

Energy-Efficient Bandwidth Allocation in Wireless Networks: Algorithms, Analysis, and Simulations

Wenye Wang, Xinbing Wang, and Arne A. Nilsson

Abstract—In this paper, we present a new energy-efficient bandwidth allocation scheme for wireless networks. First of all, we investigate the intrinsic relationship between the energy consumption and transmission rates of mobile terminals, in which transmission rate is determined through channel allocations. Then, we propose two schemes for connection admission control: Victim Selection Algorithm (VSA) and Beneficiary Selection Algorithm (BSA) with the intent to reduce energy consumption of each terminal. Moreover, we introduce an adjustment algorithm to statistically meet the demands for quality of service (QoS) during the resource allocation. The performance of the proposed schemes is evaluated with respect to energy consumption rate of each successfully transmitted bit, throughput and call blocking probabilities. An extensive analysis and simulation study is conducted for Poisson and self-similar, multi-class traffic.

Index Terms—Wireless networks, energy consumption, and connection admission control.

I. INTRODUCTION

MOBILE devices are often deployed in wireless networks, whereas most of them operate on batteries such as cellular phones, portable digital assistants (PDAs), and laptops in situations with no available power supply. 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.

The issue of supporting QoS in wireless networks has been researched extensively in the past decade via connection admission control (CAC). Many solutions have been proposed to dynamically allocate bandwidth based on different criteria, including system priority, capacity, mobility, and interference [1], [2], [3], [4], [5], [6]. One solution uses a guard channel scheme (GCS) by assigning a higher capacity limit for handoff requests than new calls [1], [4], [7]. The difference between two limits is called “guard channel”. The number of guard channels can be either fixed [1], [4], or adaptive [7] to satisfy the minimum requirements of call dropping/blocking

probabilities. In order to meet the requirements of heterogeneous traffic with time-varying characteristics, priority-based schemes were proposed in [2], which give a higher priority to handoff requests than to new calls in order to reduce handoff dropping probabilities. Moreover, prediction-based approaches are presented to reserve bandwidth in a specified location prior to the arrival of handoff requests based on mobility of users [3], [5].

Meanwhile, many wireless systems are using code division multiple access (CDMA) technology. Several admission control schemes based on interference and power control are proposed, where the acceptance of a new request depends on signal-to-interference ratio (SIR) values [8], [9], [10], [11], [12]. In [13], [14], integrated admission control uses game theory and intelligent call admission controller uses fuzzy logic theory to estimate the interference caused by new/handoff calls and other QoS parameters to make decision, respectively. Although energy conservation has been recognized as one of the critical issues in wireless networks, it is hardly considered in the CAC algorithms [8], [11]. 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 [9], [15], [16]. For instance, joint source and channel coding with power control are exploited [17] to prolong the battery life of mobile terminals. It is also suggested that energy consumption can be reduced by smartly turning off mobile devices during idle time [18]. It is noticeable that transmission energy consumption can be minimized by adapting transmission rate to channel conditions under the constraints of QoS such as delay and distortion [19].

In this paper, we aim to integrate bandwidth allocation and energy conservation, while reducing call blocking and handoff dropping probabilities for multi-class traffic. The system model and QoS requirements are introduced in Section II. We define a new parameter, *energy consumption rate*, which describes how energy consumption is influenced by transmission rates and CAC algorithms in Section III. Then, we propose adaptive admission control schemes for multi-class systems in Section IV, including *victim selection algorithm* (VSA) and *beneficiary selection algorithm* (BSA). In order to guarantee QoS requirements for multi-class services, we propose a *stochastic adjustment algorithm* (SAA) for dynamic resource allocation. The performance of multi-class services in wireless networks is analyzed and evaluated by extensive simulations of Poisson and self-similar traffic in Sections V and VI, respectively. Finally, the paper is concluded in Section VII.

Manuscript received April 07, 2004; revised December 23, 2004; accepted April 16, 2005. The associate editor coordinating the review of this letter and approving it for publication was J. Hou.

The authors are with the North Carolina State University, Raleigh, NC 27695 (e-mail: wwang@ncsu.edu)

Digital Object Identifier 10.1109/TWC.2006.05018.

II. MULTI-CLASS SYSTEM MODEL

The multi-class system model, unlike a single-class model, is very important in analyzing performance of a wireless network because of the demand for various applications in mobile environments [5], [20]. 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, with 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.

The following parameters are defined in order to describe our system model:

| | |
|----------------|--|
| C | system capacity; |
| K | number of classes, $k = 1, 2, \dots, K$; |
| 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 ; |
| $\vec{\beta}$ | 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 |
| $\vec{\delta}$ | 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, whereas data services may require from 2 to 5 channels. 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 [5], [6] in the analysis; therefore, we denote λ_k and μ_k as the mean arrival rate and service rate for class k , respectively.

III. CONNECTION ADMISSION CONTROL (CAC) AND ENERGY CONSUMPTION

In wireless networks, bandwidth allocation is critical to supporting a variety of services with desired QoS expectation such as transmission rate, delay, and loss. 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 objective of CAC is to allocate bandwidth, that is, to determine the transmission rate of a mobile device while not degrading QoS parameters. Meanwhile, according to communication theory, transmission rate is closely related to the energy consumption per bit. In [21], the relationship between transmission rate and energy consumption is explained as follows. 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.

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. Let vector $\vec{\mathbf{P}} = [P_1, P_2, \dots, P_i, \dots, P_N]$ 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 $\vec{\mathbf{R}} = [R_1, R_2, \dots, R_i, \dots, R_N]$, where R_i is the transmission rate of terminal i . The channel gain for each user is represented by a vector $\vec{\mathbf{H}} = [h_1, h_2, \dots, h_i, \dots, h_N]$. Then, the signal-to-interference and noise ratio, $SINR$, of mobile user i can be written as [22]

$$SINR = \frac{h_i P_i}{\sum_{j \neq i} h_j P_j + \eta_0 W}. \quad (1)$$

Let E_b denote the energy of transmission signal per information bit η_0 is the noise spectral density including thermal noise and interference. For existing wireless systems, we consider that bit error rate (BER), P_{ber} , as [23]

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

Since the payload of a service request will be encapsulated into frames during transmission, thus, we need to know *frame error rate* (FER), which is determined by BER and coding schemes. In this paper, we use *Reed-Solomon* (RS) code [17] 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*, $P_s(E_b)$, and *frame error rate*, $P_{fer}(E_b)$, using $RS(n, k)$ are given by

$$P_s(E_b) = 1 - (1 - P_{ber}(E_b))^n \quad \text{and} \quad (3)$$

$$P_{fer}(E_b) = \sum_{j=t+1}^n \binom{n}{j} P_s(E_b)^j (1 - P_s(E_b))^{n-j}. \quad (4)$$

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

$$\Omega(P_{fer}(E_b)) = \frac{1}{1 - P_{fer}(E_b)}. \quad (5)$$

In addition, total transmission time of mobile terminal i , T_{total}^i , depends on the total amount of data, number of

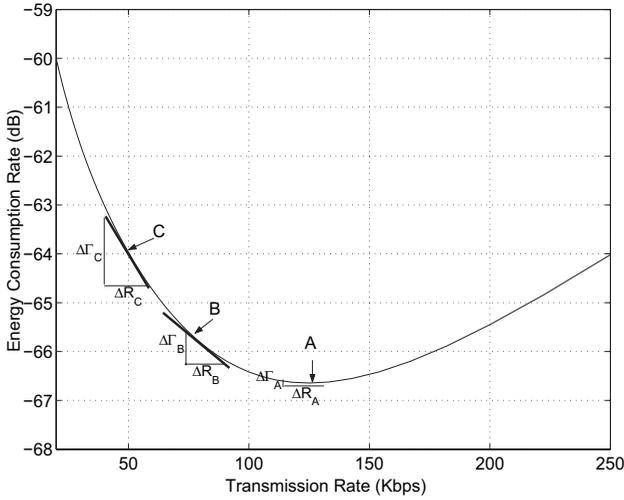


Fig. 1. An Example of Energy Variation versus Bandwidth Change.

retransmissions, and transmission rate. That is

$$T_{total}^i = \frac{\phi_i \cdot \Omega(P_{fer}(E_b))}{R_i} = \frac{\phi_i}{R_i(1 - P_{fer}(E_b))} \quad (6)$$

where ϕ_i is the total amount of data to be transmitted for mobile terminal i . And the total energy consumption, E_{total}^i , can be expressed as $E_{total}^i = T_{total}^i \cdot R_i \cdot E_b$. Therefore, we observe 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 a fixed volume of information. The ratio of total energy consumption over the data volume is the energy consumption for each bit. Thus, we define *energy consumption rate* (ECR), $\Gamma(E_b)$, as

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

$$\Gamma(E_b) = \frac{E_b}{1 - P_{fer}(E_b)}. \quad (7)$$

To demonstrate the effect of transmission rate in general, we rewrite (7) in terms of average transmission rate \bar{R} and power \bar{P} as $\Gamma_r(\bar{R}) = \frac{\bar{P}}{\bar{R} - \bar{R} \cdot P_{fer}(\frac{\bar{P}}{\bar{R}})}$.

An example with numerical demonstration of energy consumption rate, $\Gamma_r(R)$, is shown in Fig. 1, in which $RS(16, 8)$ code is used. Other parameters include channel gain $h_i = 10^{-2}$ [23], $P = 20$ dBm, and $N_o = 5 \times 10^{-9}$ WHz [22]. 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. For instance, 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”. This means that a decrease in bandwidth for “A” will result in very little

increase in energy consumption compared to the bandwidth decrease for “C”. And we can gain more reduction in energy consumption, ($\Delta \Gamma_C$), by increasing the bandwidth for terminal “C” compared to the bandwidth increase for “A”. Based on this observation, we propose a new admission control scheme and illustrate that, in reality, energy consumption can be reduced by bandwidth allocation.

Note that not only the number of channels may be changed during admission control, but also the coding scheme for the user needs to be changed. Therefore, the transmission rate is not changed continuously, but discretely. In this context, 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. In fact, in the simulation, we use discrete transmission rate to determine the necessary energy consumption. In addition, the factors affecting signal to interference and noise ratio are changing over time, such as distance between mobile users and Node B, multi-path fading and so on. In this context, we assume that power control schemes and adaptive coding schemes are used [10], [13], [15], thus average values can be used for the purpose of analysis as in other similar works on admission control.

IV. ENERGY-EFFICIENT BANDWIDTH ALLOCATION ALGORITHMS WITH QoS ADJUSTMENT

An energy-efficient, connection admission control scheme includes approaches to conserve energy, allocate bandwidth to mobile users upon requests and release multi-class services. In addition to minimizing the energy consumption of mobile terminals, we propose a QoS adjustment algorithm that takes requirements into account because, unlike single-class services, the interactions among multi-class services make a significant impact on QoS parameters.

For handling incoming requests, a *Victim Selection Algorithm* (VSA) is designed to minimize energy consumption. Upon the arrival of a new or a handoff request, if available bandwidth cannot meet the requirement, the BS may choose certain mobile terminals to acquire the potential bandwidth by decreasing bandwidth of those mobile terminals. Since the transmission rate of those terminals will be decreased, we name the chosen mobile terminals “victims.” After a connection is released, a certain bandwidth can be reassigned to the ongoing services to achieve a higher resource utilization. We call the mobile terminals that acquire extra bandwidth “beneficiaries” because their transmission rate will be increased. Accordingly, a *Beneficiary Selection Algorithm* (BSA) is proposed to achieve maximum energy conservation. In addition, to guarantee QoS and keep balance among all of the classes, we propose an algorithm to stochastically adjust bandwidth allocation for satisfying QoS requirements in multi-class systems.

A. Victim Selection Algorithm (VSA)

When a new or handoff request from a terminal, t , arrives at the BS, if the available bandwidth, BW_a , is greater than the minimum bandwidth requirement, $BWL[t]$, then the BS

```

1. While  $BW_a < BWL[t]$     *** Victim Selection Algorithm (VSA) ***
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  $BW[i] - \Delta BW \geq BWL[i]$ 
7.       then  $m \leftarrow \frac{d\Gamma_r(R)}{dR}$  of terminal  $i$ 
8.        $j \leftarrow i$ 
9.   if  $j == 1$  AND  $BW[1] - \Delta BW < BWL[1]$ 
10.    if  $t$  is a new request;           goto QoS Adjustment-Blocking
11.    else  $t$  is a handoff request      goto QoS Adjustment-Dropping
12.    else    $BW[j] \leftarrow BW[j] - \Delta BW$ ;    $BW_a \leftarrow BW_a + \Delta BW$ ;
13. ACCEPT  $t$ 
14. While  $BW_a > 0$     *** Beneficial Selection Algorithm (BSA) ***
15.   do  $m \leftarrow \frac{d\Gamma_r(R)}{dR}$  of terminal,  $i = 1$ 
16.    $j \leftarrow 1$ 
17.   for  $i \leftarrow 2$  to  $K$ 
18.     if for terminal  $i$ ,  $\frac{d\Gamma_r(R)}{dR} > m$ 
19.       AND  $BW[i] + \Delta BW \leq BWU[i]$ 
20.       then  $m \leftarrow \frac{d\Gamma_r(R)}{dR}$  of terminal  $i$ 
21.        $j \leftarrow i$ 
22.     if  $j == 1$  AND  $BW[1] + \Delta BW > BWU[1]$            break
23.      $BW[j] \leftarrow BW[j] + \Delta BW$ ;    $BW_a \leftarrow BW_a - \Delta BW$ 
24. return
25. ***** Stochastic Adjustment Algorithm (SAA) *****
26.  $NB[k] \leftarrow NB[k] + 1$ ;    $NTB[k] \leftarrow NTB[k] + 1$ 
27. if  $PB[k] \leq \beta_k$            return
28. else   while  $i \leq K$ 
29.     do   if  $PB[i] \leq PBU[i]$            break
30.     if  $i > K$            return
31.     else   randomly block class  $i$ 's request
32.      $NB[i] \leftarrow NB[i] + 1$ ;    $NTB[i] \leftarrow NTB[i] + 1$ 

```

Fig. 2. A Pseudocode of Energy-Efficient Bandwidth Allocation Scheme.

accepts this request. Otherwise, the VSA is triggered to select terminals whose derivative of energy consumption rate is the minimum, thus reducing the number of channels of victims, while minimizing the increase in energy consumption as shown in Fig. 2. 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.” On the contrary, if the bandwidth of this terminal can be decreased, then it becomes a “victim.” The BS will reduce the bandwidth, ΔBW , from a chosen victim, to a lower level to increase available bandwidth. 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 the QoS adjustment algorithm in Section IV-C to select terminals whose dropping/blocking probabilities are lower than their requirements.

B. Beneficiary Selection Algorithm (BSA)

Upon the bandwidth release of a completed connection, we can reallocate bandwidth to ongoing services; that is, to choose beneficiary terminals whose energy consumption can be reduced most with the increase in the bandwidth. 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. 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

other two terminals, “A” and “B.” A pseudocode of BSA is shown in Fig. 2.

C. Stochastic Adjustment Algorithm (SAA)

For multi-class service systems, bandwidth allocation needs to be coupled with QoS requirements to achieve a balanced QoS guarantee. Thus, we introduce a *stochastic adjustment algorithm* (SAA) in which 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. The BS reduces the system traffic load by sacrificing certain classes, thus creating a balance between the classes that beyond the QoS requirement and the classes that are under satisfied. In the implementation of SAA, the BS will pre-block/reduce traffic according to a stochastic process such as uniform distribution. For a blocked request of class k as shown in the pseudocode of SAA in Fig. 2, 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 in order to randomly block subsequent requests of class j . 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.

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. The proposed scheme can be directly applied to TDMA and FDMA systems, because those systems have fixed number of channels in terms of different time slots and frequency respectively.

For a CDMA system, the interference to one terminal is related to the transmission power of other users. Thus, the capacity is affected by the factors such as multi-path fading channels and interference including intra-cell and inter-cell interference. Here, we present an upper bound for the system capacity. Consider the system described in Section III. We have N mobile users and each user's SINR requirement is $\alpha_i, i = 1, 2, \dots, N$. In terms of a CDMA system, equation (1) becomes

$$\frac{l_i h_i P_i}{\sum_{j \neq i} h_j P_j + \eta_0 W} \geq \alpha_i, \quad i = 1, \dots, N. \quad (8)$$

where $l_i = W/R_i$ is the system processing gain. We can rewrite (8) as

$$l_i h_i P_i \geq \alpha_i (\sum_{j \neq i} h_j P_j + \eta_0 W), \quad i = 1, \dots, N. \quad (9)$$

Then, we have

$$l_i h_i P_i \geq \alpha_i (\sum_{j=1}^N h_j P_j - h_i P_i) + \alpha_i \eta_0 W, \quad i = 1, \dots, N. \quad (10)$$

Next,

$$(l_i + \alpha_i) h_i P_i - \alpha_i (\sum_{j=1}^N h_j P_j) \geq \alpha_i \eta_0 W, \quad i = 1, \dots, N. \quad (11)$$

In the specification of IS-95, α_i is 6 or 7 dB, which is far less than $l_i = W/R_i$. Therefore, we can simplify $l_i + \alpha_i \approx l_i$. Moreover, we assume that each user occupies only one channel and the transmission rate for each user is the same (say, R). Thus the system procession gain becomes $l = W/R$. By taking the summation of Equation (11) from 1 to N , we can obtain

$$l \sum_{j=1}^N h_j P_j - \sum_{i=1}^N (\alpha_i \sum_{j=1}^N h_j P_j) \geq \eta_0 W \sum_{i=1}^N \alpha_i, \quad i = 1, \dots, N. \quad (12)$$

That is

$$(l - \sum_{i=1}^N \alpha_i) \sum_{j=1}^N h_j P_j \geq \eta_0 W \sum_{i=1}^N \alpha_i, \quad i = 1, \dots, N. \quad (13)$$

Therefore, the upper bound of the number of users, N that a CDMA system can be achieved when condition $\sum_{i=1}^N \alpha_i < l$ holds. In fact, the change of system capacity may not affect the application of the proposed scheme because the transmission rate is mainly based on the number of channels because the

transmission rate increases linearly with the number of channels. Our proposed schemes are focused on the relationship between transmission rate and number channels that can be determined by underlying rate allocation and optimization schemes [24], [25], [26], [27]. Either victims or beneficiaries can use a certain transmission power as a result of power control schemes or optimization schemes of rate control. It is possible that the proposed CAC scheme can be combined with power control algorithms, which is one of our future work. Moreover, we note that for a CDMA system, different number of channels can be achieved by allocating multiple spreading codes to the mobile terminals [28].

V. PERFORMANCE ANALYSIS

In this context, we focus on two cases: *Fixed Channel Holding Time* and *Dynamic Channel Holding Time*. The former one is referred to as *Case I*, while the latter one is referred to as *Case II*. In the first case, the channel holding time is independent of the number of channels, which is an assumption widely deployed in existing work [5], [6]. It is appropriate to describe multimedia applications, such as on-line video transmission. For example, if we watch a 20-minute video clip online with more bandwidth, we will see a bigger picture and higher resolution. Thus, the bandwidth allocation affects video quality rather than transmission time. For *Case II*, channel holding time is dependent on the number of channels, which is often ignored by other works, but it can be applied to data transmission. For instance, a 100-KB file is to be downloaded to a laptop. With more bandwidth, the transmission time will be shorter and vice versa. The service rate is proportional to the number of channels. Our analysis is conducted based on the Continuous Time Markov Chain (CTMC). We first describe the Markovian state model, followed by the analysis of *Case I* and *Case II*. Then, we study the QoS impact of SAA, as well as the energy consumption for Poisson and self-similar traffic.

A. Markovian State Description

In addition to the notations described in Section II, we denote vector $\vec{N} = [n_1, n_2, \dots, n_K]$ 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 $[n_1, n_2, \dots, n_K]$ in the system, and the state space is the set of all possible states, which is denoted by \mathbf{S} as:

$$\mathbf{S} := \{ \vec{N} \in \mathcal{I}^K : \sum_{k=1}^K \sum_{i=1}^{n_k} b_k^i \leq C \} \quad \text{and} \quad b_k^i \in [b_k^L, b_k^U] \quad (14)$$

where \mathcal{I} is a set of non-negative integers and \mathcal{I}^K is a set of K -dimensional non-negative vectors. In order to consider the effect of SAA to QoS requirements, we let \mathbf{S}_k be a subset of the state space for which an arriving request of class k is blocked, that is

$$\mathbf{S}_k := \{ \vec{N} \in \mathcal{I}^K : C - b_k^L < \sum_{i=1}^K (b_i^L \times n_i) \leq C \}. \quad (15)$$

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

$$I_k^+(\vec{N}) = \begin{cases} 1 & \text{if } \vec{N} + \vec{E}_k \in \mathbf{S} \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

and

$$I_k^-(\vec{N}) = \begin{cases} 1 & \text{if } \vec{N} - \vec{E}_k \in \mathbf{S} \\ 0 & \text{otherwise.} \end{cases} \quad (17)$$

Let $q(\vec{N}_1, \vec{N}_2)$ denote the probability transition rate from state \vec{N}_1 to \vec{N}_2 , and $\vec{\tau} = [\tau_1, \tau_2, \dots, \tau_K]$ 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 due to the number of channels. Otherwise, the rate factor equals the number of channels, which is applicable to *Case II*. Therefore, we have

$$\begin{aligned} q(\vec{N}, \vec{N} + \vec{E}_k) &= \lambda_k(\vec{N}, \vec{N} + \vec{E}_k \in \mathbf{S}) \\ q(\vec{N}, \vec{N} - \vec{E}_k) &= n_k \tau_k \mu_k(\vec{N}, \vec{N} - \vec{E}_k \in \mathbf{S}) \\ q(\vec{N} - \vec{E}_k, \vec{N}) &= \lambda_k(\vec{N} - \vec{E}_k, \vec{N} \in \mathbf{S}) \\ q(\vec{N} + \vec{E}_k, \vec{N}) &= (n_k + 1) \tau_k \mu_k(\vec{N} + \vec{E}_k, \vec{N} \in \mathbf{S}) \end{aligned} \quad (18)$$

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

B. Blocking Probabilities Without Stochastic QoS Adjustment

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* both with and without QoS adjustment algorithm in this section and the subsequent one.

1) *Analysis for Case I*: 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, 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

$$\begin{aligned} & \left[\sum_{k=1}^K \lambda_k I_k^+(\vec{N}) + \sum_{k=1}^K n_k \tau_k \mu_k I_k^-(\vec{N}) \right] P(\vec{N}) \\ &= \sum_{k=1}^K \lambda_k I_k^-(\vec{N}) P(\vec{N} - \vec{E}_k) \\ &+ \sum_{k=1}^K (n_k + 1) \tau_k \mu_k I_k^+ P(\vec{N} + \vec{E}_k), \end{aligned} \quad (19)$$

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

$$\begin{aligned} \lambda_k I_k^- P(\vec{N} - \vec{E}_k) &= n_k \tau_k \mu_k I_k^- P(\vec{N}) \\ k &= 1, \dots, K, \quad \forall \vec{N} \in \mathbf{S} \end{aligned} \quad (20)$$

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

$$P(\vec{N}) = \frac{1}{G(\mathbf{S})} \prod_{k=1}^K \frac{\rho_k^{n_k}}{\tau_k^{n_k} n_k!}, \quad (21)$$

where

$$G(\mathbf{S}) = \sum_{\vec{N} \in \mathbf{S}} \prod_{k=1}^K \frac{\rho_k^{n_k}}{\tau_k^{n_k} n_k!} \quad \text{for } \vec{N} \in \mathbf{S}. \quad (22)$$

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 (15). Hence, the blocking probability of class k , $B_k(\vec{\tau})$, can be written as

$$B_k(\vec{\tau}) = \frac{G(\mathbf{S}_k)}{G(\mathbf{S})}, \quad \text{and} \quad \vec{\tau} = [1, 1, \dots, 1] \quad (23)$$

where \mathbf{S}_k and \mathbf{S} are given in (14) and (15), respectively.

Such product form of blocking probability given in (23) holds under the assumption that the channel holding time is irrelevant to the number of channels, which is explained at the beginning of this section. The assumption that the channel holding time is *exponentially* distributed with mean $1/\mu_k$ can be expanded to any *arbitrary* distribution with mean $1/\mu_k$ by so called "insensitivity" property [29]. For large systems, however, there is a potential numerical problem in computing blocking probabilities, which is the fact pointed out in traffic engineering [30]. Since the blocking probabilities for multi-class traffic has a product form solution, in some cases, the computation of blocking probabilities is slow and shows unstable numerical results. In order to avoid this problem, we calculate blocking probabilities step by step in a recursive way to achieve numerically stable computation.

2) *Analysis for Case II*: Here we consider that the service rate is proportional to the number of channels found in many wireless systems. We thereby denote the service rate of one channel for class k is μ_k . Then, the overall service rate for class k , $\mu_k(\vec{N})$, is equal to the summation of the service rate of all of the on-going connections in class k , which is

$$\mu_k(\vec{N}) = \sum_{i=1}^{n_k} \mu_k b_k^i, \quad \forall \vec{N} \in \mathbf{S} \quad (24)$$

where $b_k^i \in [b_k^L, b_k^U]$ and \mathbf{S} as given in (14).

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

$$\mu_k^U(\vec{N}) \leq \sum_{i=1}^{n_k} \mu_k b_k^U = n_k b_k^U \mu_k, \quad (25)$$

and

$$\mu_k^L(\vec{N}) \geq \sum_{i=1}^{n_k} \mu_k b_k^L = n_k b_k^L \mu_k. \quad (26)$$

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 ; that is a semi-Markov process is simplified to a CTMC. Note that the vector $\vec{\tau}$ is not a constant for *Case II* because the channel holding time is related to the number of

channels. For lower and upper bounds, we denote it as B_K^L and B_K^U , which correspond to lower and higher bandwidth requirements, respectively. Therefore, in a similar way as in the previous subsection, we can obtain the upper and lower bounds of blocking probability as

$$B_k^L(\vec{\tau}_L) = \frac{G(\mathbf{S}_k)}{G(\mathbf{S})}, \quad \text{and} \quad \vec{\tau}_L = [b_1^L, b_2^L, \dots, b_K^L]. \quad (27)$$

The upper bound, B_k^U , can be obtained by the similar expression with the exception of substituting $\vec{\tau}_U$ for $\vec{\tau}_L$ in the above equation.

C. Blocking Probabilities with Stochastic QoS Adjustment

When the blocking probability of class i cannot satisfy the QoS requirement, i.e., the blocking probability of class i is greater than β_i , a stochastic adjustment algorithm (SAA) shown in Fig. 2 is applied, in which some traffic will be blocked on purpose to reduce the traffic load of the whole system. Let vector $\vec{M} = [m_1, \dots, m_k, \dots, m_K]$ be the rate of arrival requests blocked for each class, where $m_k \geq 0$ is the blocked traffic for class k . We denote an arrival Poisson process as $\{N(t), t \geq 0\}$ and the process after pre-blocking as $\{N_b(t), t \geq 0\}$. Therefore, we justify that $\{N_b(t), t \geq 0\}$ is a Poisson process with the arrival rate $\lambda_k - m_k$ by first condition on $N(t)$ and then simplify the denotation. The actual traffic rate for class k after pre-blocking is $\rho_k = \frac{\lambda_k - m_k}{\mu_k}$. Then, the blocking probability of class k given applying stochastic QoS adjustment, $B_k|\mathbf{Q}$, can be written

$$B_k|\mathbf{Q} = B_k + \frac{m_k}{\lambda_k}, \quad (28)$$

where B_k is given in (23) with the new traffic load. Similarly, we can obtain the upper and lower bounds of blocking probability for *Case II* by applying the SAA for QoS adjustment. Next, we discuss how to determine $\vec{M} = [m_1, m_2, \dots, m_K]$ through a pricing model. Let us define the price of class k as Π_k , which is a function of blocking probability, B_k , given by

$$\Pi_k(m_1, m_2, \dots, m_K) = \mathbb{F}_k(B_k|\mathbf{Q}), \quad k = 1, \dots, K, \quad (29)$$

where $\mathbb{F}_k(\cdot)$ is the price function for class k , which depends on service requirements for each class. Accordingly, the objective function of stochastic QoS adjustment is to find vector \vec{M} such that

$$\begin{aligned} \min \quad & \{ \sum_{k=1}^K \Pi_k(m_1, m_2, \dots, m_K) \} \\ \text{s.t.} \quad & \text{constraint } X \text{ for } m_k \geq 0, \quad k = 1, \dots, K \end{aligned} \quad (30)$$

Therefore, the solution of \vec{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

$$\Pi_1(m_1, m_2) = \alpha B_1|\mathbf{Q}, \quad \text{and} \quad \Pi_2(m_1, m_2) = B_2|\mathbf{Q}. \quad (31)$$

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

$$\begin{aligned} \min \quad & \{ \alpha B_1|\mathbf{Q} + B_2|\mathbf{Q} \} \\ \text{s.t.} \quad & B_1|\mathbf{Q} \leq B_2|\mathbf{Q} \text{ for } m_1 \geq 0, \quad m_2 \geq 0. \end{aligned} \quad (32)$$

D. Energy Consumption

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

$$\mathbb{E} = \sum_{\vec{N} \in \mathbf{S}} P(\vec{N}) \times E(\vec{N}) \quad (33)$$

where $P(\vec{N})$ in (21) can be obtained by substituting different $\vec{\tau}$ for *Case I* and *Case II* as discussed in Section V-B. In order to acquire $E(\vec{N})$, we define the energy consumption rate (ECR) matrix as

$$ECR = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1C} \\ a_{21} & a_{22} & \dots & a_{2C} \\ \vdots & \vdots & \ddots & \vdots \\ a_{K1} & a_{K2} & \dots & a_{KC} \end{pmatrix} \quad (34)$$

where each element, a_{ij} , is the energy consumption rate of i^{th} class with j channels, which can be obtained directly by energy consumption rate $\Gamma_r(R)$, defined in Section III. In wireless systems, the BS determines the number of channels and transmission rate for a connection request upon system specifications [31]. Therefore, based on energy consumption rate $\Gamma_r(R)$ and the mapping of the number of channels and transmission rate, a_{ij} can be obtained for j serving channels.

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

$$\min_{k=1, \dots, K} \left\{ \sum_{k=1}^K \sum_{i=1}^{n_k} a_{kj} \right\} \quad \text{and} \quad E(\vec{N}) = \frac{\sum_{k=1}^K \sum_{i=1}^{n_k} a_{kj}}{\sum_{k=1}^K n_k}. \quad (35)$$

Then, by substituting $E(\vec{N})$ in (33), the total energy consumption can be obtained. For simplification, we assume a constant transmission rate for a certain number of channels, whereas in reality, these rates may be varied from time to time. The computation of the above solution would not be scalable for a large number of classes. However, in real systems, there are usually about 3 or 4 classes of traffic, and the number of channels for each class is not very large [31]. Therefore, the proposed solution is applicable to real systems.

VI. SIMULATION RESULTS

The results of the proposed bandwidth allocation scheme are compared with other schemes such as Non-Prioritized Scheme (NPS) [4] and Adaptive Resource Allocation Scheme (AREAS) [32]. The basic idea of NPS is that the number of channels is not changed during the transmission, which is the most simplest scheme with the smallest overhead. Thus, NPS become one of the candidate of wireless communications systems; and AREAS can dynamic the number of channels according to the traffic load with three distinct levels.

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 |

In addition to the two cases of traffic discussed in Section V, 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 [33] due to the fact that traffic pattern of multimedia services in wireless networks tends to be LRD [34]. Therefore, we examine self-similar traffic as well. To avoid repetition, we do not differentiate between handoff requests and new requests because they are handled in a similar way to that described in Section IV. We simulate the system with a total of $C = 100$ channels, which is about the capacity of a macrocell system. In our experiments, we consider two ($K = 2$) classes of service, but the effectiveness of the proposed scheme will not be affected by three or four classes. The other parameters used in our simulation are shown in Table I.

The call arrival rate of Class A varies from 1 call/min to 40 calls/min and the call arrival rate of Class B remains the same as 10 calls/min. The amount of data to be transmitted is the same for both classes, uniformly distributed [60,180] Kbytes for consistence. The transmission rate of each channel is 8 Kbps for both Classes. The number of channels for NPS is 7 for Class A and 3 for Class B. For AREAS, the number of channel can be varied with 4, 7 or 9 for Class A and with 2 or 5 for Class B. For our scheme, the number of channels can be used from 4 up to 9 for Class A and 2 up to 5 for Class B. As to the self-similar traffic, the number of sources for each class is 100. The traffic density parameter is from 0.1 to 1.5 for Class A and 1 for Class B. The power-law parameter A and γ are 500 and 1.5 for both classes.

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. 3. We can observe only four curves in the figure, although in fact, there are results of eight scenarios as the increase of λ_A , because 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 one other. Moreover, the blocking probabilities of the new scheme and AREAS for both classes overlap, which results in only four separate curves in Fig. 3. 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

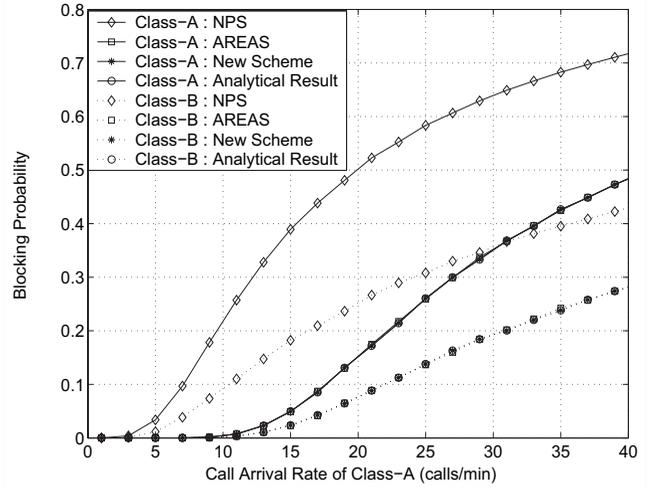


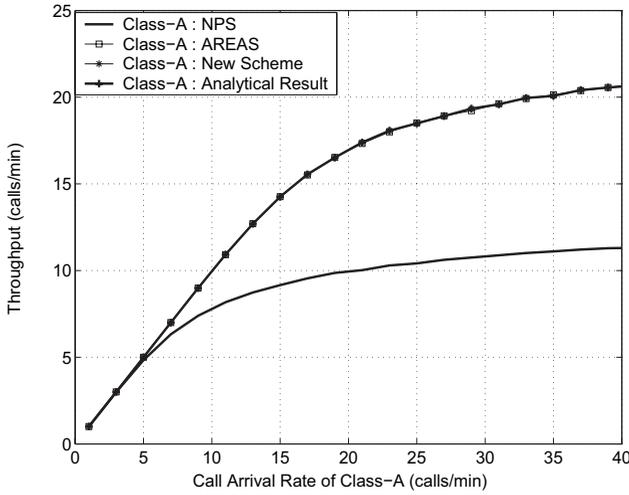
Fig. 3. Case I: Call Blocking Probability.

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. This is because the new scheme can adapt the channel allocation in accordance to the arrival traffic.

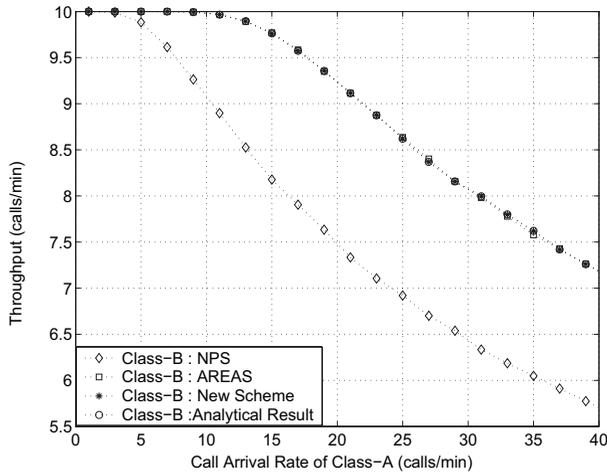
The throughput of two classes as a function of λ_A , the arrival rate of Class-A, is plotted in Fig. 4. Even though the blocking probability is increased, the throughput of three schemes in Class-A is increased, whereas for Class-B, the throughput is decreased because of the constant arrival rate and increasing arrival rate of Class-A traffic. Fig. 5 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. By taking energy consumption into consideration, the new scheme enables the systems to allocate the channels to the arrival traffic, which can yield better energy conservation at the expense of throughput.

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 is compared in Fig. 6 as the increase of λ_A . Unlike Fig. 3, the blocking probability of the new scheme is lower than that of AREAS for Case II, which means that the new scheme is better. In



(a) Class-A.



(b) Class-B.

Fig. 4. Throughput in *Case I*.

our proposed scheme, new released channels are allocated to ongoing services to speed up their transmission, thus reducing call blocking probability. Although AREAS has the capability of dynamical allocation, it has only 3 level, which limits its performance.

Similar observations can also be found for Class-B, which is omitted due to page limit. The upper and lower bounds of blocking probability, along with the simulation results of the proposed scheme are shown in Fig. 7 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.

The throughput as a function of λ_A are shown in Fig 8. For Class-B, the new scheme yields a higher throughput compared to the other two schemes, but decreasing with the increase in arrival rate of Class-A because of the increase in blocking probability and constant arrival rate. As a matter of fact, the throughput of Class-A increases according to our results, even though the figures are omitted. The energy consumption is not shown here because it is similar to that for *Case I*. This means that the energy consumption is quite robust to the variation of the incoming traffic, although there is a significant

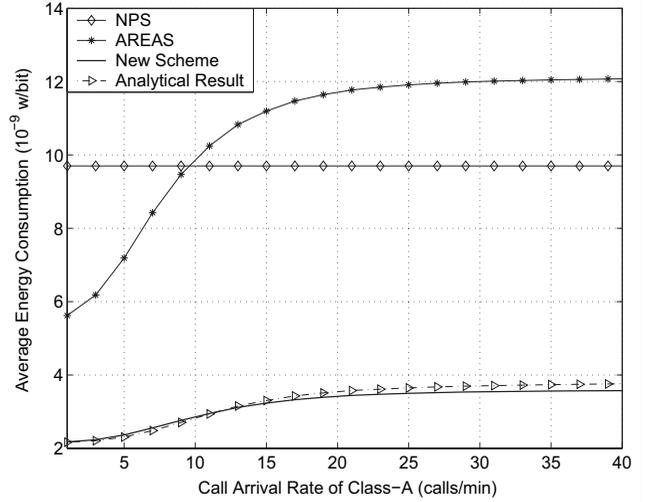


Fig. 5. *Case I*: Average Energy Consumption.

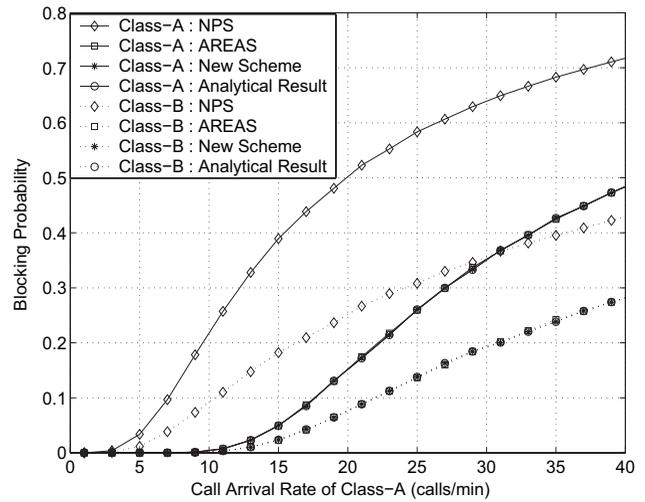


Fig. 6. *Case II*: Call Blocking Probability of Class-A.

difference in the blocking probability. Therefore, for *Case II*, the energy-efficient bandwidth allocation scheme yields better performance in blocking probability, throughput and energy consumption.

C. Stochastic QoS Adjustment

Here, we present the simulation results to show the effectiveness of “Stochastic” QoS adjustment for *Case I* only, because the results for *Case II* are very similar. Fig. 9 displays the result of stochastic QoS adjustment satisfying the pricing model in Section V-C. 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. 9 is quite different. By sacrificing Class-B, the blocking probability

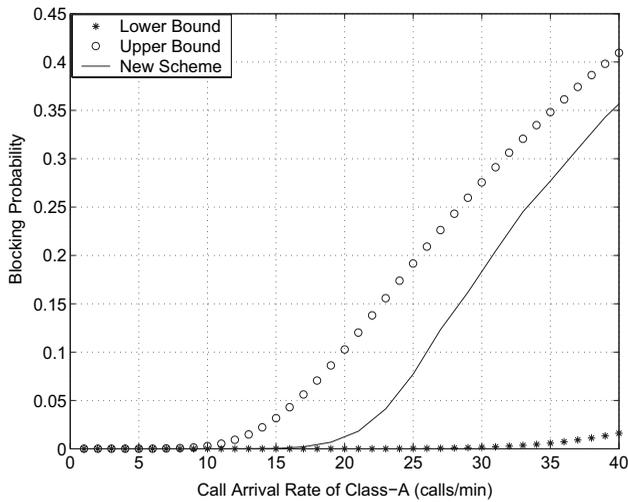


Fig. 7. Case II: Bounds of Blocking Probability of Class-A.

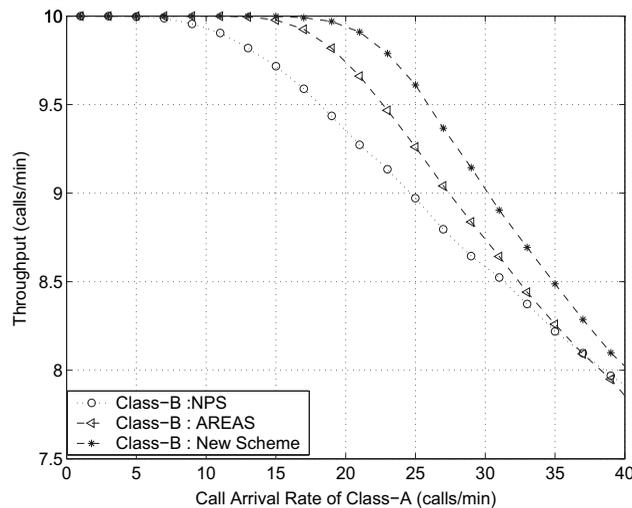


Fig. 8. Case II: Throughput of Class-B.

of Class-A has a significant decrease. A close look at Fig. 9 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 since the system can deploy different level of adjustment parameters for the best overall system performance.

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.

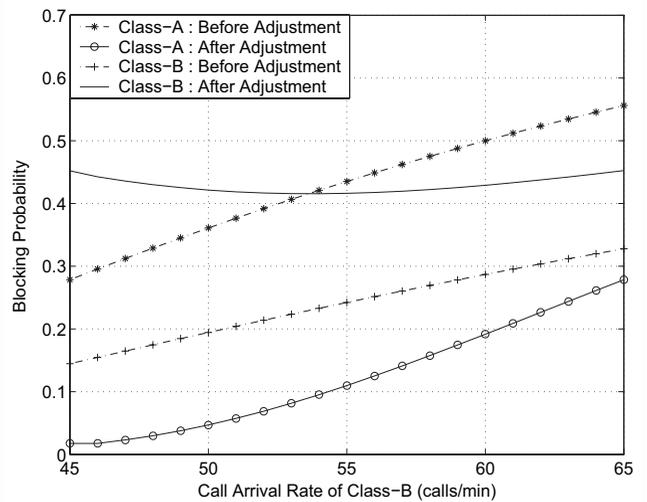


Fig. 9. QoS Adjustment with Pricing Constraint.

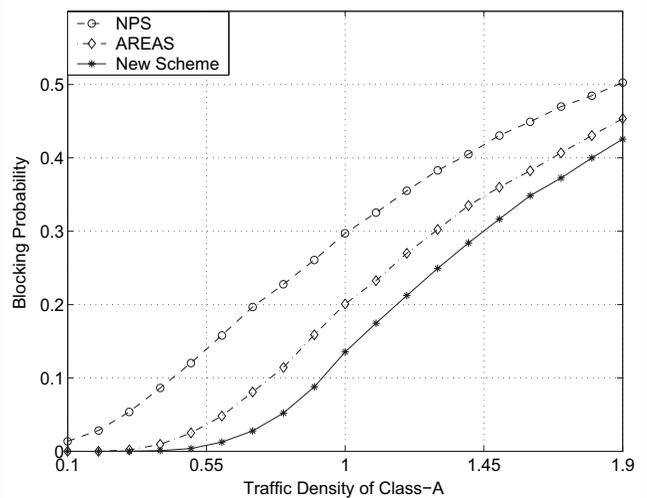


Fig. 10. Self-Similar Traffic: Blocking Probability of Class-A.

D. Self-Similar Arrival Process

We examine the performance of the proposed scheme by applying to self-similar traffic as well. Without losing generality, we deploy Power-on and Power-off mode to generate self-similar traffic and the power-law distribution. Figs. 10 and 11 depict the blocking probability and throughput of Class-A as the increase of traffic density parameter, R , of Class-A from 0.5 to 1.9. Since the results are very similar to those of *Case I* and *Case II* for Class-B, we do not show them separately here. It is easily noticed that the new scheme results in lower blocking probability and higher throughput. However, for self-similar arrival traffic, we take a longer time to obtain stable results than Poisson arrival traffic. The average energy consumption for self-similar traffic is very similar to that for *Case I* and *Case II*. Therefore, we can conclude that the effectiveness of energy conservation is not affected significantly by traffic patterns so that the proposed scheme is applicable to various traffic. To avoid repetition, we do not show the figure of energy consumption here again. Thus, with the arrival process of self-similarity, the proposed

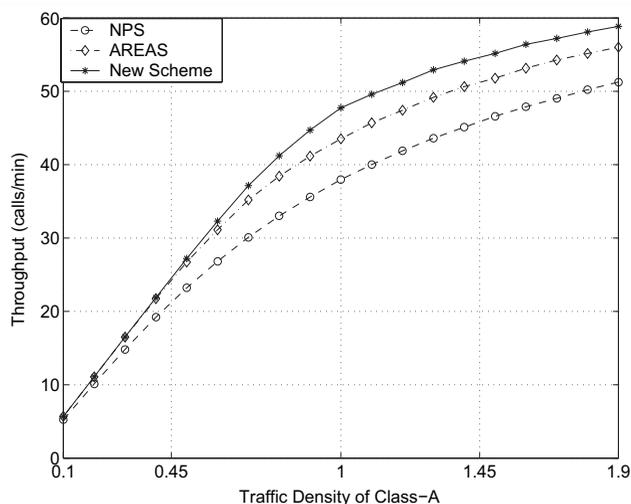


Fig. 11. Self-Similar Traffic: Throughput of Class-A.

scheme generates better performance than AREAS and NPS with regard to call blocking probability, energy consumption and throughput due to the improved utilization of channels via reallocating channels to active connections for expediting transmissions.

VII. CONCLUSIONS

We proposed a new bandwidth allocation scheme in which energy consumption is considered because battery life and radio resource are critical in mobile communications. The proposed scheme consists of three parts: Victim Selection Algorithm (VSA), Beneficiary Selection Algorithm (BSA), and Stochastic Adjustment Algorithm (SAA) with the aim to reduce energy consumption without sacrificing QoS requirements. The performance of the new scheme is analyzed with respect to call blocking probability, throughput, and energy consumption. The effectiveness of the proposed scheme is verified through extensive simulations for a variety of traffic patterns in multi-class systems, showing better energy conservation and system performance.

REFERENCES

- [1] M.-H. Chiu and M. Bassiouni, "Predictive schemes for handoff prioritization in cellular networks based on mobile positioning," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 510–522, Feb. 2000.
- [2] 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*, vol. 4, pp. 2113–2117, 2001.
- [3] W. Hou and Y. Fang, "Mobility-based call admission control schemes for wireless mobile networks," *Wireless Communications and Mobile Computing*, vol. 1, pp. 269–282, July–Sept. 2001.
- [4] L. Ortigoza-Guerrero and A. H. Aghvami, "A prioritized handoff dynamic channel allocation strategy for PCS," *IEEE Trans. Veh. Technol.*, vol. 48, pp. 1203–1215, July 1999.
- [5] P. Ramanathan, K. M. Sivalingam, P. Agrawal, and S. Kishore, "Dynamic resource allocation schemes during handoff for mobile multimedia wireless networks," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1270–1283, July 1999.
- [6] T. Zhang, E. Berg, J. Chennikara, P. Agrawal, J.-C. Chen, and T. Kodama, "Local predictive resource reservation for handoffs in multimedia wireless IP networks," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 1931–1941, Oct. 2001.

- [7] I. Panoutsopoulos, S. Kotsopoulos, and V. Tountopoulos, "Handover and new call admission policy optimization for 3G systems," *ACM-Baltzer Journal of Wireless Networks*, vol. 8, no. 4, pp. 381–389, 2002.
- [8] N. Bambos and S. Kandukuri, "Power-controlled multiple access schemes for next-generation wireless packet networks," *IEEE Wireless Communications*, pp. 58–64, June 2002.
- [9] R. Berry and R. Gallager, "Communication over fading channels with delay constraints," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1135–1149, May 2002.
- [10] R. Rao, C. Comaniciu, T. Lakshman, and H. Poor, "Call admission control in wireless multimedia networks," *IEEE Signaling Processing Mag.*, vol. 21, pp. 51–58, September 2004.
- [11] C. U. Saraydar, N. B. Mandayam, and D. J. Goodman, "Efficient power control via pricing in wireless data networks," *IEEE Trans. Commun.*, vol. 50, pp. 291–303, Feb 2002.
- [12] X. Wang, R. Ramjee, and H. Viswanathan, "Adaptive and predictive downlink resource management in next generation CDMA networks," in *Proc. IEEE INFOCOM'2004*, June 2004.
- [13] H. Lin, M. Chatterjee, S. Das, and K. Basu, "ARC: an integrated admission and rate control framework for cdma data networks based on non-cooperative games," in *ACM/IEEE MobiCom'2003*, Sept. 2003.
- [14] S. Shen, C. Chang, C. Huang, and Q. Bi, "Intelligent call admission control for wideband CDMA cellular systems," *IEEE Wireless Commun.*, vol. 3, pp. 865–878, Sept. 2004.
- [15] N. Bambos, "Toward power-sensitive network architectures in wireless communications: concepts, issues, and design aspects," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 966–976, June 2000.
- [16] M. Zorzi, "Outage and error events in bursty channels," *IEEE Trans. Commun.*, vol. 46, pp. 349–356, June 1998.
- [17] Q. Zhang, Z. Ji, W. Zhu, and Y. Zhang, "Power-minimized bit allocation for video communication over wireless channels," *IEEE Trans. Circuits and Syst. Video Technol.*, vol. 12, pp. 398–410, June 2002.
- [18] R. Kravets and P. Krishnan, "Application-driven power management for mobile communication," *ACM-Baltzer Journal of Wireless Networks*, vol. 6, pp. 263–277, 2000.
- [19] 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*, vol. 1, pp. 386–394, April 2001.
- [20] B. M. Epstein and M. Schwartz, "Predictive QoS-based admission control for multiclass traffic in cellular wireless networks," *IEEE Journal Select. Areas Commun.*, vol. 18, pp. 523–534, Mar 2000.
- [21] E. U. Biyikoglu, B. Prabhakar, and A. E. Gammal, "Energy-efficient packet transmission over a wireless link," *IEEE Trans. Networking*, vol. 10, pp. 487–499, Aug 2002.
- [22] J. G. Proakis, "Digital Communication," *McGraw-Hill*, 2002.
- [23] J. Mark and W. Zhuang, *Wireless Communications and Networking*. Upper Saddle River: Prentice Hall, NJ, 2003.
- [24] E. M. M. J. Neely and C. E. Rohrs, "Power allocation and routing in multibeam satellites with time-varying channels," *IEEE/ACM Trans. Networking*, vol. 11, pp. 138–152, February 2003.
- [25] D. N. C. Tse and S. V. Hanly, "Multiaccess fading channels part I: polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2796–2815, t 1998.
- [26] D. N. C. Tse and S. V. Hanly, "Multiaccess fading channels part II: polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2816–2831, Nov. 1998.
- [27] X. Lin and N. B. Shroff, "The impact of imperfect scheduling on cross-layer rate control in wireless networks," in *Proc. IEEE INFOCOM'05*, March 2005.
- [28] A. Huber and U. Huber, *UMTS and Mobile Computing*. Boston: Artech House, 2002.
- [29] V. A. Sposito, *Linear and Nonlinear Programming*. The Iowa State University Press, Ames Iowa 50010, 1975.
- [30] A. A. Nilsson and M. Perry, "Multi-rate blocking probabilities: numerically stable computations," in *International Telecommunications Conference (ITC)15*, vol. 2, pp. 1359–1368, 1997.
- [31] S. Nanda, K. Balachandran, and S. Kumar, "Adaptation techniques in wireless packet data service," *IEEE Commun. Mag.*, vol. 38, no. 1, pp. 54–64, 2000.
- [32] 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 Commun.*, vol. 7, pp. 42–47, Dec 2000.
- [33] 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.

- [34] Y. Fang and Y. Zhang, "Call admission control scheme and performance analysis in wireless mobile networks," *IEEE Trans. Vehic. Technol.*, vol. 51, pp. 371–382, Mar. 2002.



Wenye Wang (M'98/ACM'99) received the B.S. and M.S. degrees from Beijing University of Posts and Telecommunications, Beijing, China, in 1986 and 1991, respectively. She also received the M.S.E.E. and Ph.D. degree from Georgia Institute of Technology, Atlanta, Georgia in 1999 and 2002, respectively. She is now an Assistant Professor with the Department of Electrical and Computer Engineering, North Carolina State University. Her research interests are in mobile and secure computing, quality-of-service (QoS) sensitive networking

protocols in single- and multi-hop networks. She has served on program committees for IEEE INFOCOM, ICC, ICCCN in 2004. Dr. Wang is a recipient of NSF CAREER Award in 2006. She has been a member of the Association for Computing Machinery since 2002.



Xinbing Wang (SM'01) received the B.S. degree from the Department of Automation, Shanghai Jiaotong University, Shanghai, China, in 1998, and the M.S. degree from the Department of Computer Science and Technology, Tsinghua University, Beijing, China, in 2001. He is currently a Ph.D. candidate at the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC.

His research interests include TCP Asymptotics Analysis; Wireless Capacity; Cross Layer Call

Admission Control; Asymptotics Analysis of Hybrid systems; Congestion Control over Wireless Networks.



Arne A. Nilsson received the Master of Electrical Engineering degree and the Ph.D. degree in telecommunication systems from the Lund University of Technology, Lund, Sweden, in 1968 and 1976, respectively.

He is currently a Professor with the Department of Electrical and Computer Engineering, North Carolina State University (NCSU), Raleigh. He has more than 30 years of experience with telecommunication networks. He was a Principal Investigator for the Swedish Telecommunication Administration in the effort to build a Nordic packet switched network. He joined the faculty at NCSU in 1978. In 1986, he was a Professor at the Aeronautics Institute of Technology (ITA), San Jose dos Campos, Brazil. He was on the committee responsible for the design and architecture of the MCNC data and video network in North Carolina, and was a Principle Investigator on the VISTAnet, one of 5-Gb test beds. He has been an active participant in the industry/university cooperative research center on communications and signal processing at NCSU, Center for Communications and Signal Processing (CCSP), and the Center for Advanced Computing and Communication (CACC). He has recently become an External Evaluator for some of the European Efforts in high-speed networking. He has also worked for the U.S. Army Research Office in the wireless network engineering area.