users can accept since there is no guarantees (at the time of congestion) that users would get their services at their expected price [11]. In this paper, we propose a pricing scheme so that mobile users have choices ("alternatives") by either accepting the services with higher price based on dynamic pricing scheme or holding on to the conventional scheme (fixed low rate) with acceptable degradation in performance. We argue that our pricing scheme is superior to existing traditional dynamic pricing schemes, in terms user satisfaction.

The paper is organized as follows. In Section II, we describe the drawback of dynamic pricing in mobile networks and introduce the concept of the proposed pricing scheme as a solution to the problem. In Section III, our proposed model for dynamic pricing with "alternatives" is presented and the parameters involved are defined. Section IV shows numerical results of our pricing model. Also, discussion on the results is presented. Section V, we present the conclusion of our work and our future research direction.

0-7803-8344-3/04/$20.00 (C) 2004 IEEE

2. Dynamic Pricing with "Alternatives"

We assume that mobile users are "selfish" (in terms of network usage) and price-sensitive. If we deploy dynamic pricing scheme at the time of congestion, the users who will accept the dynamic price (higher than normal) are the ones who are in immediate need for the service. As for the rest of the users, they would find the dynamic price unacceptable and leave the system. Some of them might try accessing the system again. Some of them might not. As the system operates, the network will be less congested and possibly yield more revenue with the cost of dissatisfaction of users who are not admitted to the system. Therefore, the mechanisms that appropriately handle this type of users are needed.

In our proposed pricing scheme, namely *Dynamic Pricing with "Alternatives"*, we believe that we can trade off service blocking with service delay. Users who do not accept the dynamic price will be given an alternative choice. They can either leave the system or put their call request in a queue and wait for service. This is similar to the queue card used in restaurants. The users, known as *priority users*, would accept the higher price from the system and will be served immediately or relatively sooner. While the rest of the users, known as *conventional users*, are delayed by system for certain amount of time, then, they will be served with their expected price. We will explain the call procedure in the next section. However, with this approach, the wireless system is able to shape the incoming traffic to conform to the amount of network resource available in the system. At the same time, conventional users are satisfied assuming that they are more willing to wait for a certain amount of time than getting blocked by price barrier.

3. Proposed Model for Pricing

Call admission control (CAC) is widely used as a mean to prevent mobile networks from overloading. During system congestion, the calls admitted to the network are required to meet a certain criteria such as pricing requirement, amount of resource requested, etc. In this context, the type of CAC, namely, *Priority Call Admission control* (PCAC) is introduced in order to control incoming calls based on user priority, which is regulated by pricing policy.

As shown in Fig 1, the system consists of two types of queues. One is the queue for priority users. The other is for conventional users. Both of them store the incoming calls that are waiting for available channels. The key of this system model is two functional blocks: the pricing block and priority call admission control block (PCAC). The pricing block diagram performs as the price broadcasting point to incoming users. Unlike traditional pricing schemes where users are blocked if they cannot meet the pricing criteria, in the proposed system, the mobile users who do not accept higher price will be placed in the queue for conventional users and wait to get served. For the priority users, their call request will be placed in the priority queue and wait for service. The above procedure will only take place when the system experiences congestion. If the congestion level is not met, all calls will be placed in the queue of conventional users and served when the system is ready.

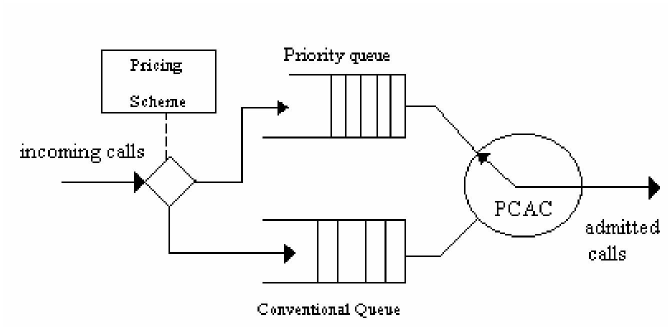

Figure 1. Dynamic Pricing with "Alternatives"

The PCAC blocks can be described as the QoS controller, which adjusts the balance of the traffic coming from both queues. Since priority users require more attention than conventional servers, priority queues will be served in such a way that the certain QoS for priority users is met. At the same time, conventional users are also served with some QoS that is, however, inferior to the priority users. The objective of this system is how we can adjust the traffic in such a way that we can meet the QoS constraint for both queues and maximize the number of calls being served at the same time. The call procedure of the system can be described as follows.

3.1. Call procedure

1. The user dials in numbers and waits for response from the mobile network.
2. The status of network is identified. If the network is not congested, the call requests will be placed in conventional queue waiting for available channels.
3. If the network is congested, the system will notify the users the approximated time they have to wait for service. Then, it will announce the price for users who want to be priority user and ask for their decision (be prioritized or stay on the line).
4. If the answer is yes, the user call request will be placed in priority queue where they are served with superior QoS.
5. For those who stay on the line, their call requests will be placed in conventional queues waiting for available channels. Once there is channels available, system notifies mobile users through control channels. As soon as mobile users confirm the notification, the system automatically occupies the traffic channel and dial in numbers as requested.

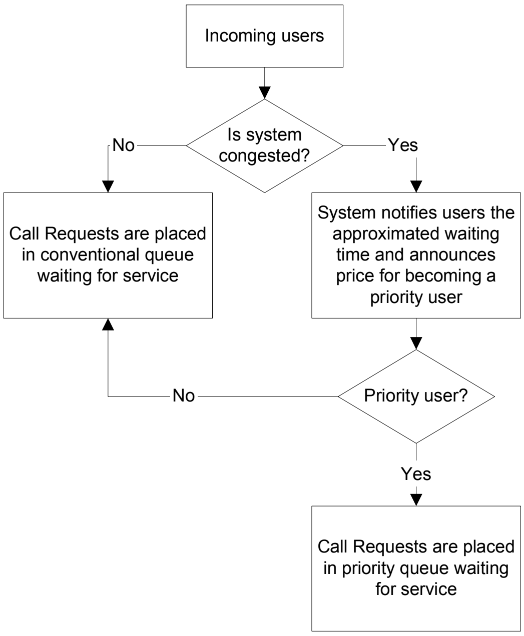

Figure 2. Call procedure

3.2. Resource partitioning

The most important parameter used in PCAC block is *priority factor* (P_s)., i.e., the probability that priority queue is served by the system over conventional queue. With higher P_s, system would serve priority queue more often than conventional queue. In our assumptions, we assume that the average holding time of priority users will be smaller than that for the conventional users since priority users pay higher price. Hence, they tend to spend less time in the system. Both types of users share the same resources (wireless channels in this case). Therefore, the average holding time that both users experience from the trunk group would be the weighted sum of holding time of priority users and conventional users, which is

$$T_{avg} = P_s T_{avg\_p} + (1 - P_s) T_{avg\_c} \tag{1}$$

Where T_{avg_p} is average holding time of priority users and T_{avg_c} is average holding time of conventional users. Since P_s regulates how much resource assigned for priority users and how much the rest of resources left assigning for conventional users. Hence, we can assume that the wireless channels are divided into two groups of wireless channels. One group is assigned to priority users and the other group is assigned to conventional users. Both of them can be considered independently. The amount of wireless channels available for each type of users is in the following.

Number of channels assigned for priority users

N_p = P_s * *(Total channels)*

Number of channels assigned for conventional users

N_c = *(1-*P_s*) * (total channels)*

3.3. Dynamic Price

The price broadcasted to mobile users when the system reaches the congestion level fluctuates by traffic demand. Percentage of mobile users who accept dynamic price (q) depends heavily on the ratio between the dynamic price and normal price (fixed rate). The q can be mostly characterized by demand function. The demand function describes the reaction of users to the change of price. We use the demand function that appears in [4] since it is used for different priority users, which fits our model. The demand function is as follows.

$$q = e^{-(\frac{p_h}{p_0} - 1)^2} \qquad p_h \geq p_0 \tag{2}$$

Where p_0 is the normal price and p_h is the price charged to priority users, therefore, we have

$$p_h = p_0 + \frac{p_0\sqrt{-4\ln(q)}}{2} \tag{3}$$

The percentage of incoming users (q) provides the information about the amount of mobile users whose their call requests are placed in each queue. Since we assume that the performance of each type of queue can be considered independently, the model can be composed of two basic queuing models with known average holding time from (1) and arrival dictated by demand function at a given price (2). These basic queuing systems are basically M/M/m system, which can be identified by erlang-C formula, i.e.

$$C^{-1}(N,a) = B^{-1}(N,a) - B^{-1}(N-1,a) \tag{4}$$

where $$B(N,a) = \frac{\frac{a^N}{N!}}{\sum_{i=0}^{N} \frac{a^i}{i!}}$$ (erlang-B formula)

The QoS of both types of queues can be identified by the user delay in the queues. That means the time that their call requests spend in the queue can be guaranteed. User delay in the queue can be seen as the tail of a delay distribution, i.e. P[user delay > R seconds] is less than a QoS requirement (such as 1%). Therefore, with the help of erlang-C formula [15] and the fact that we can consider both queue independently, we can identify the QoS requirement as follow.

$$P[W > t_p] = C(N_p, a_p)e^{-N_p\mu(1-\rho)t_p} \tag{5}$$

$$P[W > t_c] = C(N_c, a_c)e^{-N_c\mu(1-\rho)t_c} \tag{6}$$

Where C (N,a) is erlang-C formula, W is user delay (time in queue), $a_{( )}$ is the load from each type of users, $N_{( )}$ is the number of channels logically assigned for particular type of users, μ is average departure rate of users (1/Tavg), ρ is the load per server, t_p and t_c are QoS requirement for priority queue and conventional queue respectively. In another word, the system will guarantee the time that mobile user's call request would not exceed a certain amount of time. We assume that t_p would be a lot less than t_c when the system experiences congestion. Utilization of the system can be described as

$$Utilization = \frac{call\_arrival\_rate * T_{avg}}{N} \qquad (7)$$

3.4. Optimal Call Arrival Rate

As the system operates, system resources are shared in the way that the QoS for each user type can be achieved. The important parameter that we look for is the maximum number of users that the network can accommodate and yet conform to the QoS constraints. This parameter can be reflected by the optimal call arrival rate (λ_{opt}). We know that λopt is embedded in (5) and (6). With different values of percentage of users being priority users (q) and priority factor (Ps), we can achieve a certain arrival rate. However, λ_{opt} can be obtained when it maximizes the utilization of the system.

To obtain λ_{opt}, we need to consider (5) and (6). QoS constraint in (5) and (6) can be set at certain probability depending upon the user requirement. The λ_{opt} can be found by setting (5) and (6) to equal the QoS constraint. Then, we observe that arrival rate (λ) is the function of P_s and q. Therefore, with certain values of q, N, Tavg_c and Tavg_p, we can find P_s that yield maximum call arrival rate or λopt. That is

$$\lambda_{opt} \text{ for certain } (q) = f(P_s^*, q) \qquad (8)$$

Where P_s^* satisfies $\frac{df(P_s, q)}{dP_s} = 0$

4. Performance Analysis

In this section we evaluate the performance of the proposed dynamic pricing model with "alternatives". We observe that the proposed pricing model accommodates more users and meet the QoS requirements of both types of mobile users. In term of utilization, we found that we can achieve high utilization regardless of percentage of priority users (q). In addition, our pricing model also yields higher revenue with minimum percentage of priority users. In Section 4.1 we describe in detail the basic assumptions and parameters used in the dynamic pricing model with "alternatives" and the result of our analysis is shown in Section 4.2.

4.1. Assumptions and parameters

For simplicity, we assume the mobile base station handles circuit-switched voice calls. The arrival rate of incoming calls is Poisson arrival and exponentially distributed. The handoff calls into the system are handled using reserved channels at each base station for continuation of the calls and will not follow the same procedures as the call initiated within the base station. The system queues are FCFS queues. The parameters used throughout our analysis are typical in the common practice of wireless networks. We assume that:

(1) There are 30 wireless channels in total. Channels assigned for each queue is regulated by Ps.
(2) The average call holding time for priority users (*Tavg_p*) and conventional users (*Tavg_c*) are exponentially distributed with mean 120 seconds and 300 seconds respectively.
(3) The QoS parameters for our queuing system are:
    a. Probability of priority users' call request waiting in the queue for more than 1 minutes (t_p) is less than 1%
    b. Probability of conventional users' call request waiting in the queue for more than 10 minutes (t_c) is less than 1%
(4) The normal charging rate for conventional users using channels (P_o) is 8 cents per minutes. The charging rate for priority users (P_h) depends upon demand function and it is broadcasted upon arrivals.

4.2. Numerical results

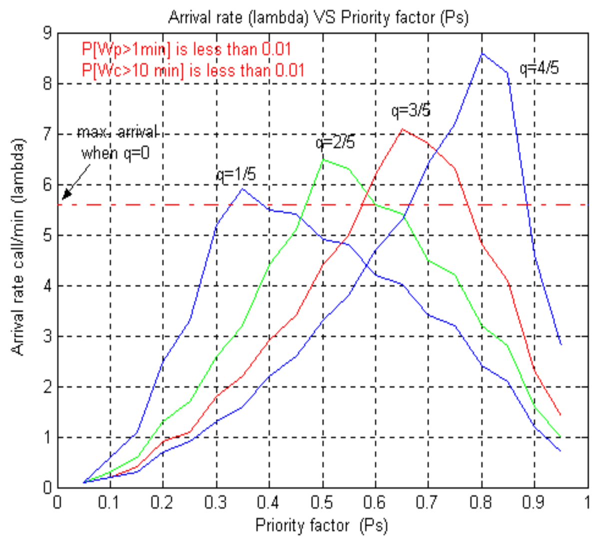

Figure 3. Optimal Arrival rate for certain percentage of priority users

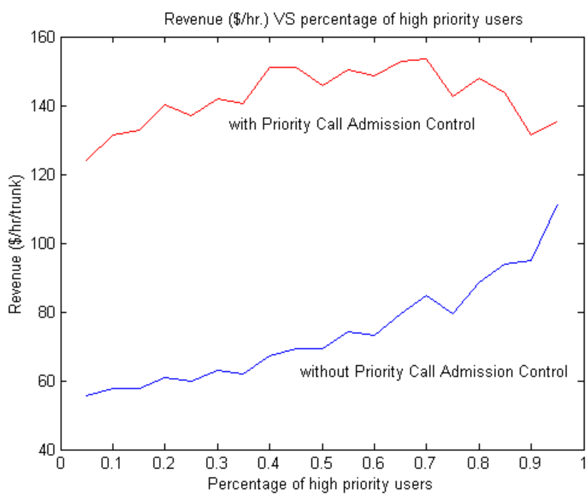

Figure 4. Revenue created by dynamic pricing with "alternatives" ($ per hour per channel)

Figure 3 shows the relationship between the arrival rate and Priority factor (P_s). For a certain percentage of priority users (q), there is an optimal call arrival rate that maximizes the number of calls (at the peak point of the curves). At the optimal call arrival rate, the system can accommodate the maximum number of mobile users while it still maintains the QoS requirement for both types of users. As the q increases, the optimal call arrival rate increases. Figure 3 also shows the improvement of call accommodation when we apply the proposed pricing model on the mobile system. Specifically, without priority scheme, the optimal arrival rate is only 5.6 calls/sec. This value of arrival rate is derived from erlang B formula when P_b=1% (our QoS requirement). However, when q increases to 80%, optimal arrival rate increases to 8.5 calls/sec. We can even achieve higher optimal call arrival rate by degrading the QoS requirement of either type of users.

Figure 4 shows the revenue created from each mobile channel using the proposed dynamic pricing scheme, and the revenue created from the traditional fixed rate system with the same amount of traffic. The revenue from our proposed model can be obtained by the sum of the revenue created from priority users and conventional users with their respected price factor. We observe that there is significant revenue increase using our pricing model over the system without it. In addition, we found that the percentage of priority users (q) does not yield significant effect on the revenue. Therefore, we can operate our system at minimum percentage of priority users and still create high revenue.

In order to present the beneficiary of dynamic pricing with "alternatives" in terms of channel accommodation, Figure 5 shows the variation of the new call arrival rate during a 24-hour period, that was used throughout our study. During 9am to 15pm, the demand for mobile usage is high. We consider this period as "peak hours". To guarantee that the system will not be overloaded, we need to deploy our pricing scheme at this period of time.

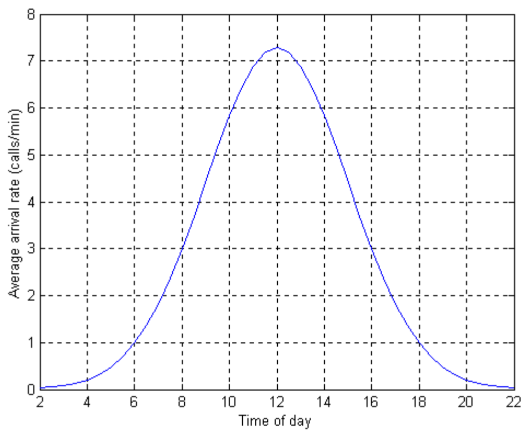

Figure 5 Call arrival rate used in the model as a function of time

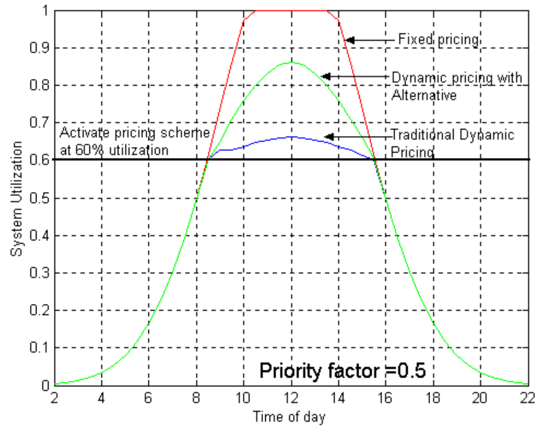

Figure 6. Average number of customers with different types of pricing scheme (priority factor =0.5)

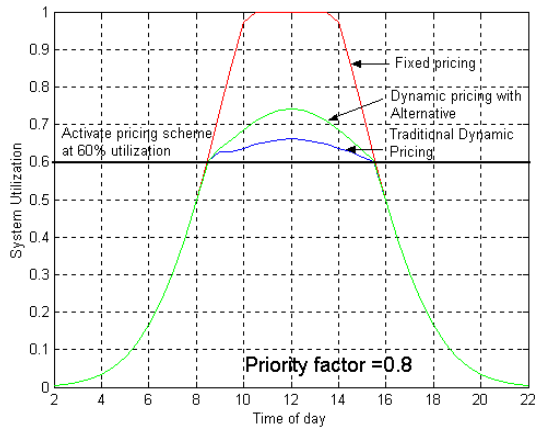

Figure7. Average number of customers with different types of pricing scheme (priority factor =0.8)

Figure 6 shows the comparison among different types of pricing scheme used in the mobile system. Fixed pricing scheme accommodates large amount of mobile users. However, without users' incentive to complete calls more quickly, the system is overloaded. Therefore, the system is forced to block mobile users, causing revenue loss and user dissatisfaction. With strict dynamic pricing scheme, the number of incoming mobile users drops significantly as price changes because mobile users are discouraged and frustrated by the price barrier. As a result, the utilization is significantly contained at a certain level (less than 70% in our model). For our proposed pricing scheme, dynamic pricing with "alternatives" also contains the utilization (at 85% in our model). This is achieved not by blocking users out, but by simply delaying their call requests and allows users to be prioritized. Priority users in the system are given the price incentive to complete their calls more quickly causing the average number of users in the system to drop. Without user blocking, user satisfaction can be significantly increased. In addition, Figure 8 shows the significance of priority factor. We can further contain the utilization of mobile users by adjusting priority factor of the system. As shown in Figure 8, by increasing priority factor up to 0.8, we can reduce system utilization by approximately 10%.

5. Conclusions

In conclusion, even though this framework only serves as a primitive starting point to improve dynamic pricing scheme in mobile network, we found that our proposed pricing model yields satisfactory results. The system can accommodate more traffic during peak hours while keeping utilization at an optimal operating point. The mobile users are more satisfied with the fact that they can choose whether they want to be served quickly by being priority users with high service charges or act as conventional users, who get the same low rate with acceptable QoS degradation. In addition, by adjusting the pricing factor and user QoS requirement, the proposed pricing model is flexible enough to alleviate traffic congestion and meet expected revenue. In the future, we are interested in extending our model considering multiple alternatives for multi-service networks. This is expected to increase both user satisfaction and system revenue. Auction-based pricing is also under current investigation.

6. References

[1] J.K MacKie-Mason and H.R. Varian, "Pricing Congestible Network Resources", IEEE Journal on Selected Areas in Communications, Vol. 13 Issue: 7, Sep 1995 pp. 1141 – 1149

[2] M. Falkner, "A user's perspective on Connection Admission Control: Integrating Traffic Shaping, Effective Bandwidths and Pricing." Doctoral thesis at Carleton University, Ottawa, Ontario, May 12,2000.

[3] E.D. Fitkov-Norris, and A. Khanifar, "Dynamic Pricing in Mobile Communication Systems", First International Conference on 3G Mobile Communication Technologies (IEE Conf. Publ. No. 471), 2000, pp. 416 –420.

[4] P.C. Fishburn, A.M. Odlyzko, "Dynamic Behavior of Differential Pricing and Quality of Service options for the Internet", ICE'98, pp. 128-139

[5] J. Hou, J. Yang, and S. Papavassiliou, "Integration of Pricing with Call Admission Control for Wireless Networks" Vehicular Technology Conference, 2001, Volume: 3, 2001 pp. 1344 -1348.

[6] V.A. Siris, B. Briscoe, and D. Songhurst, "Economic Models for Resource Control in Wireless Networks" Proceedings of IEEE PIMRC 2002, Lisbon, Portugal, SEP 15-18, 2002

[7] E. Viterbo, and C.F. Chiasserini, "Dynamic Pricing for connection-oriented services in wireless networks" 12th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications 2001, Volume: 1, Sep 2001 pp. A-68 -A-72.

[8] S.D. Patek, and E. Campos-Nanez, "Pricing of dialup services: an example of congestion-dependent pricing in the Internet", Proceedings of the 39th IEEE Conference on Decision and Control, 2000, Volume: 3, 2000 pp. 2296 - 2301.

[9] C. Courcoubetis and R. Weber, "Pricing Communication Networks: Economics, Technology and Modeling" Wiley, 2003.

[10] Y. B. Lin, S. Mohan and A. Noerepel, "Queuing Priority Channel Assignment Strategies for PCS Hand-off and Initial Access" IEEE Transactions on Vehicular Technology, Vol. 43, No.3, August 1994.

[11] A. Srivastanva, "Dynamic Pricing Models: Opportunity for Action" December, 2001

[12] I.C. Paschalidis, and J.N. Tsitsildis, "Congestion-dependent pricing of network services," IEEE/ACM Transactions on Networking, Vol.8, No.2, April 2000, pp. 416-420

[13] R. Cocchi, D. Estrin, S. Shenker, and L. Zhang." A study of priority pricing in multiclass networks", Proc. SIGCOMM'91, Sept. 1991

[14] T. Heikkinen, "Quality of service and pricing in the wireless internet", Proc. ITC, March 2000, pp. 39-50

[15] L. Kleinrock, "Queueing Systems", Vol.1, Wiley, 1975

[16] Cellular News, "Call for improved cell phone service during emergencies" [online], Available: http://www.cellular-news.com/story/9590.shtml