A New Admission Control Scheme under Energy and QoS Constraints for Wireless Networks

Wenye Wang Xinbing Wang Arne Nilsson
Department of Electrical and Computer Engineering
North Carolina State University, Raleigh, NC 27695

Abstract — **Battery capacity of mobile terminals and radio bandwidth are both limited and precious resources in wireless networks. In this paper, we present a thorough performance study of energy-based admission control scheme to make the best use of these two resources for effective mobile communications. In order to reduce energy consumption of each terminal, we introduce a** *Victim Selection Algorithm* **(VSA) and a** *Beneficiary Selection Algorithm (BSA)* **for acquiring bandwidth and releasing bandwidth, respectively. To avoid potential compromise of quality of service (QoS) due to the concern of energy consumption in connection admission control, we further propose an adjustment algorithm to statistically meet the demands for QoS. The performance of the proposed schemes is evaluated with respect to energy consumption rate of each successfully transmitted bit, throughput and call blocking probabilities for a variety of traffic such as Poisson and self-similar, multi-class traffic.**

Key Words: *Wireless networks, Quality of Service, energy consumption, and connection admission control.*

I. Introduction

Wireless networks have revealed a significant impact on the information and communication technology as more and more people are accustomed to depending on mobile devices for voice and data communications. However, a majority of mobile terminals operate on batteries, such as cellular phones, portable digital assistants (PDAs), and laptops in situations with no available power supply. Thus, energy efficiency is of particular interest in the design of wireless networks due to limited battery capacity. Along with the increasing trend of using mobile devices as a means of communication, the battery life of a mobile terminal becomes one of the bottlenecks to supporting high-quality multimedia services or huge data transmission, even affects roaming capability. The demand for universal wireless access, along with the development of wireless applications including location-aware services and mobile transactions, has motivated the research in supporting quality of service (QoS) with energy conservation in a variety of wireless networks.

Specifically, connection admission control (CAC) is one of the most important approaches to providing QoS parameters in wireless networks since radio bandwidth will be allocated for mobile communications. Many solutions have been proposed to dynamically allocate bandwidth based on different criteria, including system priority, capacity, mobility, and interference [6, 10, 13, 19, 24, 30]. With the common agreement that handoff requests should have higher priority than new arrivals because the termination of an existing connection is worse than not accepting a new request, guard channel schemes are proposed [6, 19, 21] in which handoff requests have higher priority than new incomings. To this end, the main design challenge is how to satisfy the minimum requirements of call dropping/blocking probabilities, whereas maintaining a high throughput.

Since the traffic and mobility patterns, especially for voice communications in cellular networks, can be represented by analytical models, prediction and mobility based approaches are further designed to reserve bandwidth in a specified location prior to the arrival of handoff requests [5, 7, 13, 24]. In these schemes, a base station (BS) may need to collaborate with other BSs or use only local information to make the decision for resource reservation. The tradeoff between carried traffic and QoS among multi-class traffic is pursued by using different methods such as neural networks and pricing models. The admission control schemes for code division multiple access (CDMA) systems, are mostly focused on the interference [2, 27], because the increased number of ongoing services in a CDMA system can bring signal-to-interference ratio (SIR) to an unacceptable level. Thus, power control is emphasized in many research efforts, which are mainly concerned with the improvement of system performance such as capacity and throughput [1, 2, 26].

Although battery capacity of mobile terminals and radio frequency bandwidth are both limited and precious resources in wireless networks, it is hardly considered in the CAC algorithms. At the physical layer, power management of a transmitter is studied extensively using the knowledge of channel fading and interference, thus dynamically changing transmission power [1, 3, 31]. For instance, joint source and channel coding with power control are exploited [8, 29] to prolong the battery life of mobile terminals. At the system level, it is suggested that energy consumption can be reduced by smartly turning off mobile devices during idle time [14]. Therefore, we will demonstrate that, besides power-control on uplinks and downlinks, CAC algorithms can also be designed to conserve energy consumption. In other words, total transmission energy consumption can be minimized by adapting transmission rate to channel conditions under the constraints of QoS [9, 22]. Motivated by these important work, in this paper, we develop a new admission control scheme in that it incorporates bandwidth allocation and energy conservation at the network layer.

Our contributions in this paper are as follows. First, we introduce a new parameter, *energy consumption rate*, which describes the relationship between connection admission control (CAC) and energy consumption. Then, we propose an energy-based connection admission scheme, including *victim selection algorithm* (VSA), *beneficiary selection algorithm* (BSA), and *stochastic adjustment algorithm* (SAA) for reducing energy consumption, while satisfying QoS requirements of multi-class traffic.

The rest of the paper is organized as follows. In Section II, we describe the system model and problem statement, in particular, the affect of admission control on energy consumption. Section III presents our algorithm in acquiring bandwidth from ongoing connections and re-allocating radio resources upon a connection termination for high throughput. The proposed solution is evaluated with regard to blocking probabilities, the effect of stochastic adjustment, and energy consumption for various scenarios in Section IV. Finally, simulation results for Poisson and self-similar traffic are provided in Section V, followed by conclusions in Section VI.

II. System Model and Problem Description

A. Multi-Class System Model

Unlike a single-class model, the multi-class system model is very important in analyzing performance of a wireless network because of the demand for various applications in mobile environments [5, 7, 11, 24]. In particular, resource allocation for multi-class users must consider the interaction between classes, e.g., the dropping probability of one class of users may impact that of another class of users. We consider a system model with multiple classes of traffic in which there are a total of C channels to serve K classes of services. Each class of service can be characterized by four parameters: channel requirements, arrival distribution, channel holding time, and QoS requirements, such as handoff dropping probabilities. Specifically, there parameters are defined as follows:

- C system capacity;
- K number of classes, $k = 1, 2 \ldots, K$;
- b_k^L lower bound of channel requirements in class k;
- b_k^U upper bound of channel requirements in class k;
- λ_k mean arrival rate of traffic in class k;
- μ_k mean service rate of traffic in class k;
- β_k requirement of new connections' blocking probability of services in class k;
- δ_k requirement of handoff dropping probability of services in class k;
- $\overrightarrow{\beta}$ a vector with predefined call blocking probability of each class, which means that the system needs to guarantee the blocking probability of class k being less than β_k, and
- $\overrightarrow{\delta}$ a vector with predefined handoff dropping probability of each class, which means that the system needs to guarantee the dropping probability of class k being less than δ_k.

In the general description of wireless systems, the bandwidth is denoted by the total frequency bandwidth used in a cell. On the other hand, logical channels are used for bandwidth allocation in TDMA/CDMA systems such as radio network controller (RNC) in Universal Mobile Telecommunications System (UMTS). To be consistent with the literature and specifications, we use the number of channels for bandwidth in this context. Given different types of services, the requirements of channels may be different. For example, multimedia traffic may require from 4 to 7 channels, even more, whereas data services may require from 2 to 5 channels. The definition a channel varies from system to system, but it is the minimum unit for voice service. For example, in an orthogonal frequency division multiplexing (OFDM) system, a channel can be considered as the minimum tones used to transmit voice traffic. If a service requires a fixed number of channels, it becomes a special case in our model because we can simply let $b_k^L = b_k^U =$ the number of channels. For simplicity of description, we assume the call-arrival process is the Poisson process, and the channel-holding time is exponentially distributed [24, 30]; therefore, we denote λ_k and μ_k as the mean arrival rate and service rate for class k, respectively. While multimedia traffic requires more channels, it may suffer a higher blocking probability because the request cannot be admitted if channels are not enough, resulting higher blocking/dropping probability.

B. Connection Admission Control and Energy Consumption

Connection admission control (CAC) is a technique that admits new requests and handles handoff connections intelligently in wireless systems to avoid network congestion and reduce blocking/dropping probabilities. The ultimate goal of CAC is to allocate bandwidth, that is, to determine the transmission rate of a mobile device while not degrading QoS parameters, which does not involve energy issue directly except the power control for reducing SIR. Nevertheless, transmission rate is closely related to the energy consumption per bit. In [4], it is explained that energy required to transmit a packet can be significantly reduced by lowering transmission power and transmitting the packet over a longer period of time; that is, by reducing transmission rate, energy consumption can be lowered. This is an important step to couple two problems together.In this paper, we describe how energy consumption is dependent on channel conditions, transmission time, and bandwidth allocation.

The energy consumption is determined by the total transmission time and corresponding transmission powers, which are associated with channel conditions, coding and modulation schemes. Let us denote W as the total bandwidth controlled by a BS and N as the number of terminals in a cell. For TDMA/CDMA systems, it is bandwidth that can be used by all terminals. For FDMA systems, it is the summation of each frequency channel. Let vector $\overrightarrow{\mathbf{P}} = [P_1, P_2, \ldots, P_i, \ldots, P_N]$ be the transmission power of N terminals, where P_i is the transmission power for terminal i. Similarly, we denote the transmission rate as vector $\overrightarrow{\mathbf{R}} = [R_1, R_2, \ldots, R_i, \ldots, R_N]$, where R_i is the transmission power of terminal i. The channel gain for each user is represented by a vector $\overrightarrow{\mathbf{H}} = [h_1, h_2, \ldots, h_i, \ldots, h_N]$. Then, the energy-to-noise ratio, E_b^t/N, of mobile user i can be written as [23]

$$\frac{E_b^t}{N_0} = \frac{E_b^t \cdot h_i}{\frac{\sum_{j \neq i} h_j P_j}{W} + \eta_0} \tag{1}$$

where E_b^t is transmission energy per bit and N_o is the noise spectral density including thermal noise and interference. η_0 is the white Gaussian noise level For existing wireless systems, we consider that bit error rate (BER), P_{ber}, for BPSK modulation as

$$P_{ber}(E_b^t) = Q(\sqrt{\frac{2E_b^t}{N_0}}) = \frac{2}{\sqrt{2\pi}} \int_{\sqrt{\frac{2E_b^t}{N_0}}}^{\infty} e^{-\frac{x^2}{2}} dx. \tag{2}$$

Since the payload of a service request will be encapsulated into frames during transmission, we need to know *frame error rate* (FER), which is determined by BER and coding schemes. In this context, we use *Reed-Solomon* (RS) code [29] as an example because of its considerable use in wireless systems. The RS coding scheme is represented by $RS(n, k)$, where k is the length of source symbols and $n - k$ is the length of protection symbols, which is able to correct up to $t = (n - k)/2$ symbol errors. Thus, *symbol error rate*, $P_s(E_b^t)$, and *frame error rate*, $P_{fer}(E_b^t)$, using $RS(n, k)$ are given by [23]

$$\begin{aligned} P_s(E_b^t) &= 1 - (1 - P_{ber}(E_b^t))^n \quad \text{and} \\ P_{fer}(E_b^t) &= \sum_{j=t+1}^{n} \binom{n}{j} P_s(E_b^t)^j (1 - P_s(E_b^t))^{n-j}. \end{aligned} \tag{3}$$

Then, the average number of transmissions, $\Omega(P_{fer}(E_b^t))$, can be obtained by

$$\Omega(P_{fer}(E_b^t)) = \frac{1}{1 - P_{fer}(E_b^t)}. \tag{4}$$

In addition, total transmission time for mobile terminal i, T_{total}^i, depends on the total amount of data, number of retransmissions, and transmission rate. That is

$$T_{total}^i = \frac{\phi_i}{R_i(1 - P_{fer}(E_b^t))} \tag{5}$$

where ϕ_i is the total amount of data to be transmitted for mobile terminal i. Thus, the *total energy consumption*, $E_{total}^i = \frac{E_b^t}{1-P_{fer}(E_b^t)} \cdot \phi_i$, which illustrates that the total energy consumption is dependent on data volume in transmission, energy per bit, as well as transmission errors. In particular, the transmission energy is proportional to the first item for fixed volume of information.

The ratio between the total energy consumption and the data volume is the energy consumption for each bit, which is defined as, *energy consumption rate* (ECR) $\Gamma(E_b^t)$,

Definition: *Energy consumption rate* is energy consumption of each successfully transmitted bit given by

$$\Gamma(E_b^t) = \frac{E_b^t}{1 - P_{fer}(E_b^t)}. \tag{6}$$

As an example, Fig. 1 shows energy consumption rate versus transmission rate, which is rewritten as $\Gamma_r(\bar{R}) = \frac{\bar{P}}{\bar{R} - \bar{R} \cdot P_{fer}(\bar{P}/\bar{R})}$ from the above definition. Other parameters include channel gain $h_i = 10^{-2}$ [17], $P = 20\ dBm$, and $N_o = 5 \times 10^{-9}\ WHz$ [23]. The similar curves hold true for other experiments such as $RS(128, 112)$, and $P = 20\ dBm$. This figure shows that transmission rate and energy consumption are closely related over erroneous channels. For instance, when transmission rate is lower than a threshold "A", about $120\ Kbps$, energy consumption will not decrease any more, but will increase dramatically as transmission rate decreases. Also, the variation of energy consumption is not the same for the same change in transmission rates. When we decrease the bandwidth at "A", "B", and "C" by ΔR_A, ΔR_B, and ΔR_C, respectively, the increases in energy consumption are different: $\Delta\Gamma_A$, $\Delta\Gamma_B$, and $\Delta\Gamma_C$. The increase in energy consumption of terminal "A" is lowest and highest at terminal "C". Therefore, more reduction in energy consumption, ($\Delta\Gamma_C$) can be gained by increasing the bandwidth for terminal "C" than the bandwidth increase for "A".

We note that in reality, the transmission rate is not changed continuously. For the convenience of illustration, we use the slope to represent the changing rate. In a realistic situation, we can consider the value of $\Delta\Gamma$ for different mobile terminals, which depends on the bandwidth of each logical channel. Since the same bandwidth variation may result in various changes in energy consumption, the problem of energy-based admission control is formulated how to select connections whose bandwidth can be reduced to accept more requests and how to re-allocate bandwidth to active services so that the average energy consumption of each terminal can be reduced. In addition, how to avoid the QoS compromise for multi-class traffic systems.

III. A New Admission Control Scheme Based on Energy Consumption

In this section, we present an adaptive scheme for connection admission control based on energy consumption. It distinguishes itself from the existing adaptive schemes in two aspects. First, in the connection setup phase, energy-based CAC

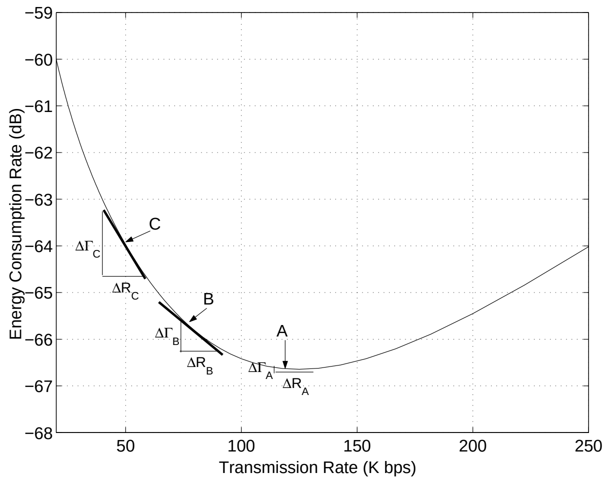

Fig. 1. An Example of Energy Consumption vs. Transmission Rate.

While $BW_a < BWL[t]$
  do $m \leftarrow \frac{d\Gamma_r(R)}{dR}$ of terminal, $i = 1$
   $j \leftarrow 1$
   for $i \leftarrow 2$ **to** K
    if for terminal i, $\frac{d\Gamma_r(R)}{dR} < m$
     AND $BW[i] - \Delta BW >= BWL[i]$
     then $m \leftarrow \frac{d\Gamma_r(R)}{dR}$ of terminal i
      $j \leftarrow i$
   if j == 1 **AND** $BW[1] - \Delta BW < BWL[1]$
    if t is a new request
     goto QoS Adjustment-Blocking
    else t is a handoff request
     goto QoS Adjustment-Dropping
   else
    $BW[j] \leftarrow BW[j] - \Delta BW$
    $BWa \leftarrow BWa + \Delta BW$
ACCEPT t

Fig. 2. A Pseudocode of Victim Selection Algorithm (VSA).

calculates the transmission power in order to have the mobile terminals operate at the minimum energy consumption rate point with the constraint of transmission rate. Second, during the transmission, the proposed scheme may change the transmission rate of a connection to adapt to incoming or outgoing requests upon energy status of mobile terminals. For example in a CDMA system, different number of channels can be achieved by allocating multiple spreading codes to the mobile terminals. In addition to the energy conservation being considered in the admission control, a QoS adjustment algorithm is proposed to stochastically adjust bandwidth allocation for satisfying QoS requirements in multi-class systems.

A. Victim Selection Algorithm

When a new or handoff request from a terminal, t, arrives at a point of attachment such as a base station, if the available bandwidth, BWa, is greater than the minimum bandwidth requirement, $BWL[t]$, then the BS accepts this request as shown in Fig. 2. Otherwise, the BS searches all ongoing services, i.e., a total number of K users, to find a *victim*, j, whose derivative of $\Gamma_r(R)$ is the minimum through a so called. *victim selection algorithm* (VSA). Through this algorithm, mobile terminals, whose derivative of energy consumption rate is the minimum, will be selected as "victims" and their transmission rate will be reduced to minimize the change in energy consumption. The BS will reduce the bandwidth, ΔBW, from a chosen victim, to a lower level to increase available bandwidth.

If a terminal already operates at its lower bound, $BWL[i]$, the bandwidth of this terminal cannot be reduced. As such, this mobile terminal cannot be treated as a "victim." This procedure will be repeated until the available bandwidth is greater than the requested bandwidth for accepting terminal, t. If no "victim" is available, VSA uses an adjustment algorithm described in Section III-C to select those terminals that their QoS are beyond their expectations.

B. Beneficiary Selection Algorithm

Once a connection is finished, the bandwidth for this session will be released. In order to utilize the bandwidth efficiently, we can reallocate the bandwidth to ongoing services. The *beneficiary selection algorithm* (BSA) is designed to choose "beneficiary" terminals because they will be allocated more bandwidth, resulting maximum decrease in energy consumption. For example, we can see in Fig. 1 that terminal "C" will be the beneficiary for BSA because of the maximum decrease in energy consumption resulting from the increase in transmission rate compared to the other two terminals, "A" and "B." A pesudocode of BSA is shown in Fig.3. The BS searches for a potential beneficiary from a total number of ongoing services, K, with ongoing services to determine a terminal, j, which can benefit most by receiving more bandwidth, i.e., $\Gamma_r(R)$ is the maximum. Definitely, a higher level of $BW[j]$ caused by the increase, ΔBW, will not exceed the upper bound of bandwidth requirement, $BWU[j]$, for the "beneficiary" terminal, j.

1 . **While** BWa > 0
2 . **do** $m \leftarrow \frac{d\Gamma_r(R)}{dR}$ of terminal, $i = 1$
3 . $j \leftarrow 1$
4 .   **for** $i \leftarrow 2$ **to** K
5 .   **if** for terminal i, $\frac{d\Gamma_r(R)}{dR} > m$
6 .   **AND** $BW[i] + \Delta BW <= BWU[i]$
7 .   **then** $m \leftarrow \frac{d\Gamma_r(R)}{dR}$ of terminal i
8 . $j \leftarrow i$
9 .   **if** j == 1 **AND** $BW[1] + \Delta BW > BWU[1]$
10.   **goto** 13
11. $BW[j] \leftarrow BW[j] + \Delta BW$
12. $BWa \leftarrow BWa - \Delta BW$
13. **return**

Fig. 3. A Pseudocode of Beneficiary Selection Algorithm (BSA).

C. Stochastic Adjustment Algorithm

For multi-class service systems, typically in current and future wireless networks, bandwidth allocation needs to be coupled with QoS requirements to achieve a balanced QoS guarantee. Thus, we introduce a *stochastic adjustment algorithm* (SAA) in that a BS partially blocks the other classes' traffic before they come into the system; we call this operation *pre-block*, even if the BS has resource available. In this way, the BS reduces the total system traffic load by sacrificing certain classes, thus creating a balance between the classes already beyond the QoS requirement and the classes that are under satisfied.

In the implementation of SAA, the BS will pre-block some traffic according to a stochastic process like a uniform distribution. For a blocked request of class k as shown in the pseudocode of SAA in Figure 4, the BS updates the number of blocked requests of class k, NB$[k]$, as well as the total number of new requests, NTB$[k]$. Then, the BS examines whether the blocking probability of class k, PB$[k]$ exceeds its QoS specification, β_k. If the blocking probability is still under the QoS requirement, the BS will not invoke any procedures. Otherwise, the BS will search for a class, j, whose blocking probability, PB$[j]$, is beyond its QoS specification to randomly block subsequent requests of class j so that the traffic load is reduced. If a handoff request is dropped, then a similar procedure of SAA will be implemented, except the dropping probabilities are considered as a measurement instead of blocking probabilities. Therefore, the QoS requirements are considered in our bandwidth allocation scheme in addition to energy conservation.

```
NB[k]  ← NB[k]  + 1
NTB[k] ← NTB[k] + 1
if PB[k] ≤ β_k
    return
else
    While i <= K
        do
            if PB[i] <= PBU[i]
                break
    if i > K
        return
    else
        randomly block class i's call request
        NB[i]  ← NB[i]  + 1
        NTB[i] ← NTB[i] + 1
```

Fig. 4. A Pseudocode of Stochastic Adjustment Algorithm (SAA).

D. Discussion

The proposed scheme for admission control has several advantages. First, our approach reduces energy consumption by CAC algorithms at the network layer, which is different from many previous schemes at the medium access control (MAC) layer. Second, the blocking probability and dropping probability incurred in our scheme are less than or equal to those of other schemes for multi-class systems. Moreover, our solution takes QoS into consideration, maintaining a balance among all the classes of traffic.

For CDMA systems, when mobile terminals change their power, the interference will also be changed. As a result, the capacity of the system may also be changed over time. However, this will not affect the application of the proposed scheme because the transmission rate is changed based on the number of channels rather than transmission power. In other words, either the victims or the beneficiaries will use the same transmission power as a result of power control schemes. It is possible that the proposed CAC scheme can be combined with power control algorithms, which is one of our future work.

One potential problem is that that decreasing bandwidth will possibly increase channel holding time. This might further increase the blocking probability because more terminals are in service. In fact, the results are more likely to be mixed. As the bandwidth of each user is decreased, more bandwidth will be available for admitting new requests. To this end, the blocking probability will be reduced. Moreover, to avoid high blocking probability, the stochastic algorithm is used to "suppress" this potential problem. From our simulation results, the blocking probability of "low-bound' case is lower.

Additional overhead to calculate the energy consumption rate for each terminal will be induced in the proposed scheme. However, in reality, it is not necessary for the BS to calculate for each terminal. We only need to compute the energy consumption rate for every possible number of channels in advance. For example, as to data service, the number of channels may range from 2 to 5. Then, the BS only needs to know the energy consumption rate for 2, 3, 4, and 5 channels, respectively. In other words, 4 points are enough for the BS, and thus, the computation overhead can be greatly reduced. Moreover, the SAA may not work for those processes that do not remain the same process after pre-blocking, such as Fractional Brownian Motion process or other long range dependence processes [16, 20]. We will investigate this problem in our future work.

IV. Performance Analysis

In this section, we analyze the performance of the proposed scheme based on the Continuous Time Markov Chain (CTMC). We do not consider handoff and new requests separately because they are processed in a very similar way from the point of view of energy conservation. In particular, we focus on two cases:

- *Case I*: Channel holding time is independent of the number of channels, which is referred to as *Fixed Channel Holding Time*. This assumption is widely deployed in existing work [24, 30], which is appropriate to describe multimedia applications, such as on-line video transmission. For example, if we watch a 20-minute video clip on-line with more bandwidth, we will see a bigger picture and/or higher resolution. Therefore, the bandwidth allocation affects video quality rather than transmission time.
- *Case II*: Channel holding time is dependent on the number of channels, which is referred to as *Dynamic Channel Holding Time* and it is often ignored by other works. This assumption can be applied to most data transmis-

sion, e.g., we want to download a 10-MB file. With more bandwidth, the transmission time will be shorter and vice versa. Since this is a common scenario in many network applications, it is necessary to investigate this case. Although it is difficult to reach a close-form formula for the blocking probability, we will present the upper bound and lower bound of the blocking probability for this case.

A. Markovian State Model

In addition to the notations described in Section II, we denote vector $\overrightarrow{N} = [n_1, n_2, \ldots, n_K]$ to represent the number of connections for each class, where n_k is the number of connections of class k in the system. Then, each state represents a possible combination of $[n_1, n_2, \ldots, n_K]$ in the system, and the state space is the set of all possible states, which is denoted by $\mathbf{S}$ as:

$$\begin{aligned} \mathbf{S} \quad := \quad & \{\overrightarrow{N} \in \mathcal{I}^K : \sum_{k=1}^{K}\sum_{i=1}^{n_k} b_k^i \leq C\} \quad \text{and} \\ & b_k^i \in [b_k^L, b_k^U] \end{aligned} \tag{7}$$

where $\mathcal{I}$ is a set of non-negative integers and $\mathcal{I}^K$ is a set of $K-dimensional$ non-negative vectors. Also, we let $\mathbf{S_k}$ be a subset of the state space for which an arriving request of class k is blocked, that is

$$\mathbf{S_k} := \{\overrightarrow{N} \in \mathcal{I}^K : C - b_k^L < \sum_{i=1}^{K} (b_i^L \times n_i) \leq C\}. \tag{8}$$

In order to keep track of the transmission rate of sessions in each class, let $\overrightarrow{E}_k$ be a K-dimensional vector of all "0" except for a "1" of the k^{th} element, e.g., $\overrightarrow{E}_2 = [0, 1, 0, \ldots, 0]$. So, we have two possible results after the change in the number of connections due to bandwidth allocation, represented by two indicators:

$$I_k^+(\overrightarrow{N}) = \begin{cases} 1 & \text{if } \overrightarrow{N} + \overrightarrow{E}_k \in \mathbf{S} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

and

$$I_k^-(\overrightarrow{N}) = \begin{cases} 1 & \text{if } \overrightarrow{N} - \overrightarrow{E}_k \in \mathbf{S} \\ 0 & \text{otherwise.} \end{cases}$$

Let $q(\overrightarrow{N_1}, \overrightarrow{N_2})$ denote the probability transition rate from state $\overrightarrow{N_1}$ to $\overrightarrow{N_2}$, and $\overrightarrow{\tau} = [\tau_1, \tau_2, \ldots, \tau_K]$ be the service rate factor to describe the dependence of channel holding time on the number of channels. If the channel holding time is independent of the number of channels, then this factor is 1, meaning the service rate will not be changed. Otherwise, the rate factor equals the number of channels, which is applicable to *Case II*. Therefore, we have

$$\begin{aligned} q(\overrightarrow{N}, \overrightarrow{N} + \overrightarrow{E}_k) &= \lambda_k(\overrightarrow{N}, \overrightarrow{N} + \overrightarrow{E}_k \in \mathbf{S}) \\ q(\overrightarrow{N}, \overrightarrow{N} - \overrightarrow{E}_k) &= n_k\tau_k\mu_k(\overrightarrow{N}, \overrightarrow{N} - \overrightarrow{E}_k \in \mathbf{S}) \\ q(\overrightarrow{N} - \overrightarrow{E}_k, \overrightarrow{N}) &= \lambda_k(\overrightarrow{N} - \overrightarrow{E}_k, \overrightarrow{N} \in \mathbf{S}) \\ q(\overrightarrow{N} + \overrightarrow{E}_k, \overrightarrow{N}) &= (n_k+1)\tau_k\mu_k(\overrightarrow{N} + \overrightarrow{E}_k, \overrightarrow{N} \in \mathbf{S}) \end{aligned} \tag{10}$$

where $k = 1, 2, \ldots, K$.

B. Blocking Probabilities

While we consider energy conservation in our CAC scheme, we do not want to sacrifice QoS requirements. Thus, we focus on discussing the blocking probabilities for *Case I* and *Case II* without using QoS adjustment algorithm in this section.

B.1 Analysis for Fixed Channel Holding Time

Because of the adaptive bandwidth allocation, the number of channels serving one connection may be changed due to the application of VSA and BSA algorithms. However, in this case we assume that the number of channels does not affect channel holding time. In other words, for one session, the number of channels, or transmission rate, is irrelevant to total transmission time. Rather, its channel holding time is exponentially distributed with mean $1/\mu_k$. Hence, inside a class, the service rate for each connection only depends on the number of terminals of the classed in the system.

Therefore, the global Markovian equilibrium balance equation can be expressed by

$$\begin{aligned} & \left[\sum_{k=1}^{K} \lambda_k I_k^+(\overrightarrow{N}) + \sum_{k=1}^{K} n_k\tau_k\mu_k I_k^-(\overrightarrow{N})\right] P(\overrightarrow{N}) \\ &= \sum_{k=1}^{K} \lambda_k I_k^-(\overrightarrow{N}) P(\overrightarrow{N} - \overrightarrow{E}_k) \\ &\quad + \quad \textstyle\sum_{k=1}^{K} (n_k+1)\tau_k\mu_k I_k^+ P(\overrightarrow{N} + \overrightarrow{E}_k) \end{aligned} \tag{11}$$

where $P(\overrightarrow{N})$ is the probability of each state $\overrightarrow{N}$. The local balance equation is represented as

$$\begin{aligned} \lambda_k I_k^- P(\overrightarrow{N} - \overrightarrow{E}_k) &= n_k\tau_k\mu_k I_k^- P(\overrightarrow{N}) \\ k &= 1, \ldots, K, \quad \forall \overrightarrow{N} \in \mathbf{S} \end{aligned} \tag{12}$$

Therefore, we can obtain $P(\overrightarrow{N})$ with $\rho_k = \lambda_k/\mu_k$ as follows:

$$P(\overrightarrow{N}) = \frac{1}{G(\mathbf{S})}\prod_{k=1}^{K}\frac{\rho_k^{n_k}}{\tau_k^{n_k} n_k!}, \quad \overrightarrow{N} \in \mathbf{S} \tag{13}$$

where

$$G(\mathbf{S}) = \sum_{\overrightarrow{N} \in \mathbf{S}}\prod_{k=1}^{K}\frac{\rho_k^{n_k}}{\tau_k^{n_k} n_k!}. \tag{14}$$

Then, the blocking probability of class k is equal to the sum of probabilities of all states to be blocked, i.e., the states in subset $\mathbf{S_k}$ given in (8). Hence, the blocking probability of class k, $B_k(\overrightarrow{\tau})$, can be written as

$$B_k(\overrightarrow{\tau}) = \frac{G(\mathbf{S_k})}{G(\mathbf{S})}, \quad \text{and} \quad \overrightarrow{\tau} = [1, 1, \ldots, 1] \tag{15}$$

where $\mathbf{S_k}$ and $\mathbf{S}$ are given in (7) and (8), respectively.

We should point out that even though the dynamics of the number of channels for a terminal do not change the blocking probability, it will make difference to the energy consumption as illustrated in Section IV-D.

B.2 Analysis for Dynamic Channel Holding Time

Here we consider that the service rate is proportional to the number of channels and denote the service rate of one channel for class k as μ_k. Then, the overall service rate for class k, $\mu_k(\overrightarrow{N})$, is

$$\mu_k(\overrightarrow{N}) = \sum_{i=1}^{n_k} \mu_k b_k^i, \quad \forall \overrightarrow{N} \in \mathbf{S} \tag{16}$$

where $b_k^i \in [b_k^L, b_k^U]$ and $\mathbf{S}$ as given in (7).

Since it is difficult to obtain the close form expression of blocking probabilities for this case, we discuss the upper bound and lower bound of the blocking probability. With the assumption above, we have

$$\begin{aligned} \mu_k^U(\overrightarrow{N}) &= \sum_{i=1}^{n_k} \mu_k b_k^i \le \sum_{i=1}^{n_k} \mu_k b_k^U = n_k b_k^U \mu_k \quad (17) \\ \mu_k^L(\overrightarrow{N}) &= \sum_{i=1}^{n_k} \mu_k b_k^i \ge \sum_{i=1}^{n_k} \mu_k b_k^L = n_k b_k^L \mu_k \end{aligned}$$

Thus, we do not need to consider the random variable b_k^i in obtaining these bounds, but to use known parameters b_k^U and b_k^L; thus, a semi-Markov process is simplified to a CTMC. Note that the vector $\overrightarrow{\tau}$ is not a constant for *Case II* because the channel holding time is related to the number of channels. In a similar way as in the previous subsection, we can obtain the lower bounds of blocking probability as

$$B_k^L(\overrightarrow{\tau_L}) = \frac{G(\mathbf{S_k})}{G(\mathbf{S})}, \quad \text{and} \quad \overrightarrow{\tau_L} = [b_1^L, b_2^L, \dots, b_K^L]. \tag{18}$$

The upper bound, B_k^U, has the same expression with the exception of substituting $\overrightarrow{\tau_U}$ for $\overrightarrow{\tau_L}$ in the above equation.

C. *Blocking Probabilities with Stochastic Adjustment*

When the blocking probability of class i cannot be satisfied, i.e., the blocking probability of class i is greater than β_i, a stochastic adjustment algorithm (SAA) shown in Fig. 4 is applied in which some traffic will be blocked on purpose to reduce the traffic load of the whole system. Let vector $\overrightarrow{M} = [m_1, \cdots, m_k, \cdots, m_K]$ be the rate of arrival requests blocked for each class, where $m_k \ge 0$ is the blocked traffic for class k.

Lemma 5.1: A Poisson process with an arrival rate of λ_k is still a Poisson process after pre-blocking with a rate of m_k, and the new process is with arrival rate $\lambda_k - m_k$.

Assume that a Poisson process is $\{N(t), t \ge 0\}$ and the process after pre-blocking as $\{N_b(t), t \ge 0\}$. First, we condition on $N(t)$

$$\begin{aligned} P\{N_b(t) = n\} &= \sum_{i=n}^{\infty} P\{N_b(t) = n | N(t) = i\} P\{N(t) = i\} \\ &= \sum_{i=n}^{\infty} \binom{i}{n} (\frac{\lambda_k - m_k}{\lambda_k})^n (\frac{m_k}{\lambda_k})^{i-n} e^{\lambda_k t} \frac{(\lambda_k t)^i}{i!} \end{aligned} \tag{19}$$

The above expression can be simplified and we have

$$P\{N_b(t) = n\} = e^{(\lambda_k - m_k)t} \frac{[(\lambda_k - m_k)t]^n}{n!} \tag{20}$$

Therefore, we justify that $\{N_b(t), t \ge 0\}$ is a Poisson process with the arrival rate $\lambda_k - m_k$. The actual traffic load for class k after pre-blocking is $\rho_k = \frac{\lambda_k - m_k}{\mu_k}$. Then the blocking probability of class k becomes $B_k|\mathbf{Q}$, which is

$$B_k|\mathbf{Q} = B_k(\overrightarrow{\tau}) + \frac{m_k}{\lambda_k}, \tag{21}$$

where B_k can be obtained by using (15).

Next, we discuss how to determine $\overrightarrow{M} = [m_1, m_2, \cdots, m_k]$ through a pricing model. Let us define the price function of class k, Π_k as

$$\Pi_k(m_1, m_2, \dots, m_K) = \mathbb{F}_k(B_k|\mathbf{Q}), \quad k = 1, \dots, K \tag{22}$$

Consequently, the objective function of stochastic QoS adjustment is to find vector $\overrightarrow{M}$ such that

$$\begin{aligned} \min \quad & \{\textstyle\sum_{k=1}^{K} \Pi_k(m_1, m_2, \dots, m_K)\} \\ \text{s.t.} \quad & \text{constraint } X \\ & m_k \ge 0, \quad k = 1, \dots, K \end{aligned} \tag{23}$$

Therefore, the solution of $\overrightarrow{M}$ can be obtained by using fractional programming for a specific requirement of X, which varies with the systems. For example, consider that there are 2 service classes in the system: data service ($k = 1$) and multimedia service ($k = 2$). The class of data service is assigned a higher priority in terms of lower blocking probability. Then, we may choose $\mathbb{F}_1(B_1|\mathbf{Q}) = \alpha \times B_1|\mathbf{Q}$ and $\mathbb{F}_2(B_2|\mathbf{Q}) = B_2|\mathbf{Q}$, where $0 < \alpha < 1$. The price function Π_1 and Π_2 can be written as

$$\begin{aligned} \Pi_1(m_1, m_2) &= \alpha B_1|\mathbf{Q}, \quad \text{and} \\ \Pi_2(m_1, m_2) &= B_2|\mathbf{Q}. \end{aligned} \tag{24}$$

The objective function of stochastic QoS adjustment becomes finding m_1, m_2 such that

$$\begin{aligned} \min \quad & \{\ \alpha B_1|\mathbf{Q} + B_2|\mathbf{Q}\ \} \\ \text{s.t.} \quad & B_1|\mathbf{Q} \le B_2|\mathbf{Q} \\ & m_1 \ge 0, \ \ m_2 \ge 0. \end{aligned} \tag{25}$$

D. Energy Consumption

The average energy consumption, $\mathbb{E}$, can be represented by the product of the probability of each state, $P(\overrightarrow{N})$ and the average energy consumption rate $E(\overrightarrow{N})$, that is

$$\mathbb{E} = \sum_{\overrightarrow{N} \in \mathbf{S}} P(\overrightarrow{N}) \times E(\overrightarrow{N}) \tag{26}$$

where $P(\overrightarrow{N})$ in (13) can be obtained by substituting different $\overrightarrow{\tau}$ for *Case I* and *Case II* as discussed in Section IV-B. In order to acquire $E(\overrightarrow{N})$, we define the energy consumption rate (ECR) matrix as

$$ECR = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1C} \\ a_{21} & a_{22} & \dots & a_{2C} \\ \vdots & \vdots & \ddots & \vdots \\ a_{K1} & a_{K2} & \dots & a_{KC} \end{pmatrix} \tag{27}$$

where each element, a_{ij}, is the energy consumption rate of class i with j channels, which can be obtained from energy consumption rate, defined in (6). In wireless systems, the BS is aware of the transmission rate of each session. For example, in cdma2000 systems with bandwidth $1.25MHz$, single channel with spreading factor 64 can provide $9.6\ Kbps$ peak transmission rate [18]. Therefore, based on energy consumption rate and the number of channels of a certain transmission rate, a_{ij} can be obtained.

For $\overrightarrow{N} \in \mathbf{S}$, we choose $j \in [b_k^L, b_k^U],\ \ k = 1, \dots, K$ such that

$$\min_{k=1,\dots,K} \{ \ \sum_{k=1}^{K} \sum_{i=1}^{n_k} a_{kj} \ \} \tag{28}$$

then the average energy consumption, $E(\overrightarrow{N})$, is given by

$$E(\overrightarrow{N}) = \frac{\sum_{k=1}^{K} \sum_{i=1}^{n_k} a_{kj}}{\sum_{k=1}^{K} n_k}. \tag{29}$$

Then, by substituting $E(\overrightarrow{N})$ in (26), total energy consumption can be obtained. For simplicity, we assume a constant transmission rate for a fixed number of channels, whereas in reality, these rates may be varying from time to time. The computation of the above solution would not be scalable for a large number of classes. However, by far, there exist only about 3 to 5 classes of traffic. Therefore, the proposed solution is feasible in real systems.

V. Simulation Results

The performance of the proposed scheme is evaluated by simulations with respect to blocking probabilities, throughput, and energy consumption. The results of the proposed bandwidth allocation scheme are compared with other schemes such as Non-Prioritized Scheme (NPS) [19] and Adaptive REsource Allocation Scheme (AREAS) [15]. In addition to the two cases of traffic discussed in Section IV, we also study the effect of the proposed scheme on self-similar traffic as the arrival process, which is modeled as long-range dependent (LRD) traffic [28] due to the fact that traffic pattern of multimedia services in wireless networks tends to be LRD [12].

We consider a system with the following parameters, which is defined in the previous section.

- The total number of channels, $C = 100$.
- The number of service classes, $K = 2$.

Although in our experiments, only two classes of service are considered, whereas the effectiveness of the proposed scheme will not be affected by more classes. The other parameters used in our simulation are shown in Table I.

TABLE I
SIMULATION PARAMETERS.

Parameters	Class-A	Class-B
Arrival Rate: λ (calls/min)	[1,40]	10
Data Volume: ϕ (KByte)	{60,180}	{60,180}
Transmission Rate (Kbps/channel)	8	8
No. of Channels for NPS	7	3
No. of Channels for AREAS	{4,7,9}	{2,3,5}
No. of Channels for New Scheme	{4,9}	{2,5}
No. of Sources: M	100	100
Traffic density: R	[0.1-1.5]	1
Power-Law: A	500	500
Power-Law: γ	1.5	1.5

A. Case I: Fixed Channel Holding Time

The blocking probabilities for Class-A and Class-B with the comparison to other schemes are shown in Fig. 5. We can observe only four curves in the figure, although in fact, they are the results of eight scenarios. This is because that the simulation and analytical results of the proposed scheme (noted as 'New Scheme' in the figure) match so closely that they cannot be distinguished from each other. Moreover, the blocking probabilities of the new scheme and AREAS for both classes overlap, which results in only four separate curves in Fig. 5. As would be expected, the blocking probabilities of adaptive bandwidth allocation schemes are much lower than those of the NPS, and they increase as the traffic load increases. The blocking probability of Class-A is higher than the corresponding blocking probability of Class-B, since the lower bound of Class-B is smaller than that of Class-A. Thus, the new scheme reduces energy consumption significantly, while it yields the same blocking probability and throughput as that of AREAS.

The throughput of two classes as a function of λ_A, the arrival rate of Class-A, is plotted in Figs. 6 and 7. Even though the blocking probability is increased, the throughput of three schemes in Class-A is increased, whereas for Class-B, the throughput is decreased because of the constant arrival rate and increasing arrival rate of Class-A traffic. Fig. 8 displays the average energy consumption as a function of λ_A. We can see that the new scheme, which is an energy-based scheme, consumes less energy compared to the other two schemes with the same blocking probability. Since NPS does not change bandwidth during transmission, the average energy consumption remains unchanged. When the traffic load is light, the energy consumption of AREAS is lower than that of NPS; and during heavy traffic, AREAS is higher than NPS. However, the proposed scheme always transmits at a lower energy consumption rate; thus, the new scheme conserves more energy, while not sacrificing call blocking probability for multi-class traffic.

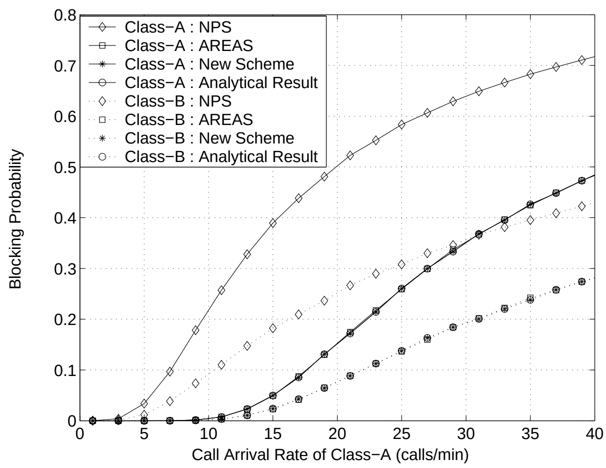

Fig. 5. *Case I*: Call Blocking Probability.

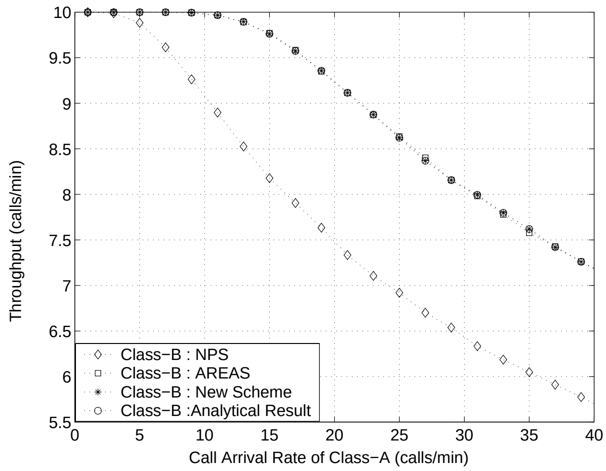

Fig. 7. *Case I*: Throughput of Class-B.

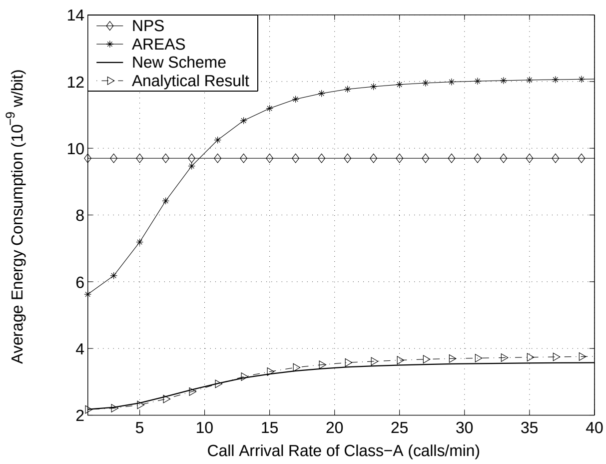

Fig. 8. *Case I*:Average Energy Consumption.

B. *Case II: Dynamic Channel Holding Time*

In most of existing work, the channel holding time is considered as independent of the number of channels, which is our *Case I*. However, for many data and data applications, more bandwidth means shorter transmission time, that is, the channel holding time is dependent on the number of channels. The call blocking probability of Class-A and Class-B are compared in Figs. 9 and 10 as the increase of λ_A. Unlike Fig. 5, the blocking probability of the new scheme is lower than that of AREAS for *Case II*, which means that the new scheme is better. The upper and lower bounds of blocking probability, along with the simulation results of the proposed scheme are shown in Fig. 11 for Class-A traffic, and similar results for Class-B are omitted due to the page limit. We can observe that the blocking probabilities of both classes fall in between the theoretical upper bound and lower bound.

At lower traffic load, the blocking probabilities of both classes tend to approach upper bound (say, the arrival rate is less than 20 calls/min). And at higher traffic load, the blocking probabilities are close to the lower bound (say, the arrival rate is greater than 35 calls/min). As the traffic increases, two bounds converge; especially, at high traffic load, the differ-

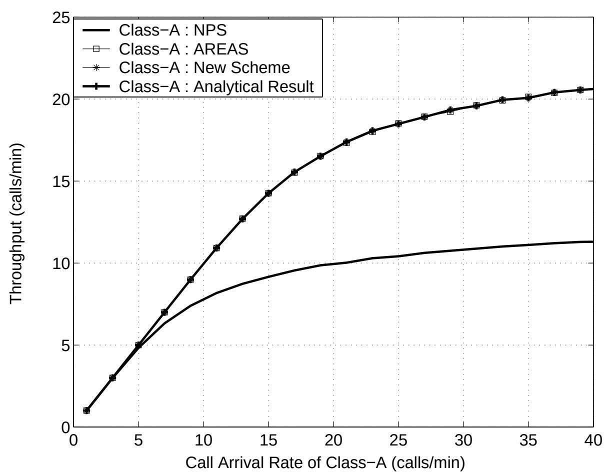

Fig. 6. *Case I*: Throughput of Class-A.

ence between the upper bound and lower bound will be less than 10%.

The throughput as a function of λ_A are shown in Fig 12. It can be observed that even though the blocking probability increases for Class-A, the new scheme yields a higher throughput compared to other two schemes. For Class-B, the new scheme also yields a higher throughput compared to the other two schemes, but decreasing with the increase in arrival rate of Class-A, even though the figures are omitted due to page limit. The energy consumption is not shown here because it is similar to that for *Case I*. This means that the new solution for energy conservation is robust to the variation of the incoming traffic, although there is a significant difference in the blocking probability.probability.

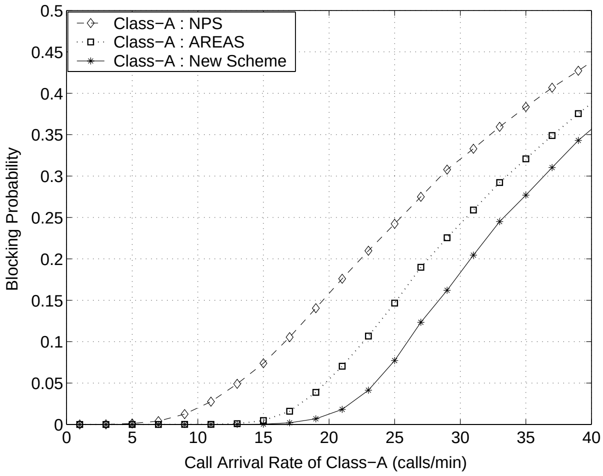

Fig. 9. *Case II*: Call Blocking Probability of Class-A.

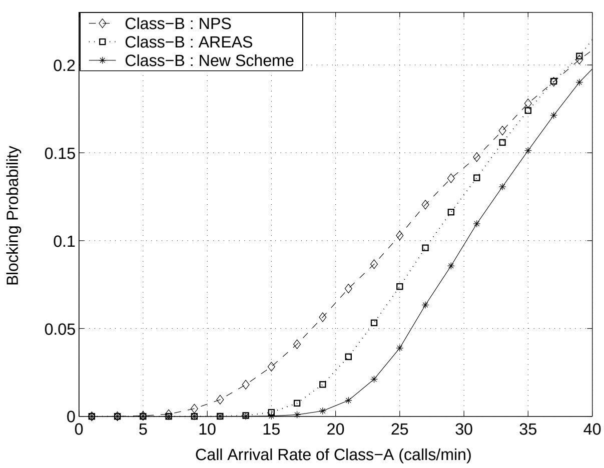

Fig. 10. *Case II*: Call Blocking Probability of Class-B.

C. *Stochastic QoS Adjustment*

Fig. 13 displays the result of stochastic QoS adjustment for *Case I* only, because the results for *Case II* are very similar

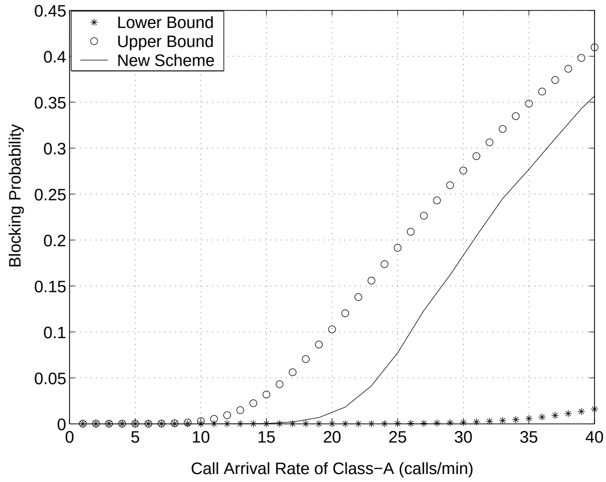

Fig. 11. *Case II*: Bounds of Blocking Probability of Class-A.

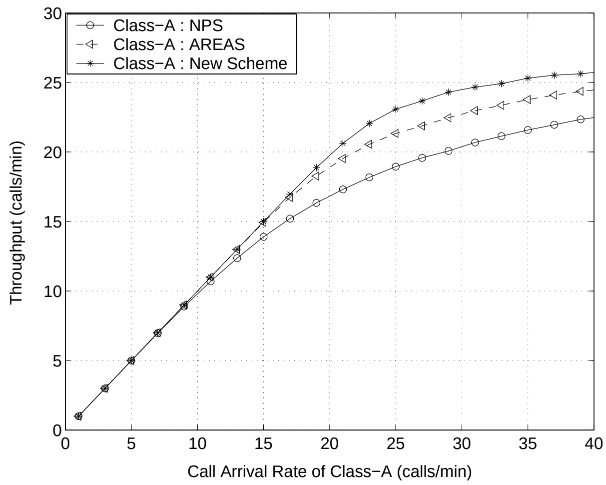

Fig. 12. *Case II* : Throughput of Class-A.

. We consider a dynamic adjustment of pre-blocking such as m_2 is changed from 9 to 11 for high traffic with $\lambda_A = 5$, and λ_B is changed from 45 to 65. From our simulations results, we realize that blocking probability is affected significantly with the change in arrival rate of Class-B. Meanwhile, the blocking probability of Class-A is further decreased. The blocking probability of two classes become very close to each other, which means a better QoS balance. The blocking probability after adjustment in Fig. 13 is quite different. By sacrificing Class-B, the blocking probability of Class-A has a significant decrease. A close look at Fig. 13 reveals that the blocking probability of Class-B changes very little with the increase of the traffic load, which means that our adjustment is adaptive to the traffic. Therefore, the proposed QoS adjustment algorithm SSA is very effective in achieving a tradeoff QoS for multi-class systems.

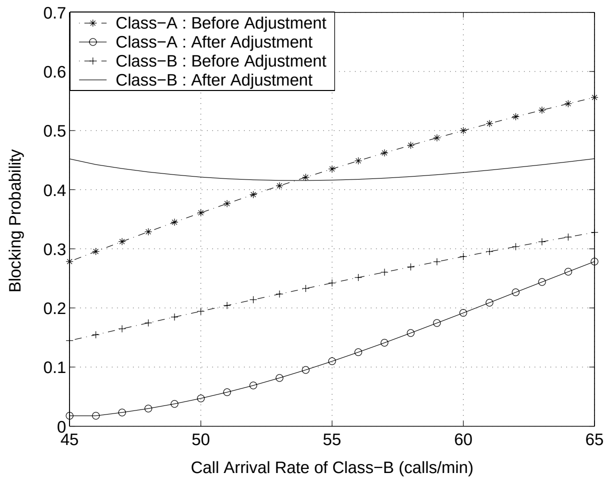

Fig. 13. QoS Adjustment with Pricing Constraint.

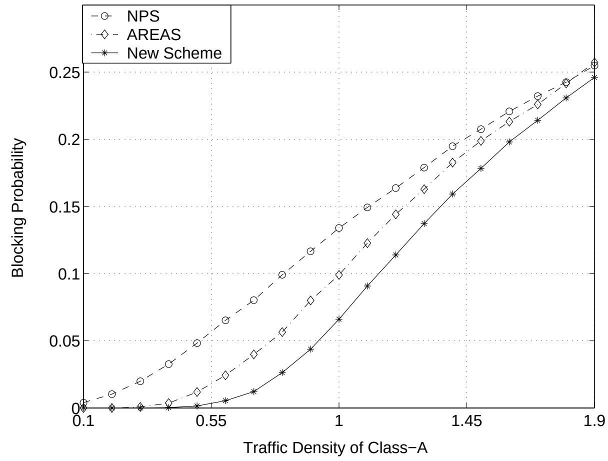

Fig. 15. Self-Similar Traffic: Blocking Probability of Class-B.

D. *Self-Similar Arrival Process*

Since the call arrival process may not be Poisson process for multimedia and data services [12], self-similar traffic, which tends to be long-range dependence (LRD) [28], is considered to be more appropriate. Without losing generality, we deploy Power-on and Power-off mode to generate self-similar traffic and the power-law distribution as [25]

$$p(t) = \begin{cases} \gamma A^{-1} e^{-\frac{\gamma t}{A}} & 0 \le t \le A \\ \gamma e^{-\gamma} A^{\gamma} t^{-(\gamma+1)} & t > A. \end{cases}$$

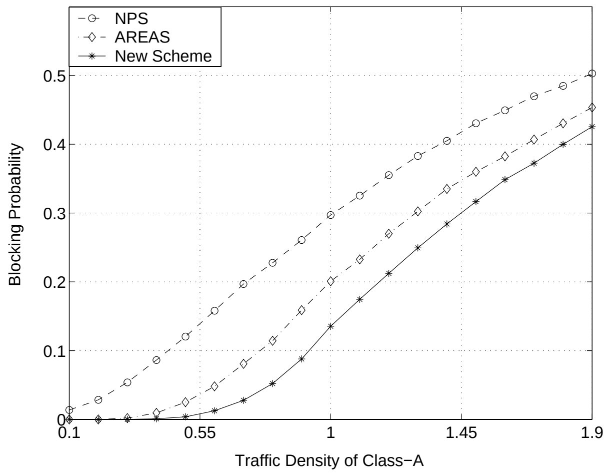

Fig. 14. Self-Similar Traffic: Blocking Probability of Class-A.

Figs. 14 and 15 depict the blocking probability of Class-A and Class-B, respectively. It is easily noticed that the new scheme results in lower blocking probability. However, for self-similar arrival traffic, we take a longer time to obtain stable results than Poisson arrival traffic because self-similar traffic is a long-range dependence traffic, and Poisson traffic is a short range dependence traffic. The average energy consumption for self-similar traffic is very similar to that for *Case I* and *Case II* based on our simulations. Thus, we conclude that the effectiveness of energy conservation is not affected significantly by traffic patterns.

The throughput of self-similar arrival process is shown in Figures 16 and 17, for Class-A and Class-B, respectively. The results illustrate that the new admission control scheme achieves higher throughput than those of other two schemes, because for the same traffic load, the new scheme serve more connections with lower blocking probability.

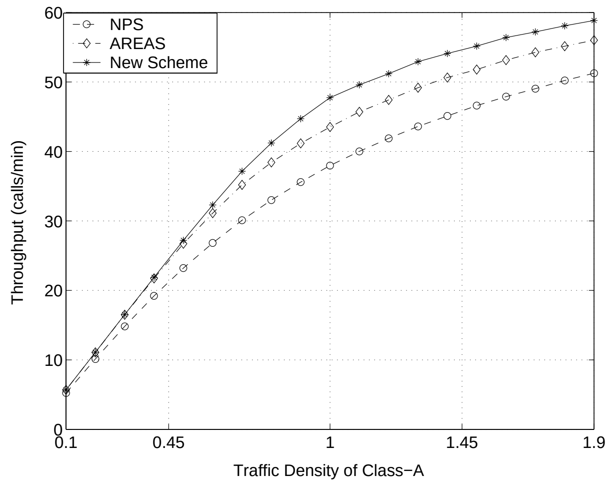

Fig. 16. Self-Similar Traffic: Throughput of Class-A.

VI. Conclusions

In this paper, we considered the problem of how to minimize the energy consumption through connection call admission control because both battery life and radio bandwidth are limited resource in mobile communications. We introduced a new parameter, energy consumption rate (ECR), by

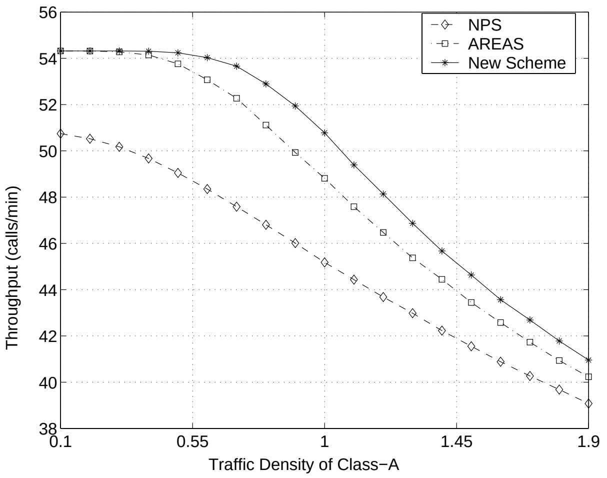

Fig. 17. Self-Similar Traffic: Throughput of Class-B.

exploring the connection between CAC and energy consumption, which represents the energy consumed for each bit that is successfully transmitted. We proposed a new admission control scheme which includes three parts: Victim Selection Algorithm (VSA), Beneficiary Selection Algorithm (BSA), and Stochastic Adjustment Algorithm (SAA), based on the fact that the same variation in transmission rate may yield different energy consumption of mobile terminals. The performance of the new scheme is analyzed with respect to call blocking probability, throughput, and energy consumption. Furthermore, the effectiveness of the proposed scheme is verified through extensive simulations for a variety of traffic patterns in multi-class systems.

References

[1] N. Bambos. Toward Power-Sensitive Network Architectures in Wireless Communications: Concepts, Issues, and Design Aspects. *IEEE Journal on Selected Areas Communications*, 18:966–976, June 2000.

[2] N. Bambos and S. Kandukuri. Power-Controlled Multiple Access Schemes for Next-Generation Wireless Packet Networks. *IEEE Wireless Communications*, pages 58–64, June 2002.

[3] N. Bambos and J. M. Rulnick. Mobile Power Management for Wireless Communication Networks. *IEEE Wireless Networks*, 3(1):3–14, 1997.

[4] E. U. Biyikoglu, B. Prabhakar, and A. E. Gammal. Energy-efficient Packet Transmission over a Wireless Link. *IEEE Transaction on Networking*, 10:487–499, Aug 2002.

[5] W.A. Chang and R.S. Ramakrishna. QoS provisioning dynamic connection-admission control for multimedia wireless networks using a Hopfield neural network. *IEEE Transactions on Vehicular Technology*, 53(1):106–117, January 2004.

[6] M-H. Chiu and M.A. Bassiouni. Predictive Schemes for Handoff Prioritization in Cellular Networks Based on Mobile Positioning. *IEEE Journal on Selected Areas in Communications (JSAC)*, 18(3):510–522, February 2000.

[7] S.K. Das, S.K. Sen, K. Basu, and H. Lin. A framework for bandwidth degradation and call admission control schemes for multiclass traffic in next-generation wireless networks. *Journal on Selected Areas in Communications*, 21(10):1790–1802, December 2003.

[8] Y. Eisenberg, C. E. Luna, T. N. Pappas, R. Berry, and A. K. Katsaggelos. Joint Source Coding and Transmission Power Management for Energy Efficient Wireless Video Communications. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(6):411–424, June 2002.

[9] Y. Eisenberg, T. N. Pappas, R. Beery, and A. K. Katsaggelos. Minimizing Transmission Energy in Wireless Video Communications. In *Proc. ICIP, Thessaloniki, Greece*, volume 1, pages 958–961, October 2001.

[10] E. Elalfy, Y. D. Yao, and H. Heffes. Adaptive Resource Allocation with Prioritized Handoff in Cellular Mobile Networks under QoS Provisioning. In *IEEE Vehicular Technology Conference (VTC), 2001*, volume 4, pages 2113 –2117, 2001.

[11] B. M. Epstein and M. Schwartz. Predictive QoS-based Admission Control for Multiclass Traffic in Celluar Wireless Networks. *IEEE Journal on Select Areas Communications*, 18:523–534, March 2000.

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

[13] W. Hou and Y. Fang. Mobility-Based Call Admission Control Schemes for Wireless Mobile Networks. *Wireless Communications and Mobile Computing*, 1(3):269–282, July-September 2001.

[14] R. Kravets and P. Krishnan. Application-driven Power Management for Mobile Communication. *Wireless Networks*, 6:263–277, 2000.

[15] J. Lee, T. June, S. Yoon, S. Youm, and C. Kang. An Adaptive Resource Allocation Mechanism Including Fast and Reliable Handoff in IP-Based 3G Wireless Networks. In *IEEE Wireless Communications*, volume 7, pages 42–47, December 2000.

[16] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson. On the Self-Similar Nature of Ethernet Traffic (Extended Version). *IEEE/ACM Transaction on Networking*, 2(1):1–15, February 1994.

[17] J. Mark and W. Zhuang. Wireless Communications and Networking. In *Prentice Hall, NJ*, January 2003.

[18] S. Nanda, K. Balachandran, and S. Kumar. Adaptation Techniques in Wireless Packet Data Service. *IEEE Communications Magazine*, 38(1):54–64, 2000.

[19] L. Ortigoza-Guerrero and A. H. Aghvami. A prioritized handoff dynamic channel allocation strategy for PCS. *IEEE Trans. on Vehicular Technology*, 48(4):1203–1215, July 1999.

[20] G. Oppenheim P. Doukhan and M.S.Taqqu. Theory and Applications of Long-Range Dependence. *Birkhauser, Boston*, 2003.

[21] I.C. Panoustsopaulos, S. Kotssopoulos, and V. Tountopoulos. Handover and New Call Admission Policy Optimization for 3G Systems. *Wireless Networks*, 8(4):381–389, 2002.

[22] B. Prabhakar, E. U. Biyikoglu, and A. E. Gamal. Energy-Efficient Transmission over a Wireless Link via Lazy Packet Scheduling. In *Proc. IEEE Infocom 2001*, volume 1, pages 386 –394, April 2001.

[23] J. G. Proakis. Digital Communication. *McGraw-Hill*, 2002.

[24] P. Ramanathan, K. M. Sivalingam, P. Agrawal, and S. Kishore. Dynamic Resource Allocation Schemes During Handoff for Mobile Multimedia Wireless Networks. *IEEE Journal on Selected Areas in Communications*, 17:1270–1283, July 1999.

[25] B. Ryu and S. Lowen. Fractal Traffic Models for Internet Simulation. In *IEEE Symposium on Computers and Communications, ISCC 2000*, pages 200–206, July 2000.

[26] C. U. Saraydar, N. B. Mandayam, and D. J. Goodman. Efficient Power Control via Pricing in Wireless Data Networks. *IEEE Trans. on Communications*, 50(2):291–303, February 2002.

[27] D.C. Schultz, S-H Oh, C.F. Grecas, and et al. A QoS Concept for Packet Oriented S-UMTS Services. In *IST Mobile and Wireless Telecommunications Summit*, June 2002.

[28] W. Willinger, V. Paxson, and M. S. Taqqu. Self-Similarity and Heavy Tails: Structural Modeling of Network Traffic. *A Practical Guide to Heavy Tails: Statistical Techniques for Analyzing Heavy Tailed Distributions*, 1998.

[29] Q. Zhang, Z. Ji, W. Zhu, and Y. Zhang. Power-Minimized Bit Allocation for Video Communication Over Wireless Channels. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(6):398–410, June 2002.

[30] T. Zhang, E.V.D. Berg, J. Chennikara, P. Agrawal, J-C. Chen, and T. Kodama. Local Predictive Resource Reservation for Handoffs in Multimedia Wireless IP Networks. *IEEE Journal on Selected Areas in Communications (JSAC)*, 19(10):1931–1941, October 2001.

[31] M. Zorzi. Outage and Error Events in Bursty hannels. *IEEE Transactions on Communications*, 46(3):349 –356, June 1998.