A New Stochastic Admission Control Scheme for Wireless Networks

Xinbing Wang, Zhuo Chen, Youyun Xu
Department of Electronic Engineering
Shanghai Jiaotong University, China
Email: {xwang8, chenzhuo, xuyouyun}@sjtu.edu.cn

Ruhai Wang
Department of Electrical Engineering
Lammar University
Email: wang@ee.lamar.edu

***Abstract*— In traditional call admission control (CAC) schemes, mobile users are always the passive roles during the admission procedures and the base station determines whether to admit or reject the call requests without the involvement of mobile users. In this paper, we propose a novel stochastic CAC framework that allows the mobile user to be an active entity during the CAC process. The objective of each mobile user is to maximize its utility function based on its own decision (i.e., to join the queue or not to join the queue). The optimal stochastic decision is in the sense that the probability for a mobile user to join the queue will maximize the expectation of the utility function. In other words, in the long run, the mobile user will benefit from the optimal joining probability p^*. We further show that the optimal join probability for a mobile user could be quite different depending on the number of mobile users waiting in the queue. Finally, we illustrate the structure of the optimal join probability p^* under various utility functions, which indicates that the optimal user's policy depends on the utility functions. Moreover, we find that even if the theoretical structure of service time in the system are quite different for various queue scheduling schemes, the optimal p_i^* are identical in most of circumstance.**

I. Introduction

Call admission control (CAC) investigates the issue of how to efficiently utilize the limited resource of wireless bandwidth. With the rapid evolvement of wireless technology and increasing demand of the mobile users, the CAC problem is gaining more and more attention.

Extensive research efforts have been conducted to investigate CAC issues and reveal plenty of insights [1], [2], [3], [4], [5], [6], [7], [8], [9], [10]. In [6], user mobility is considered by the CAC scheme with arbitrary call-arrival rate, which requires the global information on the call arrival process and user mobilities. The proposed scheme is analyzed in terms of a more realistic network topology and their results show that the CAC algorithm achieves high throughput with guaranteed Quality of Service (QoS) and call blocking probabilities. In [7], the mathematical model for on/off traffic model is obtained for cellular networks provisioning voice and best-effort data service. The traffic is characterized by a three-dimensional birth-death model that effectively captures the complicated interaction between on/off voice traffic and best-effort data traffic under a completely sharing situation. The close form for the performance metrics are derived and the minimum amount of resources to guarantee the required QoS is obtained.

A common feature shared by the existing literature is that the mobile subscribers always are the passive roles during the CAC procedures. The mobile users are either admitted to the system or denied by the system. They do not have any control or choice during such an admission control process. However, with the rapid development of CPU and DSP chip, the modern mobile terminals, like PDA, smart phone and handset etc, become more and more powerful in terms of computational capacity. In addition, mobile users have more information like their location and service requirements than the base station has. It is desirable for the system to fully utilize the mobile node's capability to improve the admission control's performance from CAC perspective.

Towards this direction, we propose a new stochastic admission control paradigm that makes mobile users actively participate into the CAC procedure. The new scheme allows the mobile users to decide whether to join the system or not, but in a probability or stochastic sense rather than a deterministic sense. For example, if the probability of the mobile users joining the system (say 'p') is greater than 0.5, then the mobile users are more willing to join the system. For 'p' is smaller than 0.5, it suggests that the mobile users are more reluctant to join the system. We identify the optimal strategy for the mobile users in sense that the optimal portfolios will maximize the mobile users's utility. In other words, without following the suggested optimal strategy, the mobile users only reduce their own utilities.

In addition, call blocking or dropping probability are not always good performance metrics for all the situations. For example, if the call waiting time is too long, some users will run of patient and leave the system. In such a case, even though the calls are not blocked, the users are not satisfactory with the admission service. Hence, call blocking probability is not a good signal to reflect the users' requirements in certain circumstance. Instead, we propose the notion of utility function for the CAC system. Utility functions are widely deployed to formulate the problems in the fields like Internet congestion [11], [12], game theory [13], [14], supplying chain and inventory theory [15], [16], etc. In this paper, We define the utility function for each user to capture the gain and loss during the CAC process. The objective of each mobile user is to maximize its own utility function. We show that for different utility functions, the optimal join probability, 'p' is quite different and even for the same utility function, 'p' could be very different as well depending on the number of mobile users in the queue. We explore certain structures of 'p' under various situations, which suggest that simply joining or not joining the queue is not optimal in most of circumstances. More interestingly, we find that even if the theoretical structure of service time are quite different according to the queue scheduling scheme, the optimal joining probability p_i^* are identical.

The rest of paper is organized as follows. We describe the system model in Section II and present our stochastic CAC scheme in Section III. In Section IV, we illustrate how to optimize the utility function and display the structure of optimal join probability according to different queue size. We provide numerical results in Section V and this paper is concluded in Section VI.

II. System Model

We consider a single cell with capacity of C channels and buffer size of n. Each mobile user requests one channel for service. The call arrival follows Poisson process with rate λ and the channel holding time is exponentially distributed with rate μ as in [17], [2]. Once there is a call arrival to the base station (BS) and there is one channel available, the call request will be served immediately. If no resource left at that time, the call will be put into the queue to wait until it can get the service. Since the total queue size is finite n, when the queue is full, the coming request will be blocked.

Note that although we present a very simple model without multiple classes of services and handoff consideration, etc, the idea in this paper can be extended to the general settings with more cumbersome illustration. The purpose of using such a simple model is to reduce the complexity of queueing analysis and highlight the idea of stochastic admission control based on utility function.

III. Stochastic CAC Scheme

In this section, we describe how our CAC scheme admits a call request stochastically. When there is a new call request arriving at the BS, if the BS has the available channels, it will allocate one of the channels to the call. If all the channels are occupied by the ongoing calls, the BS will notify the mobile user that currently there is no resource available and the new call could be put into the queue (if the queue is not full yet) to wait until the channel is available. Up to here, the our CAC scheme is indifferent from the existing ones, and the difference is the following. After receive such a notification, the mobile user will decide whether to join the queue or not. If the mobile user decide to join the queue, then the BS will put it into the queue; if not, the mobile user simply leaves the system and the BS discards the request. The pseudocode of the admission control procedures is displayed in Figure 1.

Algorithm 1: Stochastic Admission Control

Input: Acall arrival '**M**'; remaining number of channels '**C**'; the remaining buffer capacity '**B**'. optimal join probability p_i for **M**.

Output: An admission decision of a BS.

Method:

```
A call M arrives at the BS
IF C > 0
    Accept the call
    C ← C − 1
ELSE
    Notify M that there is no available channel and ask
    M whether it wants to join the queue to wait or not
    M will generate a uniform random variable x ∈ [0, 1]
    IF M wants to join the queue (i.e., x > p_i)
        IF B > 0
            put the call into the queue
            B ← B − 1
        ELSE
            Reject the call
        END
    ELSE (i.e., x ≤ p_i)
    M withdraws the call request and leaves the system
    END
END
```

Fig. 1

THE STOCHASTIC ADMISSION CONTROL SCHEME-PART I: ADMISSION CONTROL.

Whether a mobile user joins the queue or not is based on by its utility function. The objective of each mobile user is to maximize its utility function. In other words, if join the queue tends to increase the utility, the mobile user will join. If not, the mobile user will avoid joining. Since the system is random in that the arrival is random and service is random as well, the utility function is defined through the expectation of certain random variables as shown in Section IV.B. Thus, the decision of joining the queue is not deterministic as well, rather it is a stochastic decision. We therefore denote p as the probability to join the queue. For example, $p = 0.8$ means that on average, the mobile user will join the queue 8 times out of 10 times (not join the queue for 2 times). If $p = 0.2$, the mobile user is more reluctant to join the queue compared with $p = 0.8$. As a special case, for $p = 1$ ($p = 0$), the mobile user will always join (not join) the queue and hence this degenerates to the deterministic case.

Moreover, we found out depending on the utility function, the join probability p could be very different for different number of mobile users in the queue. We further denote p_i as the joining probability for a mobile user when the system has $i-1$ users waiting in the queue already. The BS calculates p_i, for $i = 1, ..., n$, which maximize the mobile user's utility in Section IV.B.

We point out that if the mobile user does not follow the join probability p_i, it will only hurt its performance in the long run, because p_i is the optimal join probability for the mobile user as shown in the next section.

IV. Optimize Utility Function

In this section, we present the definition of utility function and illustrate how to maximize the utility function under different situations.

A. Definition of Utility Function

Denote $\vec{p} = (p_1, p_2, ..., p_n)$ as the probability vector, where p_i is the probability for a mobile user to join the queue when there are already $i-1$ mobile users waiting in the queue. Let t be the average service time of the mobile user. Assume a mobile user will gain a constant 'α' as the benefit of completing this call, and pay $\beta(t)$ as the cost, which is the function of waiting time t and monotone non-decreasing in t, because from the mobile user's point of view, the more waiting time it experiences, the more cost for it pays. Moreover, let Q be number of mobile users waiting in the queue.

We define the utility function $U(t, \vec{p})$ as the expected revenue for a mobile user who arrives at the base station and has to wait in the queue [1]. For a given queue length $Q = i, i \geq 0$, if a mobile user joins the queue, the revenue for the mobile user will be $\alpha - \beta(t_i, p_i)$. The mobile user joins the queue with probability p_i, and does not join with probability of $1 - p_i$ (obtains zero revenue.) Thus, the expected revenue for a mobile user can be stated as

$$\begin{aligned} U(t, \vec{p}) &= \mathbb{E}\{\mathbb{E}\{\ \alpha - \beta(t_i) \mid Q > 0\ \}\} \\ &= \sum_{i=1}^{n} P_i \mathbb{E}\{\ \alpha - \beta(t_i) \mid Q = i\ \} \\ &= \sum_{i=1}^{n} P_i \{p_i \times [\alpha - \beta(t_i)] + (1 - p_i) \times 0\} \\ &= \sum_{i=1}^{n} P_i p_i \left[\alpha - \beta(t_i)\right], \end{aligned} \tag{1}$$

where P_i is the conditional probability $\mathbf{Prob}\{Q = i-1 \mid Q \geq 0\}$ that is the probability of $i-1$ mobile users waiting in the buffer, given the buffer is not empty ($Q \geq 0$) for $i = 1, ..., n$.

Thus, the objective of a mobile user is to maximize the utility by controlling the probability of joining the queue, $\vec{p} = (p_1, p_2, ..., p_n)$. That is,

$$\begin{aligned} \max \quad & U(t, \vec{p}) \\ \text{s.t.} \quad & 0 \preceq \vec{p} \preceq 1 \end{aligned} \tag{2}$$

TABLE I

LIST OF NOTATIONS

Notation	Representation
λ	Average arrival rate
μ	Average service rate
C	Number of channel in the system
n	Queue size
p_i	Join probability for a mobile user seeing queue length of i
P_i	Conditional probability that there are $i-1$ mobile users in the queue given the queue is not empty
Q	Number of mobile users in the queue
α	Benefit for a mobile user to complete a call
$\beta(t)$	Cost function for a mobile user to obtain the service

[1] We do not consider the situation that a mobile user arrives at the system and the base station has available channels, because it can get service right away without waiting.

B. Queueing Analysis

In order to further illustrate $U(t, \vec{p})$ in (1), we proceed queueing analysis to acquire P_i and the average waiting time t_i. We can characterize the system dynamics by a continuous time Markov Chain and state the balance equation as follows.

$$\begin{aligned} \pi_{i+1} &= \frac{\lambda}{(i+1)\mu} \pi_i, \quad 0 \leq i < C, \\ \pi_{i+1} &= \frac{\lambda \times p_{i-C}}{C\mu} \pi_i, \quad C \leq i < C + n, \end{aligned}$$

After simple calculations, we get

$$\begin{aligned} \pi_1 &= \frac{\lambda}{\mu} \pi_0, \quad \pi_2 = \frac{\lambda}{\mu} \frac{\lambda}{2\mu} \pi_0, \\ \cdots &\quad \cdots \\ \pi_C &= \left(\frac{\lambda}{\mu}\right)^C \frac{1}{C!} \pi_0, \\ \cdots &\quad \cdots \\ \pi_{C+i} &= \left(\frac{\lambda}{\mu}\right)^{C+i} \frac{\prod_{j=1}^{i} p_j}{C^i} \frac{1}{C!} \pi_0, \\ \cdots &\quad \cdots \\ \pi_{C+n} &= \left(\frac{\lambda}{\mu}\right)^{C+n} \frac{\prod_{j=1}^{n} p_j}{C^n} \frac{1}{C!} \pi_0, \end{aligned}$$

In addition, we know that

$$\pi_0 + \pi_1 + \cdots + \pi_{C+n} = 1.$$

We have π_{C+i} in terms of $\vec{p} = (p_1, ..., p_n)$ as

$$\pi_{C+i} = \frac{\left(\frac{\lambda}{\mu}\right)^{C+i} \frac{\prod_{j=1}^{i} p_j}{C^i} \frac{1}{C!}}{\left[\sum_{k=0}^{C} \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!} + \frac{1}{C!} \sum_{k=1}^{n} \left(\frac{\lambda}{\mu}\right)^{k+C} \frac{\prod_{i=1}^{k} p_i}{C^k}\right]}. \tag{3}$$

for $i = 1, 2..., n-1$.

Now, we are ready to calculate the conditional probability P_i by

$$\begin{aligned} P_i &= \mathbf{Prob}\{Q = i \mid Q \geq 0\} \\ &= \frac{\mathbf{Prob}\{Q = i, Q \geq 0\}}{\mathbf{Prob}\{Q \geq 0\}} \\ &= \frac{\pi_{C+i-1}}{\sum_{j=1}^{n} \pi_{C+j-1}}, \quad i = 1, 2..., n. \end{aligned}$$

By substituting π_{C+i} in (3), we have

$$P_i = \frac{\left(\frac{\lambda}{\mu}\right)^{C+i-1} C^{1-i} \prod_{j=1}^{i-1} p_j}{\sum_{k=1}^{n} \left(\frac{\lambda}{\mu}\right)^{C+k-1} C^{1-k} \prod_{j=1}^{k-1} p_j}, \quad i = 1, ..., n. \tag{4}$$

C. t_i in FIFO and LIFO Queue

In this subsection, we present how to obtain t_i. We further divide the discussion into two cases: (i) first-in-first-out queue (FIFO) and (ii) last-in-first-out queue (LIFO). In the long run, it is fair to each mobile user for either case, while two typical queue scheduling schemes produces different structures of t_i.

For FIFO queue, the structure of t_i is straightforward. Since the service time is exponentially distributed, the average waiting time given i users already in the queue is

$$t_i^{FIFO} = 1/\mu + i/C\mu = (C+i)/C\mu. \quad (5)$$

As to LIFO queue, the structure of t_i is more complicated than that of FIFO queue.

$$\begin{aligned}\tau_{i,1} &= \frac{\lambda p_{i+1}}{\lambda p_{i+1} + C\mu} \\ \tau_{i,2} &= \frac{C\mu}{\lambda p_{i+1} + C\mu}\end{aligned}$$

where P_1 is the probability that the coming event is an arrival and P_2 is the probability that the coming event is a departure.

Theorem 1: Assume user i is the last one in the queue currently and given the next event to the system is an arrival, i.e., user $i+1$, instead of a departure from the system. The average waiting time in the queue for user i just **before** the user $i+1$ joins the queue is $\frac{1}{\lambda + C\mu}$.

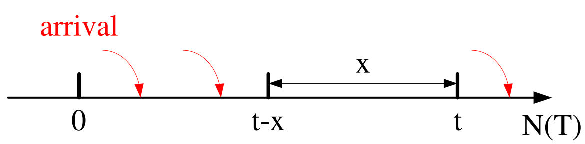

Fig. 2
THE ARRIVAL PROCESS.

Proof: Since the service time for each server is exponential distribution with average $\frac{1}{\mu}$, the service time for total C servers is exponential distribution with average $\frac{1}{\lambda+C\mu}$. In other words, the departure process from the system with C servers is Poisson process $N(t)$ with rate $\lambda + C\mu$. Suppose t is large enough and γ_t is past waiting time for user i, which is the last one in the queue and the first one to be served, if one departure occurs. Then, $\gamma_t > x$ if and only if there is no departure from time interval $[t-x, x]$, i.e., $N(t) - N(t-x) = 0$. Hence the distribution function of γ_t is given by

$$\begin{aligned}F_{\gamma_t}(x) &= \mathbf{Prob}(\gamma_t \le x) \\ &= 1 - \mathbf{Prob}(\gamma_t > x) \\ &= 1 - \mathbf{Prob}(N(t) - N(t-x) = 0) \\ &= 1 - e^{-(\lambda+C\mu)x}\frac{[(\lambda+C\mu)x]^0}{0!} \\ &= 1 - e^{-(\lambda+C\mu)x}\end{aligned}$$

$$\begin{aligned}E\{\gamma_t\} &= \int_0^\infty P(\gamma_t > x)dx \\ &= \int_0^\infty e^{-(\lambda+C\mu)x}dx \\ &= \frac{1}{\lambda + C\mu}\end{aligned}$$

■

Theorem 2: The time spent in the system for i^{th} arrival has the Erlang distribution given by

$$t_i^{LIFO} = \frac{1}{C\mu}\prod_{j=i}^{n-1}\frac{\tau_{j,1}}{\tau_{j,2}} + \sum_{j=i}^{n-1}\frac{1}{\tau_{j,2}}\frac{1}{(\lambda p_{j+1} + C\mu)}\prod_{k=i}^{j-1}\frac{\tau_{k,1}}{\tau_{k,2}}. \quad (6)$$

Proof: We know that the total time for user i spent in the system, t_i consists of two parts: (i) the waiting time in the queue, W_i and (ii) the service time in the server. That is,

$$t_i^{LIFO} = W_i + \frac{1}{\mu},$$

where $\frac{1}{\mu}$ is the average service in the server.

The average waiting time for the last user n in the queue W_n (n is the queue size) has the highest priority to join the service. So long as the system has one departure, user n will join the server, and hence the average waiting time will be $\frac{1}{C\mu}$. While all other users i ($i = 1, 2, ..., n-1$) has two possibilities

The average waiting time W_i can be stated in the recursive fashion.

$$\begin{aligned}W_n &= \frac{1}{C\mu} \\ W_{n-1} &= \tau_2\frac{1}{\lambda p_n + C\mu} + \tau_1(\frac{1}{\lambda p_n + C\mu} + W_n + W_{n-1}) \\ ... &= ... \\ W_2 &= \tau_2\frac{1}{\lambda p_3 + C\mu} + \tau_1(\frac{1}{\lambda p_3 + C\mu} + W_3 + W_2) \\ W_1 &= \tau_2\frac{1}{\lambda p_2 + C\mu} + \tau_1(\frac{1}{\lambda p_2 + C\mu} + W_2 + W_1)\end{aligned}$$

Then, after simple algebra calculation, we can have

$$\begin{aligned}W_i - \frac{\tau_{i,1}}{\tau_{i,2}}W_{i+1} &= \frac{1}{\tau_{i,2}(\lambda p_{i+1} + C\mu)} \\ \frac{\tau_{i,1}}{\tau_{i,2}}W_{i+1} - \frac{\tau_{i,1}\tau_{i+1,1}}{\tau_{i,2}\tau_{i+1,2}}W_{i+2} &= \frac{1}{\tau_{i+1,2}(\lambda p_{i+2} + C\mu)}\frac{\tau_{i,1}}{\tau_{i,2}} \\ ... &= ... \\ \prod_{j=i}^{n-3}\frac{\tau_{j,1}}{\tau_{j,2}}W_{n-2} - \prod_{j=i}^{n-2}\frac{\tau_{j,1}}{\tau_{j,2}}W_{n-1} &= \frac{1}{\tau_{n-2,2}(\lambda p_{n-1} + C\mu)}\prod_{j=i}^{n-3}\frac{\tau_{j,1}}{\tau_{j,2}} \\ \prod_{j=i}^{n-2}\frac{\tau_{j,1}}{\tau_{j,2}}W_{n-1} - \prod_{j=i}^{n-1}\frac{\tau_{j,1}}{\tau_{j,2}}W_n &= \frac{1}{\tau_{n-1,2}(\lambda p_n + C\mu)}\prod_{j=i}^{n-2}\frac{\tau_{j,1}}{\tau_{j,2}},\end{aligned}$$

By taking the summation over all the above equations, we have

$$W_i - \prod_{j=i}^{n-1}\frac{\tau_{j,1}}{\tau_{j,2}}W_n = \sum_{j=i}^{n-1}\frac{1}{\tau_{j,2}}\frac{1}{(\lambda p_{j+1} + C\mu)}\prod_{k=i}^{j-1}\frac{\tau_{k,1}}{\tau_{k,2}},$$

Then,

$$W_i = W_n \prod_{j=i}^{n-1} \frac{\tau_{j,1}}{\tau_{j,2}} + \sum_{j=i}^{n-1} \frac{1}{\tau_{j,2}} \frac{1}{(\lambda p_{j+1} + C\mu)} \prod_{k=i}^{j-1} \frac{\tau_{k,1}}{\tau_{k,2}}.$$

With $W_n = \frac{1}{C\mu}$, the result follows. ■

With P_i and t_i (t_i^{FIFO} and t_i^{LIFO}, respectively), we can substitute them into Eqn. (1) to calculate the utility function.

V. Numerical Results

In general, the utility function in (1) is not convex, nor concave function in most situations. We have to rely on the numerical methods to obtain the optimal solutions. Here, we explore some properties of the utility function to know the certain structure of the optimal solution.

We present three cases with different utility functions and use generic algorithm to search the optimal solutions. Since the solution is in a multi-dimensional Euclidian space, we cannot plot them in one figure. We thus list some typical values for $\vec{p}$, present the optimal solution p_i^* with bold fonts for the comparison purpose and explain the implications of the optimal solutions for both FIFO and LIFO queues. In addition, we provide the optimal t_i^* (for FIFO and LIFO, respectively) corresponding to p_i^*. Although the underlying rationale is still not clear, we report the interesting observation that even if the theoretical structure of t_i^{FIFO} and t_i^{LIFO} are quite different, the optimal p_i^* are identical in most of circumstance, which is robust against queue scheduling scheme.

Case 1: The system parameters are configured as $C = 20, n = 6$, $\lambda = 20$, $\mu = 1$, $\alpha = 30, \beta(t) = 2t_i$, and thus the utility function is given by $U(t,\vec{p}) = \sum_{i=1}^{6} P_i p_i(30 - 2t_i)$. We are interested in the optimal $\vec{p}$ that maximize the utility function, i.e., $\vec{p}^* = \arg\max_{0 \preceq \vec{p} \preceq 1} \sum_{i=1}^{5} P_i p_i(30-2t_i)$. The numerical results for both t_i^{FIFO} and t_i^{LIFO} are $\vec{p}^* = (1,1,1,1,1,1)$. The resulting $\vec{p}^*$ is so called "join dominant". In other words, in any situations, the mobile user prefers to join the queue with probability 1, rather than leave the system, and the expected utility for the mobile user is maximized to 27.65 compared to other strategies displayed in Tables II and III.

TABLE II
Case 1: Performance Comparison of FIFO Queue.

p_1	p_2	p_3	p_4	p_5	p_6	Utility
0.0	0.0	0.0	0.0	0.0	0.0	0.00
0.2	0.2	0.2	0.2	0.2	0.2	5.57
0.4	0.4	0.4	0.4	0.4	0.4	11.13
0.6	0.6	0.6	0.6	0.6	0.6	16.66
0.8	0.8	0.8	0.8	0.8	0.8	22.17
1.0	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**27.65**
t_1^*	t_2^*	t_3^*	t_4^*	t_5^*	t_6^*	
1.05	1.10	1.15	1.20	1.25	1.30	

Case 2: The system parameters are set as $C = 20, n = 6$, $\lambda = 20$, $\mu = 1$, $\alpha = 10, \beta(t) = 10t$, and the utility function is given by $U(t,\vec{p}) = \sum_{i=1}^{5} P_i p_i(10 - 10t_i)$. We can obtain $\vec{p}^* = \arg\max_{0 \preceq \vec{p} \preceq 1} U(t,\vec{p}) = (0,0,0,0,0,0)$. Such structure of $\vec{p}^*$ is so called "not join" dominant. The mobile users will not join the queue to wait no matter how many users wait in

TABLE III
Case 1: Performance Comparison of LIFO Queue.

p_1	p_2	p_3	p_4	p_5	p_6	Utility
0.0	0.0	0.0	0.0	0.0	0.0	0.00
0.2	0.2	0.2	0.2	0.2	0.2	5.57
0.4	0.4	0.4	0.4	0.4	0.4	11.13
0.6	0.6	0.6	0.6	0.6	0.6	16.66
0.8	0.8	0.8	0.8	0.8	0.8	22.17
1.0	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**27.65**
t_1^*	t_2^*	t_3^*	t_4^*	t_5^*	t_6^*	
1.30	1.25	1.20	1.15	1.10	1.05	

the queue [2] and leave the system immediately. In other words, to wait in the queue (no matter what is the queue length) will only reduce the mobile user's utility in the long run. We present the comparison in Tables IV and V.

TABLE IV
Case 2: Performance Comparison of FIFO Queue.

p_1	p_2	p_3	p_4	p_5	p_6	Utility
0.0	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0000**
0.2	0.2	0.2	0.2	0.2	0.2	-0.125
0.4	0.4	0.4	0.4	0.4	0.4	-0.328
0.6	0.6	0.6	0.6	0.6	0.6	-0.661
0.8	0.8	0.8	0.8	0.8	0.8	-1.147
1.0	1.0	1.0	1.0	1.0	1.0	-1.750
t_1^*	t_2^*	t_3^*	t_4^*	t_5^*	t_6^*	
0	0	0	0	0	0	

TABLE V
Case 2: Performance Comparison of LIFO Queue.

p_1	p_2	p_3	p_4	p_5	p_6	Utility
0.0	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0000**
0.2	0.2	0.2	0.2	0.2	0.2	-0.125
0.4	0.4	0.4	0.4	0.4	0.4	-0.328
0.6	0.6	0.6	0.6	0.6	0.6	-0.661
0.8	0.8	0.8	0.8	0.8	0.8	-1.147
1.0	1.0	1.0	1.0	1.0	1.0	-1.750
t_1^*	t_2^*	t_3^*	t_4^*	t_5^*	t_6^*	
0	0	0	0	0	0	

Case 3: In both previous cases, the mobile user will join with either probability 1 or probability 0. In this case, we choose utility function as $U(t,\vec{p}) = \sum_{i=1}^{6} P_i p_i[30 - 2p_i(t_i + i)^2]$ with $C = 20, n = 6$, $\lambda = 20$, $\mu = 1$, $\alpha = 30, \beta = 2p_i \times (t_i + i)^2$. We obtain numerically the optimal solution given by $\vec{p}^* = \arg\max_{0 \preceq \vec{p} \preceq 1} U(t,\vec{p}) = (1, 0.489, 0.265, 0.174, 0.109, 0.125)$. When the system does not have any available channels, the best strategy for the mobile user depends on the number of users waiting in the queue. When there is nobody in the queue, the mobile user will join the queue with probability 1; When there is one use waiting in the queue, then the mobile user will join with probability 0.489 rather than simply join or simply leave. In this situation, the mobile user will join the queue in the

[2] Even there is nobody in the queue, the mobile user may prefer not to wait in the queue, if it cannot get service right away.

stochastic sense, not in the deterministic sense. Intuitively, on average, for 1000 times, it will join 489 times and not join 511 times. So on and so forth for p_3, p_4, p_5 and p_6. The interesting observation is that given the utility function as in the this case, p_i does not monotone decrease in i. For example, p_5 is less than p_6, and this shows that the mobile user is more tentative to join a queue with a larger queue size (e.g., 5 users in the queue) than a smaller queue size (e.g., 4 users in the queue). In addition, we can clearly see that the optimal p_i^* are identical for both FIFO and LIFO queue scheduling, although t_i^* of FIFO and LIFO are different from each other. We also examine the other combinations of the system parameters and the same observations hold. We provide the detailed data in Tables VI and VII.

TABLE VI

CASE 3: PERFORMANCE COMPARISON OF FIFO QUEUE.

p_1	p_2	p_3	p_4	p_5	p_6	Utility
0.0	0.0	0.0	0.0	0.0	0.0	0.00
0.2	0.2	0.2	0.2	0.2	0.2	5.54
0.4	0.4	0.4	0.4	0.4	0.4	9.28
0.6	0.6	0.6	0.6	0.6	0.6	8.56
0.8	0.8	0.8	0.8	0.8	0.8	−0.36
1.0	1.0	1.0	1.0	1.0	1.0	−20.1
1.0	**0.489**	**0.265**	**0.174**	**0.109**	**0.125**	**13.20**
t_1^*	t_2^*	t_3^*	t_4^*	t_5^*	t_6^*	
1.05	1.10	1.15	1.20	1.25	1.30	

TABLE VII

CASE 3: PERFORMANCE COMPARISON OF LIFO QUEUE.

p_1	p_2	p_3	p_4	p_5	p_6	Utility
0.0	0.0	0.0	0.0	0.0	0.0	0.00
0.2	0.2	0.2	0.2	0.2	0.2	5.54
0.4	0.4	0.4	0.4	0.4	0.4	9.31
0.6	0.6	0.6	0.6	0.6	0.6	8.73
0.8	0.8	0.8	0.8	0.8	0.8	0.16
1.0	1.0	1.0	1.0	1.0	1.0	−18.9
1.0	**0.489**	**0.265**	**0.174**	**0.109**	**0.125**	**13.16**
t_1^*	t_2^*	t_3^*	t_4^*	t_5^*	t_6^*	
1.0825	1.0646	1.0593	1.0564	1.0573	1.05	

Traditional CAC can be interpreted as an "always join" scheme until the queue is full. In our analysis, we can see that in most of cases "always join" strategy may not be the optimal and the optimal strategy depends on the utility functions of the mobile users.

VI. CONCLUSION

In traditional CAC schemes, mobile users always become a passive party during the admission procedures and the BS decides whether to admit or reject the call request without the involvement of the mobile users. In this paper, we propose a novel stochastic CAC scheme that allows the mobile user to be an active role during the CAC process to fully utilize the mobile users' capability and hence improve the system performance. The objective of each mobile user is to maximize the utility function by its own decision (to join the queue or not to join the queue). The optimal stochastic decision is in the sense that the probability for a mobile user to join the queue will maximize the expectation of the utility function. We further show that the optimal join probability for a mobile user could be quite different, when it observers different number of mobile users waiting in the queue. Finally, we illustrate the structure of the optimal join probability $\vec{p}*$ under various utility functions, which indicates that simply to join the queue all the time is not optimal in most situations. As the future work, we may (i) consider incorporate handoff price into the utility function and analyze the impact of handoff; (ii) explore the underlying reason for the same optimal p_i^* structure against the queue scheduling mechanism.

REFERENCES

[1] D. Zhao, X. Shen, and J. W. Mark, "Soft handoff and connection reliability in cellular CDMA networks," *IEEE Transactions on Wireless Communications*, vol. 5, pp. 354–365, Feb 2006.

[2] Y. Fang and Y. Zhang, "Call Admission Control Scheme and Performance Analysis in Wireless Mobile Networks," *IEEE Transaction on Vehic. Technology*, vol. 51, pp. 371–382, March 2002.

[3] J. Hou and Y. Fang, "Mobility-based call admission control schemes for wireless mobile network," *Wireless Communications and Mobile Computing*, vol. 1, pp. 269–282, July-Sept 2001.

[4] M. Ghaderi and R. Boutaba, "Call Admission Control for Voice/Data Integration in Broadband Wireless Networks," *IEEE Transactions on Mobile Computing*, vol. 5, pp. 193–207, March 2006.

[5] W. Wang, X. Wang, and A. Nilsson, "Energy-Efficient Bandwidth Allocation in Wireless Networks: Algorithms, Analysis, and Simulation," in *IEEE Transactions on Wireless Communications*, vol. 5, pp. 1103–1114, April 2006.

[6] R. G. Akl, M. V. Hegde, and M. Naraghi-Pour, "Mobility-Based CAC Algorithm for Arbitrary Call-Arrival Rates in CDMA Celluar Systems," *IEEE Transactions on Vehicular Technology*, vol. 54, pp. 639–651, March 2005.

[7] C. W. Leong, W. Zhuang, Y. Cheng, and L. Wang, "Call Admission Control for Integrated On/Off Voice and Best-Effort Data Services in Mobile Celluar Communications," *IEEE Transactions on Communications*, vol. 52, pp. 778–790, May 2004.

[8] D. G. Jeong and W. S. Jeon, "CDMA/TDD System for Wireless Multimedia Services with Traffic Unbalance Between Uplink and Downlink," *IEEE Journal On Selected Areas in Communications*, vol. 17, pp. 939–946, May 1999.

[9] M. H. Ahmed and H. Yanikomeroglu, "SINR Threshold Lower Bound for SINR-based Call Admission Control in CDMA Networks with Imperfect Power Control," *IEEE Communications Letters*, vol. 9, pp. 331–333, April 2005.

[10] D. Zhang, X. Shen, and J. W. Mark, "Radio Resource Management for Cellular CDMA Systems Supporting Heterogeneous Services," *IEEE Transactions on Mobile Computing*, vol. 11, pp. 32–39, April-June 2003.

[11] S. H. Low, L. Peterson, and L. Wang, "Understanding Vegas: a Dulity Model," *Journal of ACM*, vol. 49, pp. 207–235, Mar 2002.

[12] J. Wang, L. Li, S. H. Low, and J. C. Doyle, "Cross-layer Optimization in TCP/IP Networks ," *IEEE/ACM Trans. on Networking*, vol. 13, pp. 582–596, June 2005.

[13] T. Basar and S. Srikant, "A Stackelberg network game with a large number of followers," *Journal of Optimization Theory and Applications*, vol. 115, pp. 479–490, Dec 2002.

[14] T. Alpcan, T. Basar, R. Srikant, and E. Altman, "CDMA Uplink power control as a noncooperative game," *Wireless Networks*, vol. 8, pp. 659–670, 2002.

[15] D. Shabtay and M. Kaspi, "Minimizing the makespan in open-shop scheduling problems with a convex resource consumption function ," *Naval Research Logistics, Publish online*, Feb 2006.

[16] Y. Chen, S. Ray, and Y. Song, "Optimal pricing and inventory control policy in periodic-review systems with fixed ordering cost and lost sales ," *Naval Research Logistics*, vol. 53, pp. 117–136, March 2006.

[17] W. Wang, X. Wang, and A. Nilsson, "A New Admission Control Scheme under Energy and QoS Constraints for Wireless Networks," in *Proceedings of IEEE Infocom'05, Miami, FL*, March 2005.