Mobile Effective Bandwidth and Its Application to Admission Control in Cellular Packet Networks

Majid Ghaderi[†] and Raouf Boutaba
School of Computer Science
University of Waterloo
Waterloo, ON N2L 3G1, Canada
{mghaderi,rboutaba}@uwaterloo.ca

Abstract

Admission control is a key element for providing quality of service in a mobile cellular network. The problem of admission control in packet-switched cellular networks is more challenging compared to their circuit-switched counterparts due to packet-level network interactions. Although packet-based architectures allow for more efficient sharing of scarce radio resources, admission of new connections into the network is more complicated. The concept known as effective bandwidth is a simple yet efficient approach for admission control and resource allocation in wireline networks. This paper introduces the notion of mobile effective bandwidth which extends the classical effective bandwidth concept introduced for wireline networks to cellular packet networks. The main idea is to consider the spacial multiplexing due to user mobility in computing an effective bandwidth for variable-bit-rate connections. Based on this concept, an admission control algorithm is proposed which guarantees a prespecified packet loss probability while achieving zero connection dropping probability. This is a sharp diversion from the existing admission control schemes which can not completely eliminate connection dropping because of the underlying circuit-based architecture.

[†] corresponding author

Mobile Effective Bandwidth and Its Application to Admission Control in Cellular Packet Networks

I. INTRODUCTION

Emerging wireless technologies such as 3G [1] and 4G [2] tend to be packet-switched rather than circuit-switched because the packet-based architecture allows for better sharing of limited wireless resources. In a packet network, connections (packet flows) do not require dedicated circuits for the entire duration of the connection. Unfortunately, this enhanced flexibility makes it more difficult to effectively control the admission of connections into the network.

The problem of connection admission control (CAC) in a packet network can be addressed by a concept known as the *effective bandwidth* [3]–[5]. When employing this concept, an appropriate effective bandwidth is assigned to each connection and each connection is treated as if it required this effective bandwidth throughout the active period of the connection. The computation of an effective bandwidth is based on desired packet-level quality of service (QoS), e.g. packet loss probability, delay and jitter. Using effective bandwidths, the problem of admission control in a packet-switched network is transferred to the problem of admission control in a circuit-switched network.

In wireless circuit-switched networks, two important parameters which determine the connection-level quality of service are *connection blocking probability* and *connection dropping probability*. When a mobile terminal requests service, it may either be granted or denied service. This denial of service is known as connection blocking. An active terminal in a cellular network may move from one cell to another by executing a *handoff* procedure. Failure to get a successful handoff at any cell in the path forces the network to discontinue service to the user. This is known as connection dropping. In general, dropping a connection in progress is considered to have a more negative impact from the user's perspective than blocking a newly requested connection. *A challenging problem in cellular networks is to devise proper mechanisms for offering a guaranteed upper bound on connection dropping probability*.

As mentioned before, the CAC problem in a packet-switched network can be transferred to the CAC problem in a circuit-switched network using the notion of effective bandwidth. Applying this transformation, it is left for CAC algorithm to address only connection-level QoS requirements. In fact, most of the researchers in wireless networking field have focused only on connection-level quality of service parameters for admission control and resource allocation [6]–[10] because the primary concern has been voice traffic support in a circuit-switched cellular network.

A major difference between our approach and existing approaches is that we take into consideration two characteristics of wireless packet networks in allocating resources:

1) *Statistical multiplexing:* per packet bandwidth allocation feature of packet-based networks,
2) *Spacial multiplexing:* mobility of users which leads to connection mobility in a wireless network.

Combining these features, we are able to develop the notion of *mobile effective bandwidth* for admission control in cellular packet networks to address both connection-level and packet-level quality of service.

The rest of the paper is organized as follows. Section II reviews the related work in the area. Our system model, assumptions and notations are described in section III. Section IV presents the analysis for computing the mobile effective bandwidth. Then, in section V, an admission control algorithm is proposed based on the computed mobile effective bandwidth. Finally, section VI concludes this paper.

II. BACKGROUND AND MOTIVATION

A. Background on Effective Bandwidth

For variable-bit-rate (VBR) traffic, the amount of bandwidth needed by each connection varies over time. The actual required bandwidth fluctuates between some minimal rate and a peak rate. The effective bandwidth of a connection is some value between its average rate and its peak rate.

To avoid complicated exact analysis which does not provide any useful result, approximate techniques for computing effective bandwidth of a connection or aggregate of connections have been developed. For the infinite buffer regime fluid-flow analysis has been applied [11], [12] while for the bufferless regime large deviation techniques [4], [13] or stationary approximations [5], [14] have been successfully used.

In this paper, we are interested in stationary techniques based on Gaussian approximation as developed in [5]. It is the simplicity and effectiveness of these techniques which make them a good candidate for real-time admission control in cellular networks. To apply stationary approximations it is assumed that the number of connections sending bursts of packets remain fixed long enough to reach steady-state equilibrium. This assumption is embedded in in fluid-flow and large deviations analysis too. The fundamental issue in cellular networks is the mobility of users which makes such an assumption unrealistic and if applied will result in higher packet loss probability than what was predicted.

B. Background on Admission Control

Connection admission control has been extensively studied in circuit-switched cellular networks (see [15], [16] and references there in). Hong and Rappaport [6] are the first who systematically analyzed the famous *guard channel* (GC) scheme, which is currently deployed in cellular networks supporting voice calls. Ramjee et al. [7] have formally defined and categorized the admission control problem in cellular networks. They showed that the guard channel scheme is optimal for minimizing a linear objective function of connection blocking and dropping probabilities. To capture the global effect of user mobility, collaborative or distributed admission control schemes have been proposed (see for example [8]–[10]). Information exchange among a cluster of neighboring cells is the approach adopted by all distributed schemes.

The main objective of these CAC schemes is to provide connection-level QoS in terms of guaranteed connection dropping probability. Most of them are based on the reservation paradigm in which a number of available channels (circuits) are reserved to be exclusively used by handoffs. Since the number of channels is limited in a circuit-switched network, if a handoff request arrives while all channels are in use then the handoff request is blocked and the connection is forced to terminate. The fundamental issue is that packet-switched networks have virtually unlimited number of channels. Therefore, in a packet-switched network it is always possible to accept the handoff request at the expense of probably increasing the number of dropped packets. While this approach completely eliminates the call dropping event, we will show that its impact on packet loss can be effectively controlled.

C. Motivation

There are several important issues in emerging wireless packet networks that existing admission control schemes have failed to address. An important issue is that existing CAC schemes address only connection-level QoS. Applying the notion of effective bandwidth, it is possible to embed packet-level QoS in these schemes. But even this approach is not without problem:

1) Using the existing techniques for computing effective bandwidth is based on the assumption that the number of connections will remain fixed. In wireless networks where users are mobile, this will underestimate the variations in aggregated network traffic. Intuitively, ignoring the variations in the number of connections reduces the uncertainty associated with the stochastic process describing the network traffic. In particular, equation (16) shows that static treatment of the number of connections will underestimate the variance of the packet arrival process.
2) Since the number of connections varies over time in a wireless network, computing the effective bandwidth for the worst case scenario, maximum number of connections, will degrade the bandwidth utilization. In a situation with variable number of connections, the actual packet loss probability is in fact an *average* of the packet loss probability conditioned on the number of connections. Let $P_L(n)$ denote the packet loss probability when there are n connections in the network. Then the average packet loss probability is given by $\widehat{P}_L = \sum_{i=0}^{\infty} \Pr\{n = i\} P_L(i)$. This is what we call *spacial multiplexing gain*.
3) While existing CAC schemes have focused on the tradeoff between reducing connection dropping probability and maximizing the bandwidth utilization, transferring a CAC problem in a packet-based network to a CAC problem in a circuit-based network will simply loose the advantage of packet-based architecture, i.e. per-packet bandwidth allocation. Due to the per-packet bandwidth allocation feature of packet-switched networks, there is no limitation on the number of connections that can be accepted. If the arrival traffic rate is more than the system capacity, e.g. too many connections are accepted in the network, then excessive traffic is dropped. In other words, system is able to distribute the service degradation, i.e. increased packet loss rate, among all the participating connections instead of dropping an individual connection. Although handoff arrival process can not be directly controlled by the network, the new connection acceptance rate can be controlled by the

network. Therefore, the packet loss rate can be controlled by the network controller through admission control for new connection requests.

Briefly speaking, the proposed approach in this paper takes into consideration a combination of both connection-level and packet-level QoS parameters in making the admission decision. Although connection-level and packet-level QoS in cellular systems have been considered by other researchers (see for example [14], [17]–[20] and references there in), they have failed to explicitly address the mobility issue. The existing literature has focused on a particular wireless technology, e.g. CDMA, to analyze the physical layer impacts on resource management. The aim of this paper is to study the mobility impact on resource management taking into consideration an abstract model of wireless channel.

The main idea is to model the bandwidth requirement in each cell of the network based on two factors: 1) mobility patterns of users, and 2) packet generation characteristics of individual connections. The main contributions of the paper are as follows:

1) Combining the user mobility with traffic variations in computing effective bandwidth. This is what we call mobile effective bandwidth.
2) Taking into consideration the packet-based network architecture to completely eliminate undesired connection dropping.

It should be emphasized that the approach in this paper is to develop a simple and yet reasonably accurate admission control scheme rather than trying to apply exact but intractable models that do not necessarily capture all the impact of complex network interactions. We believe that the proposed CAC mechanism, due to its simplicity and effectiveness, can be used in future wireless networks for efficient resource allocation and quality of service management.

III. SYSTEM MODEL

A packet-switched cellular network is considered in this paper, in which mobile users move along an arbitrary topology of M cells according to the routing probabilities r_{ij} (from cell i to cell j). In the system under consideration, no handoff connection is blocked instead excessive packets are dropped. Furthermore, it is assumed that:

- The system is homogeneous, i.e. there is one type of traffic.
- Due to the physical nature of wireless channel, channel transmission rate varies over time depending on interference level, shadowing, channel fading, etc. This dynamic behavior of

wireless channel may actually cause packets to be dropped by the network. We assume that there are appropriate underlying coding and retransmission mechanisms, e.g. combination of FEC and ARQ, to cope with packet loss due to channel effects. Therefore, each cell consists of a number of reliable channels whose transmission rate varies over time.

- Buffer overflow is approximated by cell overflow, i.e. receiving more packets that the transmission capacity. This is often a substantial overestimate of the actual buffer overflow probability since it ignores the smoothing effect of the buffer, i.e. the buffer allows the arrival rate to exceed the service rate for short periods. The significance of such inaccuracies must be tempered by the fact that even an exact model does not provide a correct measure of the loss probability seen by connections, as it can not fully capture the impact of interactions within the network. This is a common technique in approximating packet loss probability (see for example [5], [14]).
- Because of the statistical and spacial multiplexing of the shared radio resources, the QoS (packet loss probability) is based on aggregate statistical measure matching the overall system performance rather than per-connection. Allocating a fair share of resources to each connection is out of the scope of admission control mechanism and must be addressed by resource allocation and scheduling algorithms. In practice, providing per-connection QoS in wireless networks is difficult and will sacrifice the radio resource utilization.

At the level that we study the system, the only dependency between a cell and the rest of the network is through handoffs. The effect of inter-cell interferences is already embedded in variable capacity modeling. Therefore, we start analyzing the system by considering a single cell and then extend it to multiple cells when computing the handoff arrival rate. The handoff rate can be computed from other system parameters, e.g. mobility pattern. Later on, we will present a commonly used iterative method for computing the handoff arrival rate into a cell taking into consideration other cells of the network. Moreover, we assume that

1) New connection arrivals into cell i are Poisson distributed with rate $\lambda_n(i)$,
2) Handoff connection arrivals into cell i form a Poisson process with rate $\lambda_h(i)$ which shall be specified later,
3) Cell residency times in cell i are independent and exponentially distributed with mean $1/\eta(i)$,

```
if ($x$ is a handoff connection) then
    grant admission;
else /* $x$ is a new connection */
    if (cell occupancy $\leq N$) then
        grant admission;
    else
        reject;
    end if
end if
```

Fig. 1. Connection admission control algorithm.

4) Connection durations are independent and exponentially distributed with mean $1/\mu$.

The exponential connection durations and cell residency times are widely used in literature [6]–[8], [10]. In the real word, the cell residence time distribution may not be exponential but exponential distributions provide the mean value analysis, which indicates the performance trend of the system.

Each cell executes an admission control algorithm which follows the pseudo-code given in Fig. 1. In this algorithm x represents a connection request and N is a threshold value and shall be specified later in section V. In order to maximize the bandwidth utilization, we are interested to find the maximum possible value of N. As it can be seen, this algorithm has zero connection dropping probability, i.e. no connection is dropped. This is in contrast to existing admission control schemes [6]–[8], [10], [15], [16] that block handoffs if there is no free circuit in the cell. In the considered system an unlimited number of connections can be accommodated at the expense of possibly increasing the number of packets dropped. We argue that by properly choosing N, it is possible to control packet dropping rate. The goal is to find the maximum N with respect to a target packet loss probability P_{QoS}.

IV. MOBILE EFFECTIVE BANDWIDTH

A. *Connection-Level Analysis*

To simplify the notation, we focus on a single cell, say cell i, in isolation and drop the cell index in our discussion. Let random variable n denote the cell occupancy, i.e. number of active connections in the cell. In our scheme, handoff requests are always accepted regardless of the congestion situation. To control the packet dropping rate, new connection requests are subject to admission control. Fig. 2 shows a Markov chain representing the evolution of the cell occupancy

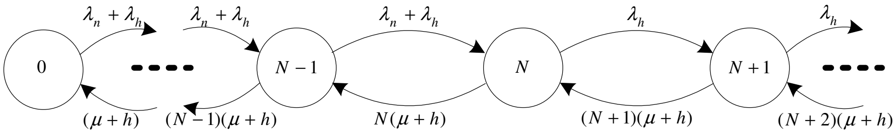

Fig. 2. Cell occupancy state transition diagram.

n. Both handoff and new connection requests are accepted while $n \leq N$. But whenever $n > N$, only handoff requests are allowed in the cell.

Define P_i to be the steady-state probability distribution of n conditioned on threshold parameter N; that is

$$P_i = \Pr\{n = i \,|N\} \,. \tag{1}$$

Using the balance equations, P_i is expressed as

$$P_i = \begin{cases} \dfrac{\rho^i}{i!} P_0 & \text{if } i \leq N \\ \left(\dfrac{\rho}{\rho_h}\right)^N \dfrac{\rho_h^i}{i!} P_0 & \text{if } i > N \end{cases} \tag{2}$$

where $\rho = \frac{\lambda_n+\lambda_h}{\mu+\eta}$ and $\rho_h = \frac{\lambda_h}{\mu+\eta}$. The normalizing condition is that $\sum_{i=0}^{\infty} P_i = 1$, which can be used to find P_0. Using (2), the blocking probability for new connection requests, P_B, is given by

$$P_B = P_N \,. \tag{3}$$

There are two parameters that will be used later in our analysis, namely the expected cell occupancy at steady-state denoted by $\mathrm{E}[n|N]$ and the variance of the cell occupancy denoted by $\mathrm{Var}[n|N]$, both conditioned on the admission threshold N. To simplify the equations, we define A and A_h as follows:

$$A = \sum_{i=0}^{N} \frac{\rho^i}{i!}, \tag{4}$$

$$A_h = e^{\rho_h} - \sum_{i=0}^{N} \frac{\rho_h^i}{i!} \,. \tag{5}$$

Using the state probabilities given by (2), the first two moments of n, i.e. $\mathrm{E}[n|N]$ and $\mathrm{E}[n^2|N]$, can be expressed as

$$\mathrm{E}[n|N] = \left[\rho A' + \rho_h A_h' \left(\frac{\rho}{\rho_h}\right)^N\right] P_0, \tag{6}$$

and,

$$\mathrm{E}[n^2|N] = \left[\rho(A' + \rho A'') + \rho_h(A'_h + \rho_h A''_h)\left(\frac{\rho}{\rho_h}\right)^N\right] P_0, \tag{7}$$

where z' and z'' denote the first and second derivatives of z. Also, P_0 can be expressed as

$$P_0 = \left[A + A_h\left(\frac{\rho}{\rho_h}\right)^N\right]^{-1}. \tag{8}$$

As it can be seen from (6) and (7), the use of A and A_h eliminates the infinite sums involved in computing $\mathrm{E}[n|N]$ and $\mathrm{E}[n^2|N]$, which will greatly simplify numerical analysis. It is left for the reader to show that all derivatives in these equations can be also expressed in terms of A and A_h. Finally, using the computed moments and noting that $\mathrm{Var}[n|N] = \mathrm{E}[n^2|N] - \mathrm{E}^2[n|N]$, $\mathrm{Var}[n|N]$ can be readily computed.

To complete the connection-level analysis, the handoff arrival rate λ_h must be determined. At this point, indices are used to distinguish among different cells in the network. Let us define $P_H(j)$ as the probability that a connection currently being served in cell j requires another handoff before completion. Also, let t_c and t_j denote the connection duration and residency time cell j. Noting that t_c and t_j are exponentially distributed with parameters μ and η_j, $P_H(j)$ can be expressed as

$$\begin{aligned} P_H(j) &= \Pr(t_c > t_j) \\ &= \frac{\eta_j}{\mu + \eta_j}. \end{aligned} \tag{9}$$

Then, the rate of handoff out of cell j is given by

$$\Big[[1 - P_B(j)]\lambda_n(j) + \lambda_h(j)\Big] P_H(j). \tag{10}$$

Finally, the handoff arrival rate into cell i is the summation of handoff arrival rates from all cells $1 \leq j \neq i \leq M$ in the network. Therefore, $\lambda_h(i)$ is given by

$$\lambda_h(i) = \sum_{j \neq i} \Big[[1 - P_B(j)]\lambda_n(j) + \lambda_h(j)\Big] P_H(j) r_{ji}, \tag{11}$$

or, equivalently, in matrix form as follows

$$\mathbf{\Lambda}_h = [(\mathbf{I} - \mathbf{B})\mathbf{\Lambda}_n + \mathbf{\Lambda}_h]\,\mathbf{\Phi}, \tag{12}$$

where, $\mathbf{\Lambda}_n = [\lambda_n(1), \ldots, \lambda_n(M)]$, $\mathbf{\Lambda}_h = [\lambda_h(1), \ldots, \lambda_h(M)]$, $\mathbf{B} = \mathrm{diag}[P_B(1), \ldots, P_B(M)]$, $\mathbf{I}$ is an $M \times M$ identity matrix, and $\mathbf{\Phi}[\phi_{ij}]$ is the handoff rate matrix with $\phi_{ij} = P_H(i)r_{ij}$. A fixed-point iteration [21] can be used to obtain the steady-state handoff arrival rate vector $\mathbf{\Lambda}_h$. Iteration starts with an initial value for $\mathbf{\Lambda}_h$, say $[0, \ldots, 0]$, to obtain a new value for $\mathbf{\Lambda}_h$. Then this new value is substituted in (12) to obtain another value. This process continues until $\mathbf{\Lambda}_h$ converges with respect to the desired precision.

B. Packet-Level Analysis

Suppose that connection $i (i \geq 1)$ sends packets, modeled as fluid, at a random rate r_i, which has mean m and variance σ^2. It is assumed that m and σ^2 are known to the admission controller a priori. Given that m and σ^2 are first order statistics, they can be estimated from measured traffic data. Since measuring statistics beyond the second moment is usually impractical [22], this traffic characterization is ideal from a measurement point of view. This is a minimal set of requirements since it does not enforce anything specific on the actual packet generating process of the individual connections. It means that individual packet generating processes can have arbitrary correlation structure and this includes self-similar processes as well [23].

Let us define $R(N)$ to be the random variable representing the total packet arrival rate into the test cell where the admission threshold is set to N. That is

$$R(N) = \sum_{i=1}^{n} r_i, \tag{13}$$

where n itself is a random variable representing the number of connections in the cell.

Let Φ_r denote the moment generating function of r_i, i.e. $\Phi_r(\theta) = \mathrm{E}[e^{\theta r_i}]$. Also, let Φ_R denote the moment generating function of $R(N)$, i.e. $\Phi_R(\theta) = \mathrm{E}[e^{\theta R(N)}]$. It is straightforward to show that

$$\Phi_R(\theta) = \mathrm{E}[\{\Phi_r(\theta)\}^n] \,. \tag{14}$$

Using (14), it is obtained that

$$\mathrm{E}[R(N)] = m\mathrm{E}[n|N], \tag{15}$$

$$\mathrm{Var}[R(N)] = \sigma^2\mathrm{E}[n|N] + m^2\mathrm{Var}[n|N] \,. \tag{16}$$

As expected, the variance of the total packet arrival rate is a function of both variance of individual connections packet generating process and the variance of the number of connections in a cell. This indicates that static treatment of the cell occupancy, e.g. assuming that there are $E[n]$ connections in a cell, is not accurate and must be avoided.

Assuming the target loss probability is sufficiently small, we approximate the packet loss probability by the cell overflow probability, i.e. $\Pr\{R(N) > c\}$. Therefore, to guarantee a target packet loss probability of P_{QoS} we must have

$$\begin{aligned} P_L &= \Pr\{R(N) > c\} \\ &\leq P_{\text{QoS}} \, . \end{aligned} \tag{17}$$

When there are many sources each small compared to the total, as is expected to be the case in future wireless networks, using central limit theorem the aggregated packet arrival process $R(N)$ from n connections can be approximated by a Gaussian process. It is expected that future wireless technologies such as 3G and 4G increase the available cell capacities to several Mbps. In such networks, the number of active connections (and consequently, the number of packets being transmitted) is so high that the central limit theorem can be successfully applied to model the packet arrival process in each cell. In fact, it has been observed that the aggregation of even a fairly small number of traffic streams is usually sufficient for the Gaussian characterization of the input process [22].

An important characteristic of wireless networks is that system capacity varies over time due to physical layer interactions. Let assume that in steady-state, the system capacity c has the probability distribution $F_c(x) = \Pr\{c \leq x\}$, or, equivalently, the probability density function $f_c(x) = \frac{d}{dx} F_c(x)$. Let $\tilde{P}_L$ denote the average packet loss probability with respect to variable system capacity c which varies over the interval $[c_1, c_2]$. That is

$$\begin{aligned} \tilde{P}_L &= \int_{c_1}^{c_2} \Pr\{R(N) > c\} \, dF_c(x) \\ &= \frac{1}{2} \int_{c_1}^{c_2} \operatorname{erfc}\left(\frac{x - \mathrm{E}[R(N)]}{\sqrt{2 \mathrm{Var}[R(N)]}} \right) f_c(x) \, dx \end{aligned} \tag{18}$$

where $\operatorname{erfc}(x)$ is the complementary error function defined as

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} \, dt \, . \tag{19}$$

Then (17) should be modified as $\tilde{P}_L \leq P_{\text{QoS}}$.

In the case that a strict packet loss probability must be enforced regardless of the capacity fluctuations, (17) may be changed as follow

$$\begin{aligned} P_L^* &= \sup_c \Big[\Pr\{R(N) > c\} \Big] \\ &\leq P_{\text{QoS}} \, . \end{aligned} \tag{20}$$

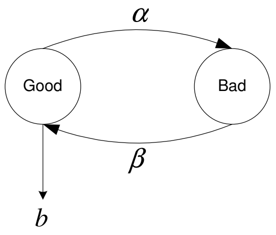

Fig. 3. Wireless channel model.

C. Cell Capacity

A simple model describing cell capacity is based on individual wireless channel behavior. Let assume that the bandwidth in each cell is channelized and channels (codes in CDMA, frequencies in FDMA or time slots in TDMA) behave independently of each other. Let assume that there are J channels in a cell where each channel is represented by a two-state model, as depicted in Fig. 3.

When in Good state, the channel's transmission rate is b but in Bad state it goes down to zero. The average time spent in Good and Bad states is $1/\alpha$ and $1/\beta$, respectively. Let j denote the number of channels in Good state. Then, j has a binomial distribution with parameter $(1+\omega)$, where $\omega = \beta/\alpha$. Therefore,

$$\Pr\{j = x\} = \frac{1}{(1+\omega)^J}\binom{J}{x}\omega^x \,. \tag{21}$$

This is a discreet model for cell capacity which can be easily extended to more complicated models [24] using the same idea. Rewriting (18) for discrete capacity distribution gives us

$$\begin{aligned}\tilde{P}_L &= \sum_{x=0}^{J} \Pr\{R(N) > xb\} \Pr\{j = x\} \\ &= \frac{1}{2(1+\omega)^J} \sum_{x=0}^{J} \operatorname{erfc}\left(\frac{xb - \mathrm{E}[R(N)]}{\sqrt{2\mathrm{Var}[R(N)]}}\right)\binom{J}{x}\omega^x \,.\end{aligned} \tag{22}$$

D. Effective Bandwidth

In order to have a fair comparison between fixed and mobile effective bandwidths, in this section, we ignore the handoff priority. In other words, new connections as well as handoff requests are subject to admission control. If there are N connections already in a cell, then the connection request is blocked no matter the connection request is for a handoff or a new connection. To restrict the comparison only to mobility impact, we assume that cell capacity is fixed and equal to c.

Using the results obtained in section IV-B, the packet loss probability can be obtained using the tail of the Gaussian distribution. Assuming that $\alpha = \Phi^{-1}(P_{\mathrm{QoS}})$, where $\Phi(.)$ is the integral over the tail of a Gaussian distribution which can be expressed in terms of the error function [25]. Therefore,

$$\begin{aligned} c &= \mathrm{E}[R(N)] + \alpha\sqrt{\mathrm{Var}[R(N)]} \\ &= m\mathrm{E}[n|N] + \alpha\sqrt{\sigma^2\mathrm{E}[n|N] + m^2\mathrm{Var}[n|N]}\,. \end{aligned} \tag{23}$$

We define $e^* = \dfrac{c}{N}$ to be the mobile effective bandwidth given that the admission threshold is set to N. Then using (23), it is obtained that

$$e^* = \frac{1}{N}\left[m\mathrm{E}[n|N] + \alpha\sqrt{\sigma^2\mathrm{E}[n|N] + m^2\mathrm{Var}[n|N]}\right]\,. \tag{24}$$

It is interesting to see the behavior of mobile effective bandwidth in the following special cases:

1) **Static treatment of cell occupancy:** In this approach, the average number of connections in a cell is used to compute the effective bandwidth. To compute the required capacity in this case, it is enough to set $\mathrm{Var}[n|N] = 0$ in (23). Then,

$$c = m\mathrm{E}[n|N] + \alpha\sigma\sqrt{\mathrm{E}[n|N]}, \tag{25}$$

which is clearly smaller than the capacity required using (23), hence, achieved packet loss will be higher using this approach.

2) **Fixed treatment of cell occupancy:** This is the traditional approach for computing effective bandwidth in wireline networks. In this approach, the number of connections is assumed to be fixed at N. Therefore, $\mathrm{E}[n|N] = N$ and $\mathrm{Var}[n|N] = 0$. Substituting these values in (23), it is found that

$$c = mN + \alpha\sigma\sqrt{N}, \tag{26}$$

which is the famous expression for classical effective bandwidth. This is clearly larger than the capacity required using (23), hence, achieved bandwidth utilization is lower using this approach.

```
$N_l \leftarrow 0$ , $N_u \leftarrow c/m$
$c_l \leftarrow c(N_l)$ , $c_u \leftarrow c(N_u)$
if $c_l \geq c$ then
    $N \leftarrow N_l$
else if $c_u \leq c$ then
    $N \leftarrow N_u$
else
    $\delta \leftarrow (N_u - N_l)$
    for $j = 0$ to $\log(\delta)$ do
        $\delta \leftarrow \delta/2$
        $N \leftarrow N_l + \delta$
        $x \leftarrow c(N)$
        if $x = c$ then
            break
        else if $x < c$ then
            $N_l \leftarrow N$
        end if
    end for
end if
```

Fig. 4. Pseudo-code for finding optimal N.

V. CONNECTION ADMISSION CONTROL

We have to determine N to guarantee the target packet loss probability P_{QoS}. To maximize the bandwidth utilization, we should find the maximum value for N. Fig. 4 shows the pseudo-code for finding N. In this algorithm, $c(N)$ is the capacity required to achieve packet loss P_{QoS} with new connection acceptance threshold N and is given by (23).

This algorithm basically performs a binary search to find the optimal value of N. The optimal value for N under mobile effective bandwidth allocation scheme will be somewhere between the values obtained for peak allocation and average allocation schemes. Since the peak rate for individual connections is unknown, the algorithm simply search the entire interval $[0, c/m]$.

VI. CONCLUSION

We introduced the notion of mobile effective bandwidth to extend the classical effective bandwidth concept introduced for wireline networks to cellular packet networks. Mobility effect is explicitly embedded in the computation of the effective bandwidth. We showed that the classical approach is not accurate for mobile networks and leads to poor network utilization. The proposed technique can be used in cellular packet networks for effective admission control.

A challenging problem is to extend the proposed technique to multiple classes of service. One approach is to have different admission thresholds for different classes of service. Then to solve the problem one must resort to multidimensional Markov chains which increase the computational complexity and may not be suitable for real-time admission control.

REFERENCES

[1] B. Sarikaya, “Packet mode in wireless networks: Overview of transition to third generation,” *IEEE Commun. Mag.*, vol. 38, no. 9, pp. 164–172, Sept. 2000.

[2] T. Zahariadis and D. Kazakos, “(R)evolution toward 4G mobile communication systems,” *IEEE Wireless Commun. Mag.*, vol. 10, no. 4, pp. 6–7, Aug. 2003.

[3] C.-S. Chang and J. A. Thomas, “Effective bandwidth in high-speed digital networks,” *IEEE J. Select. Areas Commun.*, vol. 13, no. 6, pp. 1091–1100, 1995.

[4] F. P. Kelly, “Notes on effective bandwidths,” in *Stochastic Networks: Theory and Applications*, F. P. Kelly, S. Zachary, and I. Ziedins, Eds. Oxford, UK: Oxford University Press, 1996, pp. 141–168.

[5] R. Guerin, H. Ahmadi, and M. Naghshineh, “Equivalent capacity and its application to bandwidth allocation in high-speed networks,” *IEEE J. Select. Areas Commun.*, vol. 9, no. 7, pp. 968–981, 1991.

[6] D. Hong and S. S. Rappaport, “Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures,” *IEEE Trans. Veh. Technol.*, vol. 35, no. 3, pp. 77–92, Aug. 1986, see also: CEAS Tech. Rep. No. 773, College of Engineering and Applied Sciences, State University of New York, June 1999.

[7] R. Ramjee, D. Towsley, and R. Nagarajan, “On optimal call admission control in cellular networks,” *ACM/Kluwer Wireless Networks*, vol. 3, no. 1, pp. 29–41, Mar. 1997.

[8] M. Naghshineh and M. Schwartz, “Distributed call admission control in mobile/wireless networks,” *IEEE J. Select. Areas Commun.*, vol. 14, no. 4, pp. 711–717, May 1996.

[9] D. Levine, I. Akyildiz, and M. Naghshineh, “A resource estimation and call admission algorithm for wireless multimedia networks using the shadow cluster concept,” *IEEE/ACM Trans. Networking*, vol. 5, no. 1, pp. 1–12, Feb. 1997.

[10] S. Wu, K. Y. M. Wong, and B. Li, “A dynamic call admission policy with precision QoS guarantee using stochastic control for mobile wireless networks,” *IEEE/ACM Trans. Networking*, vol. 10, no. 2, pp. 257–271, Apr. 2002.

[11] D. Anick, D. Mitra, and M. M. Sondhi, “Stochastic theory of a data-handling system with multiple sources,” *Bell System Tech. J.*, vol. 61, no. 8, pp. 1871–1894, Oct. 1982.

[12] A. I. Elwalid and D. Mitra, “Effective bandwidth of general markovian traffic sources and admission control of high speed networks,” *IEEE/ACM Trans. Networking*, vol. 1, no. 3, pp. 329–343, June 1993.

[13] A. Shwartz and A. Weiss, *Large deviations for performance analysis*. London, UK: Chapman & Hall, 1995.

[14] J. S. Evans and D. Everitt, “Effective bandwidth based admission control for multi-service CDMA cellular networks,” *IEEE Trans. Veh. Technol.*, vol. 48, no. 1, pp. 36–46, Jan. 1999.

[15] J. M. Peha and A. Sutivong, “Admission control algorithms for cellular systems,” *ACM/Kluwer Wireless Networks*, vol. 7, no. 2, pp. 117–125, Mar. 2001.

[16] Y. Fang and Y. Zhang, “Call admission control schemes and performance analysis in wireless mobile networks,” *IEEE Trans. Veh. Technol.*, vol. 51, no. 2, pp. 371–382, Mar. 2002.

[17] J. S. Evans and D. Everitt, “On the teletraffic capacity of cdma cellular networks,” *IEEE Trans. Veh. Technol.*, vol. 48, no. 1, pp. 153–165, Jan. 1999.

[18] M. M. Krunz and J. G. Kim, “Fluid analysis of delay and packet discard performance for QoS support in wireless networks,” *IEEE J. Select. Areas Commun.*, vol. 19, no. 2, pp. 384–395, Feb. 2001.

[19] J. Cai, X. Shen, and J. W. Mark, “Downlink resource management for packet transmission in OFDM wireless communication systems,” in *Proc. IEEE GLOBECOM’03*, vol. 6, San Francisco, USA, Dec. 2003, pp. 2999–3003.

[20] L. Xu, X. Shen, and J. W. Mark, “Dynamic fair scheduling with QoS constraints in multimedia W-CDMA cellular networks,” *IEEE Trans. Wireless Commun.*, vol. 3, no. 1, pp. 60–73, Jan. 2004.

[21] F. P. Kelly, “Fixed point models of loss networks,” *Australian Mathematical Society*, vol. 31, pp. 204–218, 1989.

[22] D. Y. Eun and N. B. Shroff, “A measurement-analytic approach for QoS estimation in a network based on the dominant time scale,” *IEEE/ACM Trans. Networking*, vol. 11, no. 2, pp. 222–235, Apr. 2003.

[23] T. Karagiannis, M. Molle, and M. Faloutsos, “Long-range dependence: Ten years of Internet traffic modeling,” *IEEE Internet Comput.*, vol. 8, no. 5, pp. 57–64, Sept. 2004.

[24] M. Hassan, M. M. Krunz, and I. Matta, “Markov-based channel characterization for tractable performance analysis in wireless packet networks,” *IEEE Trans. Wireless Commun.*, vol. 3, no. 3, pp. 821–831, May 2004.

[25] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge, UK: Cambridge University Press, 1992.