

ON BANDWIDTH RESERVATION POLICIES IN BROADBAND WIRELESS NETWORKS

Jelena Mišić and Tam Yik Bun

*Department of Computer Science, Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
{jmisic,csben}@cs.ust.hk*

Abstract Adaptive admission control reserves bandwidth for handoff calls by distributing bandwidth reservation values for each call to the neighboring base stations. The series of bandwidth reservation values distributed after k -th handoff determines the bandwidth reservation policy. This paper analyzes relationship between bandwidth reservation policies and call level QoS when user mobility is changing. We define classes of non-aggressive and aggressive bandwidth reservation policies. We show that bounded call level QoS under heterogeneous user mobilities can be achieved only by using aggressive bandwidth reservation policies. Examples of non-aggressive and aggressive bandwidth reservation policies are shown.

Keywords: multimedia wireless networks, adaptive call admission control, quality-of-service

1. INTRODUCTION

Third generation wireless networks should provide controlled call level QoS under the wide range of offered load and user mobilities. This should be provided through bandwidth reservation mechanism within wireless call admission control algorithm (WCAC). The WCAC algorithm must adapt to the traffic intensities in the surrounding cells, which can change when either new call arrival rates or user mobilities change. As the result, the handoff dropping probability and forced call termination probability must be bounded under different network conditions. Adaptive bandwidth reservation under uniform traffic conditions and under homogeneous user mobilities has been addressed through three main research directions:

Periodical, short-term handoff load estimation. New call admission decision is made according to the estimated short-term handoff load. Handoff load is estimated periodically based on channel occupancy in the neighboring cells [4].
Life-time handoff load estimation calculated over a number of equal time periods. Adaptive admission policy based on the *shadow cluster* concept has

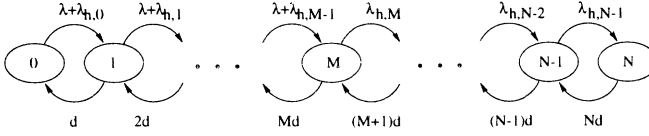


Figure 1 Markov chain for single type of traffic (where $d = h + \mu$)

been proposed in [1]. It is based on the calculation of probabilities that the mobile terminal will visit cells within its shadow cluster in consecutive, equal time periods during its life-time.

Event-based, short-term handoff load estimation. In [2], an adaptive WCAC algorithm is proposed in which the calculation of the estimated handoff load is triggered by certain events like handoff, origination and termination of calls in the surrounding cells.

Recently, adaptive WCAC with event-based handoff load estimation has been analyzed under non-uniform new call arrival rates and homogeneous mobilities in [3]. It determines the ranges of call level QoS when the call arrival rates are varying in the group of cells in the network.

This paper relates the bandwidth reservation policy within WCAC with the state probabilities of the underlying Markov chain, and derives its properties needed for the bounded call level QoS. The paper is organized as follows. Section 2 describes the Markov chain model for wireless cell with admission control. In Section 3 the bandwidth reservation process is described. Section 4 presents the performance analyses of the admission algorithm under uniform load with variable user mobility. In Section 5 we analyze the required properties of the bandwidth reservation policies and propose some of them. Section 6 concludes the paper.

2. SYSTEM MODEL

Consider a wireless cellular system with hexagonally tessellated cells. We use the assumptions that the probability to handoff to any of the six surrounding cells is equal to $1/6$ and that the call duration and dwell time are exponentially distributed with parameters μ and h respectively. The cell capacity is N channels (in all further calculations we will use $N = 50$). Only narrow-band traffic is assumed consuming one channel per user under uniform traffic intensity in all the cells.

The Markov chain that represents system for the arbitrary admission algorithm for a single type of traffic is given in Fig. 1. The states in the chain represent the number of ongoing calls in the target cell. New call arrival rate is represented by λ . In this approach we address the observation that λ_h in the target cell should be decreasing function of the number of utilized channels

in the target cell. Therefore we assume that the sum of current utilizations of the target cell and the cells from the surrounding ring is constant and equal to the sum of their average utilizations. In this case, for target cell with k ongoing calls, there are $7N\rho_{av} - k$ users in the surrounding ring with the “arriving parameter” h giving the handoff arrival rate: $\lambda_{h,k} = \frac{(7\rho_{av}N-k)h}{6}$.

The value $M = \rho_{max}N$ is the state where the call admission algorithm starts rejecting new calls. Beyond state M , only handoff calls are accepted. State probabilities of the Markov chain are then equal to:

$$P_k = \frac{\left(\frac{c}{6}\right)^k \binom{A_1}{k}}{\sum_{i=0}^M \left(\frac{c}{6}\right)^i \binom{A_1}{i} + \frac{\binom{A_1}{M}}{\binom{A_2}{M}} \sum_{i=M+1}^N \left(\frac{c}{6}\right)^i \binom{A_2}{i}} \quad (0 \leq k \leq M) \quad (1)$$

$$P_k = \frac{\left(\frac{c}{6}\right)^k \binom{A_2}{k} \frac{\binom{A_1}{M}}{\binom{A_2}{M}}}{\sum_{i=0}^M \left(\frac{c}{6}\right)^i \binom{A_1}{i} + \frac{\binom{A_1}{M}}{\binom{A_2}{M}} \sum_{i=M+1}^N \left(\frac{c}{6}\right)^i \binom{A_2}{i}} \quad (M < k \leq N) \quad (2)$$

where $c = h/(h + \mu)$, $A_1 = 7N\rho_{av} + \frac{6\lambda}{h}$ and $A_2 = 7N\rho_{av}$. The average utilization and probabilities of new call blocking and handoff call dropping for this system are therefore given by:

$$\rho_{av} = \frac{1}{N} \sum_{k=0}^N k P_k, \quad P_B = \sum_{k=M}^N P_k, \quad P_{hd} = P_N. \quad (3)$$

Since we consider systems with hard bounds on the handoff dropping probability $P_{hd} < 10^{-2}$, the number of new calls from both expressions can be approximated with $n_0 = \rho_{av}N \frac{\mu}{h+\mu}$. Furthermore, the number of calls in the target cell which have executed k handoffs is: $n_k = n_0 \left(\frac{h}{h+\mu}\right)^k$.

3. BANDWIDTH RESERVATION PROCESS

Every handoff call from the finite user population surrounding the target cell arrives independently from the others with the average rate h . In order to implement the bandwidth reservation to match the handoff arrival rate λ_h , we create the *bandwidth reservation rate* in the target cell using the whole user population in the first surrounding ring. Since the average dwell time $1/h$ may be different for various users and we don't know which mobile users will handoff to the target cell, mobile users contribute to the bandwidth reservation rate using some tentative numbers which we will denote as *bandwidth reservation values*. Every call that arrives to the cell either as a new call or handoff call will send bandwidth reservation value to all the surrounding cells. When the call leaves the cell that number is cleared from the surrounding cells. Every base station maintains the sum of those numbers associated with the bandwidth requirement

of the call. By analogy to the handoff call arrival rate, the sum of bandwidth reservation values corresponds to the average bandwidth reservation rate of the *bandwidth reservation process*. Therefore, the bandwidth reservation rate in the target cell is:

$$\nu = \sum_{k=0}^{\infty} 6n_k a_k \quad (4)$$

where n_k denotes the number of calls with k executed handoffs, and a_k denotes bandwidth reservation value sent after k -th handoff. After substituting value for n_k the average reservation rate becomes:

$$\nu = \rho_{av} N 6 \sum_{k=0}^{\infty} \frac{\mu}{\mu + h} \left(\frac{h}{\mu + h} \right)^k a_k. \quad (5)$$

The average amount of bandwidth reservation values sent out by any connection from the ring surrounding the target cell is:

$$B = \frac{\nu}{\rho_{av} N} = 6 \sum_{k=0}^{\infty} \frac{\mu}{\mu + h} \left(\frac{h}{\mu + h} \right)^k a_k. \quad (6)$$

Due to uniform cell capacity utilization and symmetry, connections from the target cell will send out the same average amount of bandwidth reservation values. In the text that follows we shall refer to B as the *bandwidth reservation parameter*. If we denote mobility parameter $\frac{h}{h+\mu} = c$, the bandwidth reservation parameter B becomes $B = 6(1 - c) \sum_{k=0}^{\infty} c^k a_k$.

In case the value of mobility parameter c and bandwidth reservation policy a_k are known, the value of reservation rate can be calculated as: $\nu = \rho_{av} N B(c, a_k)$. The bandwidth reservation parameter B depends on the mobility parameter of the call, and on the policy for bandwidth reservation a_k .

3.1 TUNING THE BANDWIDTH RESERVATION

In order to do proper reservation of the bandwidth for various values of user's average dwell times, and for various bounds on handoff dropping probability, we choose bandwidth reservation values such that average bandwidth reservation rate is much larger than the handoff call arrival rate. In this case the regulating parameter is needed to determine which portion of the bandwidth reservation rate is needed to achieve the required QoS. We choose this regulating parameter to be the probability that "the number of the arrivals of the bandwidth reservation process" will exceed the currently unused capacity of the cell. This parameter is denoted as the *overload probability* and for the single traffic type it is defined

as:

$$P_{ov} = e^{-v} \sum_{q=\lceil N(1-\rho_{curr.}) \rceil}^{\infty} \frac{v^q}{q!} = \frac{1}{\Gamma(N(1-\rho_{curr.}))} \int_0^v e^{-x} x^{N(1-\rho_{curr.})} dx. \quad (7)$$

where $\rho_{curr.}$ denotes portion of the cell capacity currently occupied by the ongoing calls. By varying the parameter P_{ov} , it is possible to control the utilization in the cell, which means that handoff dropping probability can be controlled. Also, if v changes due to the traffic fluctuations in the surrounding cells, for constant P_{ov} , $\rho_{curr.}$ will fluctuate in order to respond to different bandwidth reservation.

3.2 ADMISSION ALGORITHM BASED ON P_{OV}

We shall illustrate the algorithm using cell (0, 0); but due to symmetry of the hexagonal network any other cell can be considered. The main steps of this algorithm are:

Step 1. Variables v_j are initialized to zero and current utilization $\rho_{curr.}$ is set to zero. The values $a_{j,k}$ are chosen according to some QoS guaranteeing policy.

Step 2. When a new type j call arrives, the overload probability has to be calculated using expr. (7) in cell (0, 0) and in the surrounding cells.

P_{ov} checking in cell (0, 0) : current utilization is incremented by α_j and the overload probability $P_{ov}(0, 0)$ is recalculated with the new value of $\rho_{curr.}$. If the updated value is larger than the predetermined limit, the new call is blocked.

P_{ov} checking in surrounding cells : In each cell (1, i), ($0 \leq i \leq 5$), the value of local v_j is incremented by $a_{j,0}$ and the overload probability is recalculated. If the overload probability bound in any surrounding cell (1, i) is violated, the new call must be blocked.

Step 3. The calls which are admitted in cell (0, 0) can freely handoff to neighboring cells without invoking the call admission algorithm again. The handoff call is dropped only when there is insufficient bandwidth for serving the call. When a type j call executes k -th handoff from cell (0, 0) to a neighbor cell (1, i), and given that available bandwidth in the cell is sufficient for the call, the following actions must be performed:

Step 3.1. Update load PGF in cell (1, i), which means that $\rho_{curr.}$ is incremented by α_j , v_j is decremented by the bandwidth reservation value $a_{j,k}$.

Step 3.2. Update load PGF in the neighborhood of cells (0, 0) and (1, l). This action consists of the following:

Step 3.2.1. The new reservation values are distributed to cells surrounding cell (1, i) and the corresponding values of v_j are updated. Note that some cells not previously included in the neighborhood of cell (0, 0) will receive the call's reservation factor for the first time. This reservation factor will be included in the execution of the admission algorithm for future new calls in these cells.

Step 3.2.2. The cells bordering the cell that was hosting the call before the handoff (0, 0), but which do not belong to the 1-ring neighborhood of the cell currently hosting the call, should delete the appropriate reservation factor.

Step 4. When a call terminates, all corresponding bandwidth reservation values in the neighboring cells are deleted and $\rho_{curr.}$ in the hosting cell is decremented by α_j .

4. QOS UNDER CHANGING MOBILITY

In this section we will relate the Markov chain presented in Fig. 1 with the admission algorithm from section 3.2. We use the threshold value of overload probability P_{ovT} to determine the threshold utilization ρ_{max} at which admission algorithm starts rejecting new calls.

Given the threshold value of overload probability P_{ovT} the admission equation derived from eqn. (7) is:

$$P_{ovT} = \frac{1}{\Gamma(N(1 - \rho_{max}))} \int_0^{\rho_{av}NB} e^{-x} x^{N(1 - \rho_{max})} dx. \quad (8)$$

The relationship between ρ_{av} and ρ_{max} through the clipping state in the Markov chain $M = \rho_{max}N$ is:

$$\rho_{av} = \frac{1}{N} \sum_{k=0}^N k P_k \quad (9)$$

With knowledge of the mobility parameters h , μ and B , and given the new call arrival rate λ , the pair (ρ_{max}, ρ_{av}) can be found by numerically solving the system of equations (9, 8).

Behavior of the QoS when the user mobility, or bandwidth reservation policy is changing (given constant new call arrival rate) can be obtained by considering either the handoff dropping probability or the forced call termination probability. The handoff dropping probability is determined by the expr. 2 for $k = N$. The forced call termination probability is determined by the following expression [5]:

$$P_{fct} = \frac{P_{hd}}{\mu/h + P_{hd}} \quad (10)$$

Note that P_{fct} as the call level QoS guarantee can be of more interest to the mobile user than P_{hd} , since the user is interested in the success of the entire call.

The problem of controlling call level QoS when the user mobility is changing can be therefore defined as the following:

Determine the bandwidth reservation policy a_k which will offer

$$P_{fct} = \chi(h, \mu) \quad (11)$$

(or $P_{hd} = \psi(h, \mu)$) when handoff rate is changing in the range $(h_{min}..h_{max})$ and call departure rate is changing in the range $(\mu_{min}.. \mu_{max})$. The QoS shaping function $\chi(h, \mu)$ is determined by the network operator.

The task of finding the suitable bandwidth reservation policy requires solving of the system of equations (8, 9, 11) for the variables ρ_{av} , ρ_{max} , B . When the set of the values for the bandwidth reservation parameter are known, regression technique can be used to transform the set of points $[h, \mu, B]$ to the appropriate function $B(h, \mu)$. After that, $B(h, \mu)$ is transformed in series $B(c) = 6(1 - c) \sum_{k=0}^{\infty} c^k a_k$.

5. TYPES OF BANDWIDTH RESERVATION POLICIES

Better understanding of the role of the bandwidth reservation policy under changing handoff rates can be obtained if we look at the derivatives of the system of eqns. (9, 8, 11). Those derivatives can be obtained by considering state probabilities P_k given in expressions (1) and (2) as $P_k(\rho_{av}, \rho_{max}, h)$ and P_{ov} as $P_{ov}(\nu, \rho_{max})$:

$$\frac{\partial \rho_{av}}{\partial h} = \frac{1}{N} \sum_{k=0}^N k \frac{\partial P_k}{\partial \rho_{av}} \frac{\partial \rho_{av}}{\partial h} + \frac{1}{N} \frac{\partial (\sum_{k=0}^N k P_k)}{\partial \rho_{max}} \frac{\partial \rho_{max}}{\partial h} + \frac{1}{N} \sum_{k=0}^N k \frac{\partial P_k}{\partial h} \quad (12)$$

$$\frac{\partial P_{ov}}{\partial \nu} \frac{\partial \nu}{\partial h} + \frac{\partial P_{ov}}{\partial \rho_{max}} \frac{\partial \rho_{max}}{\partial h} = 0 \quad (13)$$

$$\psi'(h) = \frac{\partial P_N}{\partial \rho_{av}} \frac{\partial \rho_{av}}{\partial h} + \frac{\partial P_N}{\partial \rho_{max}} \frac{\partial \rho_{max}}{\partial h} + \frac{\partial P_N}{\partial h} \quad (14)$$

Here we again observe simplifying approximation applicable to the eqn. (13) similar to the one used in the analysis of QoS when new call arrival rate is changing [3]. The P_{ov} value corresponds to the area below the integrand function $f(x) = \frac{e^{-x} x^{N(1-\rho_{max})}}{\Gamma(N(1-\rho_{max}))}$ in the segment $x = [0, \nu]$, depicted in Fig. 2. The function $f(x)$ has the maximum at $x = N(1 - \rho_{max})$. When handoff rate h changes, the average bandwidth reservation rate will change due to the changes of average utilization ρ_{av} and bandwidth reservation parameter B , that is $\Delta \nu = N(\Delta \rho_{av} B + \Delta B \rho_{av})$. The character of the change $\Delta \nu$ depends on the particular bandwidth reservation policy, i.e. on the sign and absolute value of ΔB . Therefore, we have to consider two cases:

Aggressive bandwidth reservation policies. In such policies $\Delta \nu = N(\Delta \rho_{av} B + \Delta B \rho_{av}) > 0$ when handoff rate h grows by Δh . In this case, the end of the integrating segment ν tends to shift to the right towards the higher values of P_{ov} . However, P_{ov} regulation maintained by the admission algorithm has to keep the area below the function at the constant value P_{ovT} , so it will then shift the integrand function $f(x)$ to the right which is equivalent to the decrease of

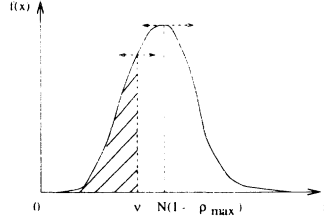


Figure 2 Representation of the utilization regulation property of the admission algorithm

the value of ρ_{max} .

Non-aggressive bandwidth reservation policies. In such policies $\Delta v = N(\Delta\rho_{av}B + \Delta B\rho_{av}) \leq 0$ when handoff rate h grows by Δh . In this case, the end of the integrating segment v tends to shift to the left towards the lower values of P_{ov} . In order to keep the value of overload probability at level of P_{ovT} , the admission algorithm will increase the value of ρ_{max} and admit more calls.

Since $\Delta v \approx N\Delta(1 - \rho_{max})$ when handoff rate h changes the relationship between v and ρ_{max} is:

$$\frac{\partial v}{\partial h} \approx -\frac{N\partial\rho_{max}}{\partial h}. \quad (15)$$

which shows that clipping utilization ρ_{max} is approximately linearly proportional to the bandwidth reservation rate v . Since $v = \rho_{av}NB$, and by applying approximation (15) the equation (13) becomes:

$$\left(\rho_{av}\frac{\partial B}{\partial h} + \frac{\partial\rho_{av}}{\partial h}B\right) + \frac{\partial\rho_{max}}{\partial h} = 0 \quad (16)$$

By combining eqns. (12), (14) and (16) the following differential equations are obtained:

$$\begin{aligned} \frac{\partial\rho_{av}}{\partial h} \left(1 - \frac{1}{N} \sum_{k=0}^N k \frac{\partial P_k}{\partial\rho_{av}} + \frac{B}{N} \frac{\partial(\sum_{k=0}^N k P_k)}{\partial\rho_{max}}\right) = \\ -\rho_{av} \frac{1}{N} \frac{\partial(\sum_{k=0}^N k P_k)}{\partial\rho_{max}} \frac{\partial B}{\partial h} + \frac{1}{N} \sum_{k=0}^N k \frac{\partial P_k}{\partial h} \end{aligned} \quad (17)$$

$$\frac{\partial\rho_{av}}{\partial h} \left(B \frac{\partial P_N}{\partial\rho_{max}} - \frac{\partial P_N}{\partial\rho_{av}}\right) = -\rho_{av} \frac{\partial P_N}{\partial\rho_{max}} \frac{\partial B}{\partial h} + \frac{\partial P_N}{\partial h} - \psi'(h) \quad (18)$$

From the equations (16) and (18), $B(h)$ function can be determined such to satisfy the required criterion. This is the system of two differential equations

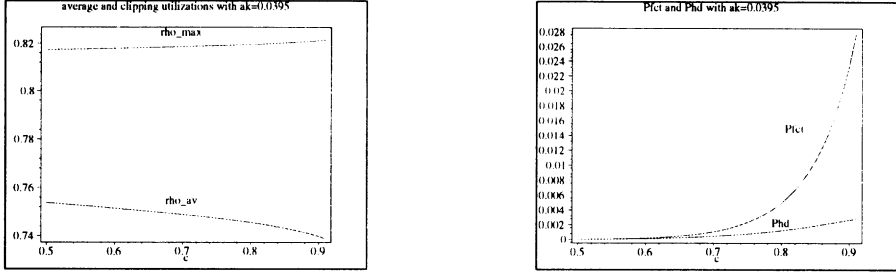


Figure 3 Call QoS behavior with constant bandwidth reservation parameter B , $\lambda = 0.1$, $c = \frac{h}{h+\mu}$, $\mu = 0.002$.

with changing parameters which again can be solved only numerically. Conversely, if we explicitly choose function $B(h, \mu)$, then we will obtain QoS characteristics $P_{fct} = \chi(h, \mu)$, as the result of bandwidth reservation policy.

In the following examples, we shall investigate explicit bandwidth reservation policies, i.e. the cases when the function $B(h, \mu)$ is known, and analyze the obtained P_{fct} and P_{hd} . In our analysis, the handoff rate h varies, while the call departure rate is fixed at the level $\mu = 0.002$. In all calculations we will assume that network operates at the nominal load $\frac{\lambda_{nom}}{N\mu} = 1$. Threshold for the admission algorithm is set to $P_{ovT} = 0.5$. The calibrating point for all policies is $B = 0.237$ under $h = 0.01$, $\mu = 0.002$ (i.e. $c = 0.833$).

Non-aggressive bandwidth reservation policy. If bandwidth reservation values are constant (independent of the number of previously executed handoffs) i.e. $a_k = a_0$, the bandwidth reservation factor becomes $B = 6a_0$. The value of $a_0 = 0.395$ is chosen to calibrate the admission system to give the desired tuple (B, h, μ) . In this approach both handoff dropping probability and the forced call termination probability will increase with the growth of mobility parameter c as shown in Fig. 3 which prevents us from giving any guarantee on QoS as mobility changes. The consequence of the increasing forced call termination probability under fixed bandwidth reservation scheme is the decreasing average utilization as shown in Fig. 3.

Aggressive bandwidth reservation policies. Assume that a_k is proportional to the number of executed handoffs. Such approach will increase bandwidth reservation in two cases:

- a) Calls with short dwell times and moderate duration times will increase bandwidth reservation according to their mobility dynamics.
- b) Very long calls with moderate dwell times will implement the "call aging" concept to decrease the probability of forced termination when call has already executed a number of handoffs.

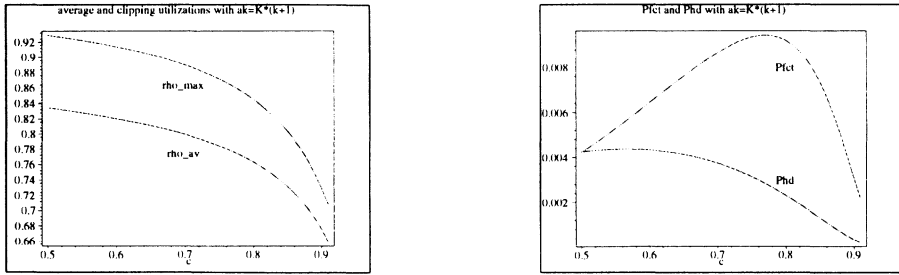


Figure 4 Call QoS behavior with bandwidth reservation parameter $B = \frac{6K}{1-c}$, $\lambda = 0.1$, $c = \frac{h}{h+\mu}$, $\mu = 0.002$.

As an example let us assume that $a_k = K(k+1)$. In this case, $B = 6K(1-c) \sum_{k=0}^{\infty} (k+1)c^k = \frac{6K}{1-c}$. This policy is easy to implement, since it does not involve computation, or storage at the mobile terminals.

6. CONCLUSION

In this paper we have analyzed the problem of controlling call level QoS under adaptive admission algorithm when the user mobility changes. We have shown that aggressive bandwidth reservation policies are needed to provide bounds on the forced call termination probability. In the future work, we will analyze QoS when both handoff rate and call departure rate are varying.

References

- [1] D. A. Levine, I. F. 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, 1997, pp. 1–12., 1997.
- [2] J. Mišić, S. T. Chanson and F. S. Lai. "Event Based resource Estimation in Admission Control for Wireless Networks with Heterogeneous Traffic." *Mobile Computing and Communications Review*, vol. 1, no. 4, 1997.
- [3] J. Mišić and Y. B. Tam. "About the problem of hot-spots under adaptive admission control in the multimedia wireless networks." *to appear in proceedings of WCNC'99*, Sep. 1999.
- [4] M. Naghshineh and M. Schwartz. "Distributed call admission control in mobile/wireless networks." *IEEE Journal on Selected Areas in Communications*, vol. 14, no. 4, pp. 711–717, May 1996.
- [5] T.S. Rappaport. "Models for call handoff schemes in cellular communication networks." In *3rd WINLAB Workshop: Third Generation Wireless Information Networks*", Workshop Record, Apr. 28, 1992.