Joint Call and Packet QoS in Cellular Packet Networks

Majid Ghaderi [a,*] Raouf Boutaba [a] Gary W. Kenward [b]

[a] *University of Waterloo, ON N2L 3G1, Canada*

[b] *Nortel Networks, Ottawa, ON K1Y 4H7, Canada*

Abstract

This paper proposes a novel call admission control scheme capable of providing a combination of call and packet level quality of service requirements in cellular packet networks. Specifically, we propose a distributed call admission control scheme called PFG, which maximizes the wireless channel utilization subject to a predetermined bound on the call dropping and packet loss probabilities for variable-bit-rate traffic in a packet-switched wireless cellular network. We show that in wireless packet networks, the undesired event of dropping an ongoing call can be completely eliminated without sacrificing the bandwidth utilization. Extensive simulation results confirm that our scheme satisfies the hard constraint on call dropping and packet loss probabilities while maintaining a high bandwidth utilization.

Key words: Call admission control, quality of service, call dropping, packet loss.

1. Introduction

The problem of call admission control (CAC) in a packet network can be addressed by a concept known as the *effective bandwidth* [1] of a call. When employing this concept, the problem of admission control in a packet-switched network is mapped to an equivalent admission control problem in a circuit-switched network. In fact, most of the researchers in wireless networking field have focused only on call-level quality of service (QoS) parameters, e.g. *call blocking and dropping probabilities*, for admission control and resource allocation [2–7] because the primary concern has been efficient resource management in a circuit-switched cellular network.

The idea is that, once there is a call-level admission control in place, using effective bandwidth concept, packet-level quality of service requirements, e.g. *packet loss probability*, can be addressed independently. The main difference between our approach and existing approaches is that we take into consideration a combination of both call-level and packet-level traffic dynamics in making the admission decision. We believe that call-level and packet-level quality of service are correlated, and hence, treating them independently neglects the flexibility of packet-based resource management and its impact on call-level quality of service.

Call admission control has been extensively studied in circuit-switched (voice) wireless cellular networks (see [8–10] and references there in). Hong and Rappaport [2] 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. [3] showed that the guard channel scheme is optimal for minimizing a linear objective function of call blocking and dropping probabilities while the *fractional guard channel* scheme (FG) is optimal for minimizing call blocking probability subject to a hard constraint on call dropping probability. Instead of explicit bandwidth reservation as in the GC, the FG accepts new calls according to a randomization parameter called the *acceptance ratio*.

* Corresponding author; Email: mghaderi@cs.uwaterloo.ca

Preprint submitted to Elsevier

Because of user mobility, it is impossible to describe the state of the system by using only local information, unless we assume that the network is uniform and approximate the overall state of the system by the state of a single cell in isolation. To include the global effect of mobility, collaborative or distributed admission control schemes have been proposed [4, 6, 7, 11]. Information exchange among a cluster of neighboring cells is the approach adopted by all distributed schemes. Interested readers are referred to [8] for a comprehensive survey on call admission control in cellular networks.

None of these papers has considered a wireless packet-switched network. There is no packet-level consideration in these works. Call dropping and blocking probabilities are the only QoS parameters considered. In circuit-switched networks, when a handoff call arrives while there is no idle circuit (wireless channel), the handoff fails and hence the call is dropped. In contrast, in a packet-switched network it is still possible to accept the handoff call 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.

Although call-level and packet-level QoS in cellular systems have been considered by other researchers [12–15], they have failed to explicitly address the mobility of users and its impact on admission control performance. The existing literature has focused on a particular wireless technology, e.g. CDMA systems, to analyze the physical layer impacts on resource management. Recently, the admission control problem with joint call and packet QoS support has been studied in [16], where the impact of user mobility is captured by a simple Markov chain corresponding to the number of active calls in a cell. The aim of this paper is to study the impact of mobility on resource management taking into consideration an abstract model of a wireless network. We consider general mobility and traffic models that can easily capture the behavior described in [16].

We introduce a *packetized fractional guard channel* (PFG) call admission control mechanism for cellular packet networks that achieves a high bandwidth utilization while satisfying a target packet loss probability without dropping any ongoing call. To the best of our knowledge, PFG is the first to address achieve these objectives. The main features of PFG are as follows:

(i) PFG achieves zero percent call dropping.
(ii) PFG is dynamic, therefore, adapts to a wide range of traffic conditions.
(iii) PFG is distributed and takes into consideration the information from direct neighboring cells in making admission decisions,
(iv) The control mechanism is stochastic and periodic to reduce the overhead associated with distributed control schemes.

The rest of the paper is organized as follows. Our system model, assumptions and notations are described in section 2. Section 3 describes the high-level operation of the proposed admission control algorithm while in section 4, detailed analysis of the algorithm is presented. Simulation results are presented in section 5 and section 6 concludes this paper.

2. System Model

The considered system is a packet-switched cellular network, in which the users move along an arbitrary topology of $\mathcal{B}$ cells according to the routing probabilities r_{ij} (from cell i to cell j). Each cell i has a set of adjacent cells denoted by $\mathcal{A}_i$. We assume that there is one type of calls in the system. Although the dynamic behavior of wireless channel may cause packets to be dropped, we assume that there are appropriate underlying coding and retransmission mechanisms to cope with packet loss due to channel effects. Buffer overflow is approximated by cell overflow, i.e. receiving more packets that the cell transmission capacity. This is often an overestimate of the actual buffer overflow probability since it ignores the smoothing effect of the buffer. However, it is a common technique for approximating packet loss probability [12, 17] because even an exact model does not provide a correct measure of the loss probability, as it can not fully capture the interactions within the network. Therefore, cell overflow is considered the only source of packet loss. This paper considers constant cell capacity, however, the approach that we propose next can be extended to include variable capacity cases using a technique similar to the one proposed in [18]. Moreover, we assume that

(i) New call arrivals to a cell are independent and Poisson distributed with rate λ_i.
(ii) Cell residency times are independent and exponentially distributed with mean $1/h$. However, we show that the proposed algorithm is insensitive to this assumption.

(iii) Call durations are independent and exponentially distributed with mean $1/\mu$.

These assumptions are widely used in literature [2–4,6,7,9,10,19] for the mean value analysis of cellular systems.

2.1. *Multiple Handoffs Probability*

To reduce the overhead, distributed CAC schemes typically have a periodic structure in which only at the beginning of control periods information exchange is triggered. Moreover, information exchange is typically restricted to a cluster of neighboring cells. In this paper, we set the control interval in such a way that the probability of having multiple handoffs in one control period becomes negligible. Therefore, we can effectively assume that only those cells directly connected to a cell can influence the number of calls in that cell during a control period. In [20], we showed that using this technique, the signalling overhead will not increase, while the collected information on the network status will be sufficiently accurate for the purpose of a stochastic admission control. The problem of choosing the proper control interval will be further addressed in section 4.4.

2.2. *Maximum Occupancy in a Cell*

Let M_i denote the maximum occupancy, i.e. maximum number of calls, in cell i under the so-called *average bandwidth assignment* scheme. This scheme allocates to each VBR call a share of bandwidth equal to the call's average bandwidth requirement. Let m denote the average bandwidth requirement of a call, then

$$M_i = \frac{c_i}{m}, \tag{1}$$

where c_i denoted the capacity of cell i. Although this scheme achieves a high bandwidth utilization, it leads to a high rate of packet loss [21]. If there are more than M_i calls in cell i, then we say that the cell is in *overloaded state*. In the overloaded state, probability of packet loss is very high. Our scheme, PFG, rejects new call requests when a cell is in overloaded state.

2.3. *Time-Dependent Handoff Probability*

Let $P_h(t)$ denote the probability that a call hands off to another cell by time t and remains active until t, given that it has been active at time 0. Also, let $P_s(t)$ denote the probability that a call remains active in its home cell until time t, given that it has been active at time 0. It is obtained in [20] that $P_h(t) = (1 - e^{-ht})\, e^{-\mu t}$ and $P_s(t) = e^{-(\mu+h)t}$. On average, for any call which arrives at time $t' \in (0, t]$, the average handoff and stay probabilities $\widetilde{P}_h$ and $\widetilde{P}_s$ are expressed as

$$\widetilde{P}_h(t) = \frac{1}{t}\int_0^t P_h(t - t')\, dt', \tag{2}$$

$$\widetilde{P}_s(t) = \frac{1}{t}\int_0^t P_s(t - t')\, dt' . \tag{3}$$

Similar to [4] and [6], we assume that during a control period each call experiences at most one handoff. This assumption is justified by setting the length of the control period T reasonably shorter than the average cell residency time. Discussion on the appropriate control interval is not included here due to paper length restriction. For the optimal control interval T, please refer to [20].

Finally, let $P_{ji}(t)$ denote the time-dependent handoff probability that an active call in cell j at time 0 will be in cell i at time t, where $j \in \mathcal{A}_i$. Since each call experiences at most one handoff during the control period, it is obtained that $P_{ji}(t) = P_h(t)\, r_{ji}$. Similarly, the average handoff probability $\widetilde{P}_{ji}(t)$ for a call which arrives at any time $t' \in (0, t]$ is given by $\widetilde{P}_{ji}(t) = \widetilde{P}_h(t)\, r_{ji}$.

3. Admission Control Algorithm

The proposed distributed algorithm, PFG, consists of two components. The first component is responsible for retrieving the required information from the neighboring cells and computing the acceptance ratio. Using the computed acceptance ratio, the second component enforces the admission control locally in each cell. The following sections describe these two components in detail.

3.1. *Distributed Control Algorithm*

To reduce the signalling overhead all the information exchange and acceptance ratio computations happen only once at the beginning of each control period of length T. Several steps involved in PFG distributed control are described below:

(i) At the beginning of a control period, each cell i sends the number of active calls in the cell at the beginning of the control period denoted by $N_i(0)$ to its adjacents and receives the number

```
if (x is a handoff call) then
   accept call;
else /* x is a new call */
   if (rand(0,1) < a_i) & (N_i(t) ≤ M_i) then
      accept call;
   else
      reject call;
   end if
end if
```

Fig. 2. Local call admission control algorithm in cell i.

of new calls, N_i, which were admitted in the last control period by each adjacent cell.

(ii) Now, cell i uses the received information and those available locally to compute the acceptance ratio a_i using the technique described in section 4.

(iii) Finally, the computed acceptance ratio a_i is used to admit call requests into cell i using the algorithm presented in section 3.2.

3.2. *Local Control Algorithm*

Let s_i denote the state of cell i, where there are s calls active in the cell. Let $a_i(s)$ denote the acceptance ratio where the cell state is s_i. Fig. 1 shows the state transition diagram of the PFG scheme in cell i. In this diagram, ν_i is the handoff arrival rate into cell i, and M_i is the maximum occupancy given by (1).

The pseudo-code for the local admission control in cell i is given by the algorithm in Fig. 2. In this algorithm, x is a call requesting a connection into cell i. The acceptance ratio for the respective control period is a_i. Also, $\texttt{rand}(0,1)$ is the standard uniform random generator function. In the next section, we will present a technique to compute the acceptance ratio a_i in order to complete this algorithm.

4. Computing the Acceptance Ratio

Assuming the target loss probability is sufficiently small, we approximate the packet loss probability by the overflow probability in each cell. This is often an 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 calls, 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 [21]). However, as the overflow probability decays to zero, both measures converge to the same value and the difference becomes negligible. Therefore, the time-dependent packet loss probability at time t in cell i denoted by $L_i(t)$ is given by

$$L_i(t) = \mathbb{P}\{R_i(t) > c_i\}, \tag{4}$$

where $R_i(t)$ denotes the total (new and handoff) packet arrival rate into cell i at time t. The main idea is to describe $R_i(t)$ using a Gaussian distribution. The motivation behind Gaussian traffic characterization is that it is very natural when a large number of traffic sources are multiplexed (motivated by the central limit theorem), as is expected to be the case in future wireless networks.

4.1. *Traffic Characterization*

Suppose that call $n (n \geq 1)$ sends packets, modeled as fluid, at a random rate r_n, which has mean m and variance σ^2. 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 calls. It means that individual packet generating processes can have arbitrary correlation structure and this includes self-similar processes [23] too.

It is assumed that individual packet generating processes are independent and identically distributed (iid) random variables. Let $R_i(t)$ denote the total packet arrival rate in cell i at time t, which is expressed as the summation of packet generating process of individual calls. That is

$$R_i(t) = \sum_{n=1}^{N_i(t)} r_n, \tag{5}$$

where $N_i(t)$ denotes the number of calls at time t. Our objective is to apply the central limit theorem to approximate $R_i(t)$ by a Gaussian distribution. At first we have to specify the parameters of random process $R_i(t)$, namely, mean and variance. Using moment generating functions it was shown in [20] that

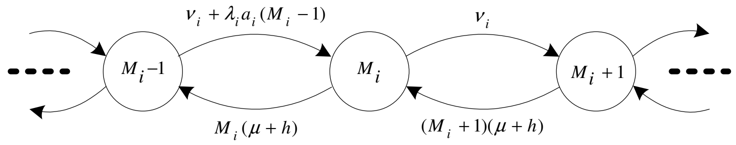


Fig. 1. Packetized fractional guard channel transition diagram.

$$\mathbb{E}\left[R_i(t)\right] = m\mathbb{E}\left[N_i(t)\right], \tag{6}$$

$$\mathrm{Var}[R_i(t)] = \sigma^2\mathbb{E}\left[N_i(t)\right] + m^2\mathrm{Var}[N_i(t)]. \tag{7}$$

As expected, the variance of the total packet arrival rate is a function of both variance of individual call packet generating process and the variance of the number of calls at time t. This indicates that static treatment of the number of calls in a cell, i.e. assuming that there are $\mathbb{E}\left[N_i(t)\right]$ calls in a cell, is not accurate and must be avoided.

4.2. *Mobility Characterization*

The number of calls in cell i at time t is affected by two factors: (1) the number of background (existing) calls which are already in cell i or its adjacent cells, and, (2) the number of new calls which will arrive in cell i and its adjacent cells during the period $(0,t]$ $(0 < t \le T)$. Let $g_i(t)$ and $n_i(t)$ denote the number of background and new calls in cell i at time t, respectively.

A background call in cell i will remain in cell i with probability $P_s(t)$ or will handoff to an adjacent cell j with probability $P_{ij}(t)$. A new call which is admitted in cell i at time $t' \in (0,t]$ will stay in cell i with probability $\widetilde{P}_s(t)$ or will handoff to an adjacent cell j with probability $\widetilde{P}_{ij}(t)$. Therefore, the number of background calls which remain in cell i and the number of handoff calls which come into cell i during the interval $(0,t]$ are binomially distributed. Using this property, the time-dependent variance of stay and handoff processes denoted by $V_s(t)$ and $V_{ji}(t)$ and their average counterparts denoted by $\widetilde{V}_s(t)$ and $\widetilde{V}_{ji}(t)$ can be computed (please refer to [20] for details).

The number of calls in cell i is the summation of the number of background calls, $g_i(t)$, and new calls, $n_i(t)$. Therefore, the mean number of active calls in cell i at time t is given by

$$\mathbb{E}\left[N_i(t)\right] = \mathbb{E}\left[g_i(t)\right] + \mathbb{E}\left[n_i(t)\right], \tag{8}$$

where,

$$\mathbb{E}\left[g_i(t)\right] = N_i(0)p_s(t) + \sum_{j\in\mathcal{A}_i} N_j(0)p_{ji}(t), \tag{9}$$

$$\mathbb{E}\left[n_i(t)\right] = \big(a_i\lambda_n(i)t\big)\widetilde{p}_s(t) + \sum_{j\in\mathcal{A}_i}\big(a_j\lambda_n(j)t\big)\widetilde{p}_{ji}(t). \tag{10}$$

Similarly the variance is given by

$$\mathrm{Var}[N_i(t)] = \mathrm{Var}[g_i(t)] + \mathrm{Var}[n_i(t)], \tag{11}$$

where,

$$\mathrm{Var}[g_i(t)] = N_i(0)V_s(t) + \sum_{j\in\mathcal{A}_i} N_j(0)V_{ji}(t), \tag{12}$$

$$\mathrm{Var}[n_i(t)] = \big(a_i\lambda_n(i)t\big)\widetilde{V}_s(t) + \sum_{j\in\mathcal{A}_i}\big(a_j\lambda_n(j)t\big)\widetilde{V}_{ji}(t) \tag{13}$$

Note that given the arrival rate λ_i and the acceptance ratio a_i, the *actual new call arrival rate* into cell i is given by $\bar{\lambda}_i = \lambda_i a_i$. Therefore, the expected number of call arrivals during the interval $(0,t]$ is given by $a_i\lambda_i t$. Instead of using the actual value of $\bar{\lambda}_j$ which is not known at the beginning of the new control interval (time 0), each cell i estimates the actual new call arrival rates of its adjacents for the new control period using an exponentially weighted moving average technique, i.e. $\bar{\lambda}_j \leftarrow (1-\epsilon)\frac{N_j}{T} + \epsilon\bar{\lambda}_j$. In our simulations we found that $\epsilon = 0.3$ leads to a good estimation of the actual new call arrival rate.

4.3. *Packet Loss Probability*

As mentioned earlier, the packet arrival distribution in each cell can be approximated by a Gaussian distribution:

$$R_i(t) \sim \mathbb{G}\Big(\mathbb{E}\left[R_i(t)\right],\ \mathrm{Var}[R_i(t)]\Big), \tag{14}$$

where, $\mathbb{E}\left[R_i(t)\right]$ and $\mathrm{Var}[R_i(t)]$ are given by (6) and (7), respectively. Using (4) and (14) it is obtained that

$$L_i(t) = \frac{1}{2}\,\mathrm{erfc}\left(\frac{c_i - \mathbb{E}\left[R_i(t)\right]}{\sqrt{2\,\mathrm{Var}[R_i(t)]}}\right), \tag{15}$$

where, $\mathrm{erfc}(c) = \frac{2}{\sqrt{\pi}}\int_c^\infty e^{-t^2}\,dt$. Using (15), the average packet loss probability over a control period of length T is given by

```
if L̃_i(0) ≥ P_L then
  a_i ← 0
else if L̃_i(1) ≤ P_L then
  a_i ← 1
else
  Δ ← 1
  x ← 0
  n ← log (Δ/ξ)
  for i = 0 to n do
    Δ ← Δ/2
    a_i ← x + Δ
    if L̃_i(a_i) < P_L then
      x ← a_i
    end if
    if L̃_i(a_i) = P_L then
      break
    end if
  end for
end if
```

Fig. 3. Algorithm for computing a_i.

$$\widetilde{L}_i = \frac{1}{T} \int_0^T L_i(t)\, dt \,. \tag{16}$$

Then, the acceptance ratio, a_i, can be found by numerically solving equation

$$\widetilde{L}_i = P_L, \tag{17}$$

where P_L is the target packet loss probability. The boundary condition is that $a_i \in [0, 1]$, hence if $\widetilde{L}_i$ is less than P_L even for $a_i = 1$ then a_i is set to 1. Similarly, if $\widetilde{L}_i$ is greater than P_L even for $a_i = 0$, then a_i is set to 0.

Fig. 3 shows a binary search algorithm for computing a_i. In this algorithm, $\widetilde{L}_i(a_i)$ denotes the packet loss probability in cell i given by (16) assuming that the acceptance ratio is a_i. Let ξ denote the required numerical precision. Then, the computational complexity of this algorithm is $O(\log 1/\xi)$, given that all mathematical operations (including exponentiation and integration) can be performed in $O(1)$.

4.4. *Control Interval*

The idea behind at-most-one handoff assumption is that by setting control interval appropriately, the undesired multiple handoffs during a control period can be avoided. As discussed in section 2.1, this minimizes the signalling overhead and operational complexity of PFG. Consider a symmetric network where each cell has exactly $\mathcal{A}$ neighbors, and the probability of handoff to every neighbor is the same. Let $q(n)$ denote the probability that an active call experiences n handoffs during time interval T. Also, let $q_{ij}(n)$ denote the probability that a call originally in cell i moves to cell j over a path consisting of n handoffs during time interval T. Define δ as the multiple handoffs probability from cell i to cell j. We then can write

$$\delta = \sum_{n=2}^{\infty} q_{ij}(n)\,. \tag{18}$$

Assuming $\delta \ll \frac{1}{\mathcal{A}}(\frac{\eta}{\mu+\eta})$, it was shown in [20] that

$$T \approx \frac{\sqrt{2\mathcal{A}\delta}}{\eta}\,. \tag{19}$$

5. Simulation Results

5.1. *Simulation Parameters*

Simulations were performed on a two-dimensional cellular system consisting of 19 hexagonal cells where opposite sides wrap-around to eliminate the finite size effect. As the basic traffic type, packetized voice calls are generated for simulation purposes. For packetized voice, a packet loss probability of $P_L = 0.01$ is acceptable. The common parameters used in the simulation are as follows. All the cells have the same capacity c. Target packet loss probability is $P_L = 0.01$, control interval is set to $T = 20\,s$ and $r_{ji} = 1/6$. In all of the cases simulated, *normalized load* $\rho = \frac{1}{M_i}(\frac{\lambda}{\mu})$ is used, where M_i is given by (1). For each load, simulations were done by averaging over 8 samples, each for $10^4\,s$ of simulation time. Unless otherwise mentioned, call duration and cell residency times are exponentially distributed with means $\mu^{-1} = 180\,s$ and $h^{-1} = 100\,s$, respectively. We found this set of parameters more or less common and reasonable for a simulation setup (see for example [7]).

A two-state Markov model is used to describe the traffic generation process of voice calls. In this model, α and β are transition rates to OFF and ON states, respectively, from ON and OFF states. While in the ON state, traffic is generated at a constant rate of A packet/sec. For this traffic model, the mean and variance of the traffic generated are given by $E[r] = \frac{\beta}{\alpha+\beta}A$ and $V[r] = \frac{\alpha\beta}{(\alpha+\beta)^2}A^2$. Commonly used parameters for human speech representation are $\alpha^{-1} = 1.2\,s$ and $\beta^{-1} = 1.8\,s$ [21]. Using an 8

Kbps encoded voice source, it is obtained that $A = 100$ packet/sec and hence, $E[r] = 40$ packet/sec and $V[r] = 50$ packet/sec assuming that each packet is 80 bits.

5.2. *Conservative PFG*

As mentioned earlier, PFG does not drop any handoff call, instead some packets may be dropped to accommodate the incoming handoff packets. To have an intuition about the impact of accepting handoffs even during the overloaded state, we have also implemented a slightly different version of PFG in addition to the original PFG represented in Fig. 1. This modified version drops handoffs during the overloaded state. We refer to the original algorithm by PFG-D0 and the modified one by PFG-DP where D0 and DP stand for zero dropping probability and P dropping probability, i.e. if we use PFG-DP instead of PFG-D0 then there will be P percent call dropping. Our purpose is to find the value of P for some simulated scenarios to see how far it is from zero. Notice that, having $N_i(t) > M_i$ ($t \in (0, T]$) indicates that cell i is in the overloaded state at time t.

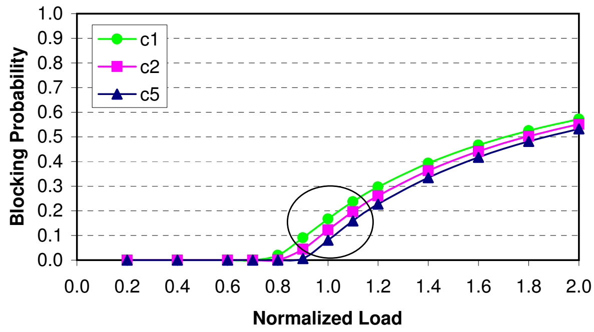
5.3. *Results and Analysis*

As mentioned earlier, PFG is the first to achieve zero call dropping while guaranteeing a hard constraint on packet loss probability. To the best of our knowledge there is no existing scheme which takes into consideration a combination of call-level and packet-level QoS parameters while taking into consideration the *mobility of users*. Therefore, we are not able to compare the performance of PFG with any other scheme. Instead, by doing extensive simulations, we have shown that PFG can achieve its defined goals.

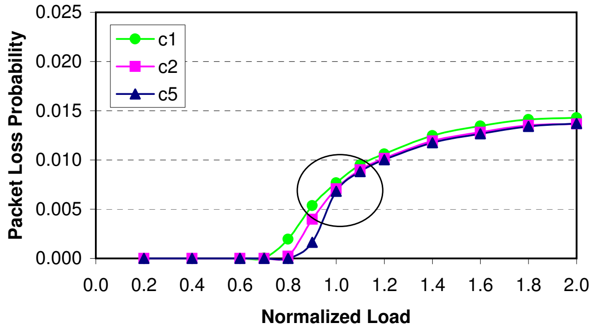
In the rest of this section, we present our simulation results. Several scenarios have been considered to investigate the impact of major factors such as cell capacity, control interval, cell residency and mobility on the performance of PFG. In all the simulated cases, PFG is stable and achieves accurate results.

5.3.1. *Effect of cell capacity*

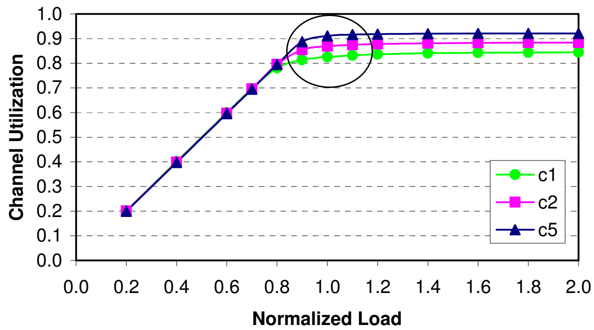
Intuitively, increasing the cell capacity leads to a better Gaussian approximation, and hence, a more accurate admission decision. To investigate the effect of cell capacity, we considered three different



(a) Blocking probability.



(b) Packet loss probability.



(c) Channel utilization.

Fig. 4. PFG-D0 performance results.

Table 1
Cell capacity profiles.

Profile	Capacity
$C1$	1 Mbps
$C2$	2 Mbps
$C5$	5 Mbps

capacity configurations as shown in Table 1. Keep in mind that the system under consideration is a broadband wireless network such as 3G and 4G systems [24, 25]. Normalized loads in range $[0 \dots 2]$ are simulated, where the normalized load is defined as before.

In Figs. 4(a), 4(b) and 4(c) we have circled a re-

Table 2
PFG-DP call dropping probability.

Load	$C1$	$C2$	$C5$
0.2	0.000000	0.000000	0.000000
0.6	0.000000	0.000000	0.000000
1.0	0.000000	0.000000	0.000000
1.4	0.000007	0.000002	0.000001
1.8	0.000012	0.000006	0.000005

gion around load $\rho = 1.0$. This is the most interesting part of the system which is likely to happen in practice. In the following discussion we refer to this region as the *operating region* of the system.

Fig. 4(a) shows the new call blocking probability. It is clear from the figure that as the cell capacity increases the blocking probability decreases which can be explained from the central limit theorem and Gaussian approximation used in section 4.3. As the system capacity increases, the Gaussian modeling leads to more accurate approximation and hence, decreased call blocking probability.

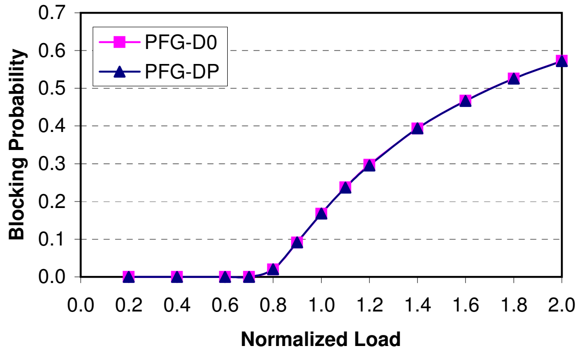
The packet loss probability, $\widetilde{L}_i$, is depicted in Fig. 4(b). Although $\widetilde{L}_i$ goes beyond the target limit for high system loads, it is completely satisfactory for the operating region. Nevertheless, it is quite possible to modify PFG-D0 in order to make it more conservative for high loads. Similar to call blocking, as the capacity increases the PFG-D0 efficiency improves.

Fig. 4(c) depicts the wireless bandwidth utilization under the three different system capacities. As explained before, increased accuracy of the Gaussian approximation for high system capacity leads to a better channel utilization. After all, $C1$ produces rather accurate results and increasing the capacity beyond it produces only marginal improvements.

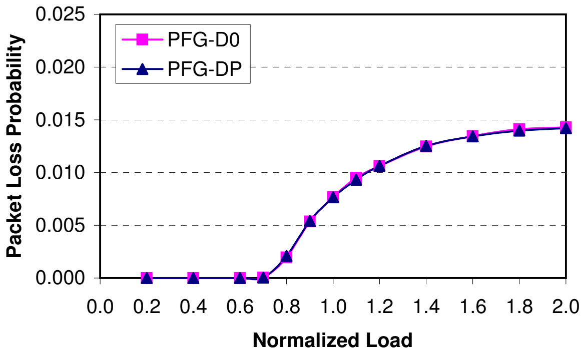
5.3.2. *Effect of accepting handoffs in overloaded state*

To investigate the impact of accepting handoffs during the overloaded state (in which $N_i(t) > c_i$), we ran PFG-DP for the same simulation configuration we ran PFG-D0. Table 2 shows the call dropping probabilities for different loads and capacities. It is observed that the call dropping probability is almost zero in all the simulated configurations. It means that basically there is no difference between two schemes in terms of the call dropping probability.

Fig. 5 shows the call blocking and packet loss probabilities of PFG-D0 versus PFG-DP when the system capacity is set to $C1$ (1 Mbps). Overall, there is no difference between the two schemes. It can be seen from Fig. 5(b) that the packet loss probability is almost the same for both schemes indicating that accepting handoffs during the overloaded state has a negligible effect on the admission control performance. Fig. 5(a) further confirms the same result.



(a) Blocking probability.



(b) Packet loss probability.

Fig. 5. PFG-D0 vs. PFG-DP.

5.3.3. *Effect of mobility*

To increase the capacity of cellular networks, micro/pico cellular architectures will be deployed in the future. The smaller cell size of these architectures leads to a higher handoff rate. Define the *mobility factor* to be $\alpha = \eta/\mu$. Intuitively, α shows the average number of handoff attempts a call makes during its life time. As the mobility factor increases the handoff arrival rate increases as well. To investigate the impact of mobility on PFG, we have simulated three mobility cases for the base capacity $C1$ as shown in Table 3. In this table, $\alpha = 9.00$ represents a highly mobile scenario such as vehicular users in a high way; $\alpha = 1.80$ is a common scenario typically used in similar research papers [7] and shows an urban area mobility, and finally, $\alpha = 0.36$ represents

Table 3
Mobility profiles.

Profile	$1/\mu\,(s)$	$1/\eta\,(s)$	α
Mob: high	180	20	9.00
Mob: mod	180	100	1.80
Mob: low	180	500	0.36

a low mobility case.

Observed from Fig. 6, PFG is almost insensitive to the mobility rate of users. As shown in Figs. 6(a) and 6(c), the call blocking probability and channel utilization are almost match. Furthermore, Fig. 6(b) shows that the effect of mobility on packet loss probability is not very significant. In all three cases, PFG is able to satisfy the target packet loss probability in the operating region of the system. In general, handoff degrades the performance of cellular systems.

5.3.4. *Effect of Control Interval*

Both signalling overhead and accuracy of PFG are affected by the control interval. Although increasing the control interval reduce the signalling overhead, the admission control accuracy will deteriorate. Therefore, there must be a compromise between the incurred overhead and the achieved accuracy. As we showed in subsection 4.4, this compromise depends on the mobility of users.

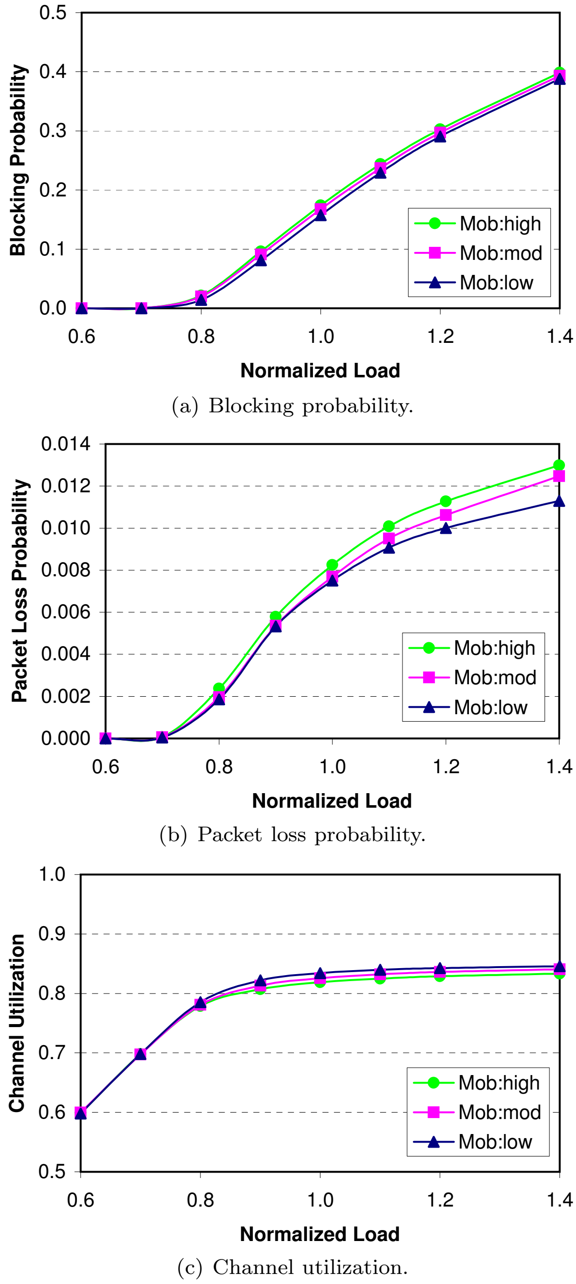
Fig. 7 shows the effect of control interval on the performance of PFG. The simulated scenarios consider the high mobility profile in Table 3, where the mobility factor is set to $\alpha = 9$. It is observed that by reducing the control interval T, the accuracy of PFG in terms of the achieved packet loss probability increases.

An interesting question is that what is the appropriate control interval for high mobility scenario to achieve the same performance as moderate mobility scenario? Fig. 6 shows that there is a small discrepancy between two scenarios when the control interval is the same and equal to $T = 20\,s$.

Using (19), it is obtained that $\frac{T_{\text{Mob: high}}}{T_{\text{Mob: mod}}} = \frac{\alpha_{\text{Mob: mod}}}{\alpha_{\text{Mob: high}}}$. Therefore, $T_{\text{Mob: high}}$ must be set to $\frac{1}{5}T_{\text{Mob: mod}}$ in order to see the same performance results. Fig. 8 shows the simulation results for high mobility and moderate mobility scenarios where $T_{\text{Mob: high}} = 4\,s$ and $T_{\text{Mob: mod}} = 20\,s$.

5.3.5. *Effect of non-exponential cell residence times*

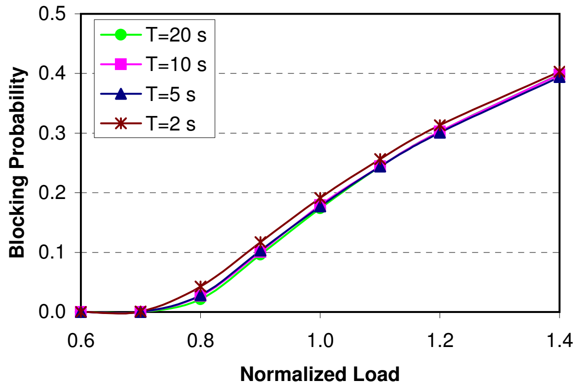
The first part of our analysis, which gives the equations describing the mean and variance of the traffic generation process, is based on the assumption of the exponential cell residency time. As mentioned earlier, exponential distributions provide the mean value analysis, which indicates the performance trend of the system. However, in practice, cell residence times are usually non-exponentially distributed. In this section, we investigate the sensitivity of PFG to exponential cell residency assumption.



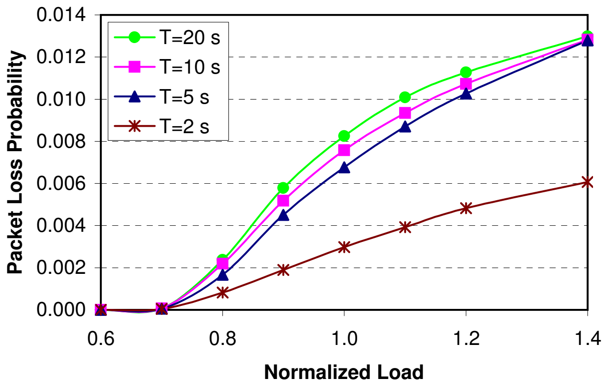
(a) Blocking probability.
(b) Packet loss probability.
(c) Channel utilization.

Fig. 6. Mobility impact on PFG-D0 performance.

Using real measurements, Jedrzycki and Leung

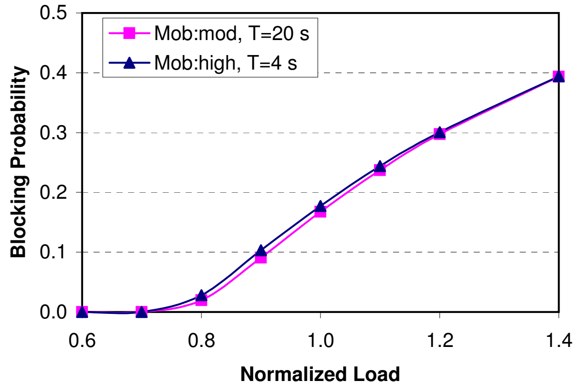


(a) Blocking probability.

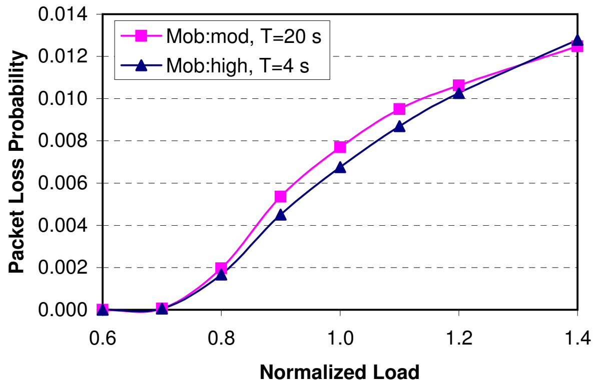


(b) Packet loss probability.

Fig. 7. Effect of control interval.



(a) Blocking probability.



(b) Packet loss probability.

Fig. 8. Robustness to mobility patterns.

[26] showed that a lognormal distribution is a more accurate model for cell residency time. We now compare the results obtained under exponential distribution with those obtained under more realistic lognormal distribution. The mean and variance of both distributions are the same. Fig. 9 shows the call blocking and packet loss probability of exponential cell residency versus lognormal cell residency. It is observed that the exponential cell residency achieves sufficiently accurate control. In other words, the control algorithm is rather insensitive to this assumption due to its periodic control in which the length of the control interval is much smaller than the mean residency time.

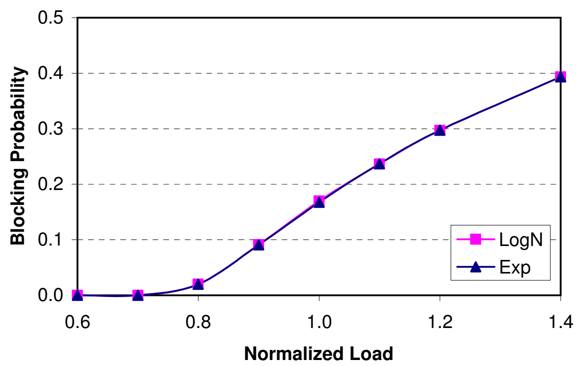
6. Conclusion

In this paper we presented a novel scheme for admission control and hence QoS provisioning for packet-switched cellular systems. In essence, our approach is the natural generalization of the well-known effective bandwidth [17] proposed for wireline networks. Through analysis and simulation, we showed that the proposed scheme, PFG, is able not only to improve utilization of scarce wireless bandwidth thanks to the statistical multiplexing of VBR traffic sources but also to eliminate the undesirable call dropping event inherent to circuit-switched cellular systems.

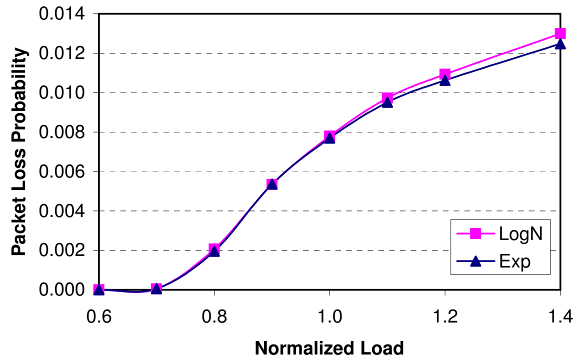
In wireless multimedia networks, there are different service classes, each of which has its own packet and call level QoS constraints. We are currently investigating the extension of PFG to multiple service classes where each service class has its own QoS requirements. Also, a preliminary work shows that by embedding the loss rate into equation (16), PFG is able to have a more precise control on actual packet loss probability.

References

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



(a) Blocking probability.



(b) Packet loss probability.

Fig. 9. Lognormal vs. Exponential residence time.

[2] D. Hong, S. S. Rappaport, Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures, IEEE Transactions on Vehicular Technology 35 (3) (1986) 77–92, see also: CEAS Tech. Rep. No. 773, College of Engineering and Applied Sciences, State University of New York, June 1999.

[3] R. Ramjee, D. Towsley, R. Nagarajan, On optimal call admission control in cellular networks, ACM/Kluwer Wireless Networks 3 (1) (1997) 29–41.

[4] M. Naghshineh, M. Schwartz, Distributed call admission control in mobile/wireless networks, IEEE Journal on Selected Areas in Communications 14 (4) (1996) 711–717.

[5] D. Levine, I. Akyildiz, M. Naghshineh, A resource estimation and call admission algorithm for wireless multimedia networks using the shadow cluster concept, IEEE/ACM Transactions on Networking 5 (1) (1997) 1–12.

[6] B. M. Epstein, M. Schwartz, Predictive QoS-based admission control for multiclass traffic in cellular wireless networks, IEEE Journal on Selected Areas in Communications 18 (3) (2000) 523–534.

[7] S. Wu, K. Y. M. Wong, B. Li, A dynamic call admission policy with precision QoS guarantee using stochastic control for mobile wireless networks, IEEE/ACM Transactions on Networking 10 (2) (2002) 257–271.

[8] M. Ghaderi, R. Boutaba, Call admission control in mobile cellular networks: A comprehensive survey, Wireless Communications and Mobile Computing (WCMC) 6 (1) (2006) 69–93.

[9] Y. Fang, Y. Zhang, Call admission control schemes and performance analysis in wireless mobile networks, IEEE Transactions on Vehicular Technology 51 (2) (2002) 371–382.

[10] Q. Gao, A. Acampora, Performance comparisons of admission control strategies for future wireless networks, in: Proc. IEEE WCNC'02, Vol. 1, Orlando, USA, 2002, pp. 317–321.

[11] M. Ghaderi, R. Boutaba, Call admission control for voice/data integration in broadband wireless networks, IEEE Transactions on Mobile Computing 5 (3) (2006) 193–207.

[12] J. S. Evans, D. Everitt, Effective bandwidth based admission control for multi-service CDMA cellular networks, IEEE Transactions on Vehicular Technology 48 (1) (1999) 36–46.

[13] M. M. Krunz, J. G. Kim, Fluid analysis of delay and packet discard performance for QoS support in wireless networks, IEEE Journal on Selected Areas in Communications 19 (2) (2001) 384–395.

[14] J. S. Evans, D. Everitt, On the teletraffic capacity of CDMA cellular networks, IEEE Transactions on Vehicular Technology 48 (1) (1999) 153–165.

[15] L. Xu, X. Shen, J. W. Mark, Dynamic fair scheduling with QoS constraints in multimedia W-CDMA cellular networks, IEEE Transactions on Wireless Communications 3 (1) (2004) 60–73.

[16] L. Huang, C.-C. J. Kuo, Joint connection-level and packet-level quality-of-service support for VBR traffic in wireless multimedia networks, IEEE Journal on Selected Areas in Communications 23 (6) (2005) 1167–1177.

[17] R. Guerin, H. Ahmadi, M. Naghshineh, Equivalent capacity and its application to bandwidth allocation in high-speed networks, IEEE Journal on Selected Areas in Communications 9 (7) (1991) 968–981.

[18] M. Ghaderi, R. Boutaba, Mobile effective bandwidth and its application to admission control in cellular packet networks, Tech. Rep. CS-2004-61, School of Computer Science, University of Waterloo (November 2004).

[19] J. M. Peha, A. Sutivong, Admission control algorithms for cellular systems, ACM/Kluwer Wireless Networks 7 (2) (2001) 117–125.

[20] M. Ghaderi, R. Boutaba, Stochastic admission control for quality of service in wireless packet networks, Tech. Rep. CS-2004-60, School of Computer Science, University of Waterloo (November 2004).

[21] M. Schwartz, Broadband Integrated Networks, Prentice Hall, New Jersey, USA, 1996.

[22] D. Y. Eun, N. B. Shroff, A measurement-analytic approach for QoS estimation in a network based on the dominant time scale, IEEE/ACM Transactions on Networking 11 (2) (2003) 222–235.

[23] T. Karagiannis, M. Molle, M. Faloutsos, Long-range dependence: Ten years of Internet traffic modeling, IEEE Internet Computing 8 (5) (2004) 57–64.

[24] T. Zahariadis, D. Kazakos, (R)evolution toward 4G mobile communication systems, IEEE Wireless Communications Magazine 10 (4) (2003) 6–7.

[25] S. Y. Hui, K. H. Yeung, Challenges in the migration to 4G mobile systems, IEEE Computer 41 (12) (2003) 54–59.

[26] C. Jedrzycki, V. C. M. Leung, Probability distribution of channel holding time in cellular telephone systems, in: Proc. IEEE VTC'96, Vol. 1, Atlanta, GA, 1996, pp. 247–251.