The Shadow Cluster Concept for Resource Allocation and Call Admission in ATM-Based Wireless Networks

David A. Levine Ian F. Akyildiz

Broadband and Wireless Networking Lab
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332
Tel: (404)-894-5141; Fax: (404)-853-9410
E-Mail: dlevine@nemesis.mirc.gatech.edu; ian@armani.gatech.edu

Mahmoud Naghshineh

IBM T.J. Watson Research Center
Yorktown Heights, NY 10598
Tel: (914)-784-6231
E-Mail: mahmoud@watson.ibm.com

Abstract

This paper describes the *Shadow Cluster* concept, a novel idea that can be used to improve the resource allocation and the call admission procedures in wireless networks. Shadow clusters can be used to allocate resources that need to be reserved for call hand-offs, and to determine if a new call should be admitted to a wireless network based on the call's requirements and local traffic conditions. The shadow cluster concept is targeted for ATM-based wireless networks with a micro/nano-cellular architecture, where service will be provided to users with very diverse requirements. In these networks, and as a consequence of the small cell sizes, mobile users will typically experience a high number of cell hand-offs during their connections' lifetime. With shadow clusters, the quality of service of mobile calls can be improved by reducing the number of dropped calls during hand-offs, and by disallowing the establishment of new calls that are highly likely to later result in a dropped call. The framework of a shadow cluster system is completely distributed, and can be viewed as a message system where a mobile terminal informs the base stations in the neighborhood about its requirements, position, and movement parameters, so that the base stations project future demands, reserve resources accordingly, and admit only those calls that can be supported adequately. In this paper, we describe how base stations define and maintain shadow clusters by multicasting probabilistic information on the future position of their mobiles with active calls. In addition, we propose resource allocation and call admission algorithms based on the information provided by the shadow clusters.

MOBICOM 95 Berkeley CA USA

Keywords: Shadow Cluster, Active Mobile Probability, Resource Allocation, Call Admission.

1 Introduction

ATM technology has been introduced in wireline networks due to its intrinsic ability to handle traffic with very diverse characteristics and requirements. It appears that wireless networks based on ATM technology will also be ideal for handling the traffic produced by emerging and future mobile applications. Not surprisingly, the expansion of ATM to the wireless environment has already begun; several ATM-compatible wireless systems have already been proposed [1,2,4]. These and future ATM-compatible systems will consist of a wireless network connected to an ATM wireline infrastructure. The resulting hybrid network will enable a mobile terminal to access practically all of the services and applications available to terminals connected directly to a wireline network. Obviously, ATM-based wireless networks will be compatible with ATM-wireline networks only to a certain extent, primarily due to the limitations in the available bandwidth and the high bit error rate of the wireless systems.

The bandwidth in a wireless network is perhaps the most precious and scarce resource of the entire communication system. It is of great importance to use this resource in the most efficient manner. A base station sometimes may need to reserve resources, even if this means denying access to a mobile terminal requesting admission to the network, in order to keep enough resources to support active users currently outside of its coverage area, but who may soon decide to emigrate to its cell. Thus, base stations must maintain a balance between two conflicting requirements: 1) maintain maximum resource (bandwidth) utilization, and 2) reserve enough bandwidth resources so that the maximum rate of unsuccessful incoming hand-offs (due to insufficient resources) is kept below an acceptable level. For the second requirement, the maximum rate of unsuccessful hand-offs can be established in terms of a Quality of Service (QoS) parameter, e.g., call blocking probability, that the network agrees to maintain.

An accurate determination on the number of resources that a base station must reserve (to maintain a certain call blocking probability) is likely to become a very important issue in future wireless networks. First, and in contrast to current systems, future wireless networks will support a wide range of applications, each with its own bandwidth requirements. The demand on bandwith resources within a cell may change abruptly in a short lapse, e.g., video-voice and high rate data users entering or leaving a cell, while at present, it usually varies gradually, and thus, it is much easier to handle. Also, future wireless networks should provide customized QoS parameters (such as call blocking probability) on a per call and/or on a service basis.

Issues and relationships between resource reservation, channel assignment, call admission, and traffic intensity have been studied before [3,5,6]. Admission control policies that determine the number of new voice or data users that are accepted in a packet radio network are presented in [6]. For these policies, voice users are accepted only if a long-term blocking probability is satisfied, while data users are accepted only if the mean packet delay and the packet loss probability are maintained below certain levels. In [5], a "flexible" channel assignment scheme is proposed based on the analysis of offered traffic distributions or blocking probabilities. A distributted call admission control procedure is proposed in [3], which takes into consideration the number of calls in adjacent as well as in the cell where a new call request is being made, in addition to the knowledge on the mean call arrival, call departure, and call hand-off rates. We propose the so-called "shadow cluster concept", which is a predictive resource allocation scheme that provides high wireless network utilization by dynamically reserving only those resources that are needed to maintain the call blocking probability (a QoS parameter) requested by the connection. The shadow cluster scheme is *dynamic and pro-active*, i.e., the amount of resources to be reserved is determined "on-the-fly", and the control functions on call admissions are aimed at preventing congestion conditions. In addition, the shadow cluster scheme is designed to operate in future ATM-based wireless networks. By using shadow clusters, the network's quality of service can be improved by reducing the number of dropped calls and by disallowing the establishment of new calls that are highly likely result later in a dropped call. Shadow clusters are best suited for wireless networks with a small cell size and with a high number of cell hand-offs during the life time of the average mobile connection. The shadow cluster concept is completely distributed; it can be implemented over the *virtual connection tree* described in [1], and requires some communication between a mobile's base station and its neighbors. Since base stations will be linked by ATM switches that use high-bandwidth optical fiber, the amount of extra packets generated by the mechanism should be easily manageable by the wireline network.

This paper is organized as follows. In Section 2, we describe physical and virtual reference architectures for an ATM-compatible wireless network. In Section 3, we present a general description of the shadow cluster scheme. In Section 4, we study how shadow clusters can be used to estimate future resource demands in a cell. Also in this section, we present the *active mobile probability* concept, used by base stations to inform their neighbors about the probable future locations of active mobiles in the network. Using the procedures outlined in Section 4, we develop a call admission algorithm in Section 5. Finally, in Section 6 we discuss some future research directions.

2 ATM-Based Wireless Network Architecture and the Virtual Connection Tree

A number of ATM-based architectures suitable for integrating wireless networks to a B–ISDN network have been proposed in recent years [1,2,4]. These ar-

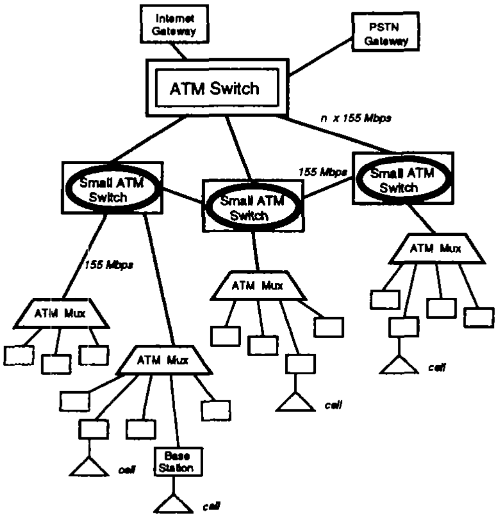

Figure 1: An ATM-Based Wireless Network Architecture

chitectures typically consist of a backbone network structure with three hierarchical layers (from bottom to top): 1) Base stations with their associated mobile stations, 2) ATM multiplexers interconnecting a collection of base stations, and 3) ATM switches interconnecting a collection of ATM multiplexers. The topmost layer is usually further divided into a sublayer of small ATM switches, and above them an ATM switch for the topmost sublayer. Some of the small ATM switches can be interconnected among them, and the access to the wireline B–ISDN is provided by the larger ATM switch, which belongs to the wireline B–ISDN. An example architecture is shown in Figure 1.

On top of this physical arrangement, a *virtual connection tree* architecture is adapted in [1]. The virtual connection tree can be regarded as a tree where the root is the ATM switch providing access to the B–ISDN, while the leaves are the base stations served by the ATM switch. This virtual architecture is created when a mobile user tries to get access to the network by executing a call setup procedure. The process, if successful, establishes a *virtual connection*, and is accomplished in two steps. First, the fixed portion of the virtual connection is established between the root of the tree (topmost ATM switch connected to the mobile's base station) and the appropriate fixed point of the wired network (this is the same process already required in B–ISDN before being extended to mobile scenarios). Second, a virtual connection tree between the root and its leaves (the base stations connected to the ATM switch) is established by assigning two sets of virtual circuit numbers (VCNs). These numbers define the paths between the root and each one of the base stations attached to that root. Two sets of numbers are required because each link requires two VCNs (one for each direction). In any case, for each active call, only two VCIs (for the uplink and downlink connections with the base station chosen by the mobile) are in use at any time.

Once its connection demand in the wireless portion of the network is admitted, a mobile user can freely hand-off to any base station within the virtual connection tree. As long as handoffs occur in cells that are within a given connection tree, no significant additional processing and delay overheads are implied. The mobile only needs to use the corresponding pair of VCNs that have been reserved for the mobile in that cell. However, when a mobile crosses connection-tree boundaries, the previously described call set up procedure must be executed again to establish a new connection tree, just as is the case when a new mobile wants to enter the network. In the following section, we describe how the shadow cluster concept uses virtual connection trees.

3 The Shadow Cluster Concept

Consider a micro/nano-cell wireless network system that can support mobile terminals running applications that demand a wide range of bandwidth resources. Mobile users can freely roam within the network's coverage area, and experience a large number of hand-offs during a typical connection. The wireless network users expect good quality of service from the system, e.g., small connection set up times, low delays, and small call dropping and packet loss probabilities. The wireless network has to provide the level of requested service even if an active mobile user moves to a crowded cell. When this happens, the corresponding base station must provide the expected service even if this implies denying network access to new connection requests. Ideally, base stations should deny network access to certain connection requests only when it is *strictly necessary*. This constitutes a problem that could only be optimized if knowledge were available regarding the future movement and holding times of the active mobiles[1] in the wireless network, as well as the future movements and holding times of the mobiles with connection requests. Like in similar problems, solutions close to the optimal can be obtained by using

[1]By active mobile, we mean a mobile that has a wireless connection and is consuming wireless resources.

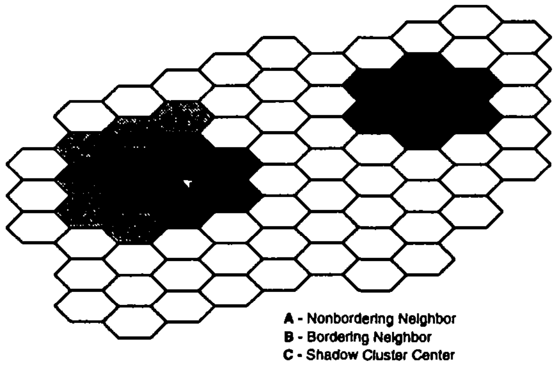

Figure 2: Shadow Clusters Produced by Two Sample Mobiles

knowledge of *past events* to predict future behavior.

The fundamental idea of the shadow cluster scheme is that *every mobile with an active wireless connection establishes an influence upon the cells (and their base stations) in the vicinity of its current location and its direction of travel.* As an active mobile travels and hand-offs to other cells, the region of influence also moves, following the active mobile to its new location. The base stations (and their cells) currently being influenced are said to conform a *shadow cluster*, because the region of influence follows the movements of the active mobile like a shadow, as shown in Figure 2. The shadow (and therefore the level of influence) is strongest near the active mobile, and fades away as a function of factors such as the distance to the mobile, current call holding time and priority, bandwidth resources being used, and the mobile's current trajectory and velocity. Because of these factors, the shape of a shadow cluster is usually not circular and can change over time.

We say that the center of a shadow cluster is not the geometric center of the area described by the shadow, but the cell where the mobile is currently located. We also refer to this cell as the mobile's *current home cell.* A *bordering neighbor* is a cell that shares a common border with the shadow cluster's center cell. In contrast, a *nonbordering neighbor* cell, although being a part of the shadow cluster, does not share a border with the shadow cluster's center cell. In a manner consistent with the above definitions, we also use the terms *current home base station*, *bordering base station*, and *nonbordering base station.* Conceptually, the amount and "darkness" of the shadows that cover a cell (as shown in Figure 2) reflect the amount of resources that the cell's base station needs to support the active mobiles currently in its own and in neighboring cells. With this information, base stations can determine, on an individual basis, if new call requests can be supported by the wireless network. In practice, a shadow cluster is a virtual message system where a base station informs its neighbors about the probability that its active mobiles will move to their neighbor cells in the future. With this information, neighbor base stations project future demands and reserve resources accordingly. Base stations reserve resources by denying network access to new call requests, and by *waiting* for active mobiles to end their calls.

The decision process for the acceptance of a new call request also involves a shadow cluster. Every new call request results in the implementation of a *tentative* shadow cluster, and the establishment of one or more virtual connection trees, where two or more virtual connection trees are required if the shadow cluster falls over two or more disjoint virtual connection trees. Base stations exchange information on their new call requests, and decide, based on this and other information, which requests should be accepted and which requests should be denied. When implementing shadow clusters, it is important to consider that the amount of information shared among neighbor base stations should be limited, so that the wireline network is not overburdened with control messages. In practice, after the existence of a mobile call has been reported, only a limited amount of information needs to be shared.

When an active mobile hand-offs to another cell, the shadow cluster moves. After a hand-off, all of the base stations within the old shadow cluster must be notified about this movement, and the mobile's new current base station has to assume the responsibility of supplying the appropriate information to the base stations within the new shadow cluster. In addition, new virtual connection trees may need to be established if the new shadow cluster falls over cells that belong to a virtual connection tree different from the one(s) already established for the call. Likewise, if the shadow cluster no longer falls over a virtual connection tree previously used by the call, this virtual connection tree is torn down. Base stations that were in an old shadow cluster but that are not in a new one must delete any entries corresponding to the active mobile establishing the shadow cluster, and free reserved resources if appropriate. Base stations that enter into a shadow cluster that has recently moved must be given appropriate information on the shadow cluster's active mobile, such as the respective QoS requirements, e.g., bandwidth demands, and any other useful information, e.g., the wireless connection's elapsed time, for the new shadow cluster.

4 Resource Demand Prediction and Shadow Clusters

A concept central to the implementation of shadow clusters is the projected probability that an active mobile will be in a certain cell at a particular time. We explore this issue as well as a procedure to establish future resource demands in this section.

4.1 Active Mobile Probability

In a wireless network that is implementing shadow clusters, every base station must inform its neighbors about the future location probabilities of each of the active mobiles currently under its control. For each active mobile, only those base stations that belong to that particular mobile's shadow cluster are informed about these probabilities. The information that each base station receives consists primarily of the probabilities that active mobiles will be in that base station's cell at future times.

Consider a wireless network where time is discretized in steps of length τ. Let j denote a base station in the network, where $j \in J$, and J is the set of all base stations in the network. Since there is a one-to-one correspondence between the set of base stations and the set of cells in the wireless network being considered, we use the same notation to denote the cells in the network. Let i be a mobile with an active wireless connection, where $i \in I$, and I is the set of all active mobiles at the present time. Excluding the current home base station, the set of base stations that form the shadow cluster of active mobile i is denoted by $K(i)$, where $k \in K(i)$ is one of the base stations that form the shadow cluster. If active mobile i is currently under the control of base station j, then it is the task of this base station to determine the projected *active mobile probabilities* $\mathbf{Pam}(i,j,k,\tau) = \{P_{am}(i,j,k,t_1),\ P_{am}(i,j,k,t_2),\ \ldots,\ P_{am}(i,j,k,t_n)\}$ that mobile i, currently in cell j, will be active in cell k (and therefore under the control of base station k) at times $t_1, t_2, \ldots, t_n$. Thus, the active mobile probabilities P_{am} are the projected probabilities that a mobile will remain active in the future *and* at a particular location (within the shadow cluster).

The parameters $\mathbf{Pam}(i,j,k,\tau)$ can be interpreted as the percentage of the total amount of resources currently being used by mobile i that base station j recommends to base station k to have available at times $t_1, t_2, \ldots, t_n$ in the event that active mobile i advances to cell k. As time goes on, mobile i's current home base station recomputes and refines the estimates on the probabilities $\mathbf{Pam}(i,j,k,\tau)$, and sends the respective updates to all base stations within the shadow cluster $K(i)$. Active mobile probability estimates for the short term future are likely to be more accurate than the corresponding estimates for the distant future. Also, bordering neighbor base stations as well as the base stations that are near the most probable paths of the active mobile are likely to receive more active mobile probability points and updates than the rest of the base stations in the active mobile's shadow cluster. In general, the further the base station is from the active mobile, the less information and updates it is likely to receive.

Evidently, the probability that active mobile i, currently in cell j, will be in cell k at times $t_1, t_2, \ldots, t_n$ constitutes a random process that is a function of several parameters. The more knowledge about the dynamics of a mobile (past and present) and call holding patterns, the more complex and accurate the probability function is likely to become. For example, if precise knowledge about the current dynamics of a mobile is available, the active mobile probability becomes a function of the position, velocity, and acceleration vectors of the mobile. The active mobile probability function can also be refined with information such as the mobile's past history, and with detailed knowledge of the geographical features of the region where the mobile is currently located. For example, if the mobile's position coincides with that of a highway, both the mobile's general velocity and direction can be predicted to a certain extent.

Next, we proceed to describe in detail how the active mobile probabilities P_{am} can be used to determine the amount of resources required by a base station, so that most current active mobiles within the base station's cell and nearby cells can engage in successful hand-offs.

4.2 Resource Demand Calculations

Consider a simplified wireless network system where the principal responsibilities of a base station are to provide (with a reasonable probability of success): 1) successfull incoming hand-offs from active mobiles with different bandwidth requirements, and 2) that if a new connection request is accepted by the base station, it will not be dropped soon due to a lack of bandwidth resources in nearby cells. In this section, we show how the shadow cluster concept can help with these objectives by predicting the amount of resources that a station will require in its near future.

We consider that every base station has a total of *Ct bandwidth units* (BU), where a bandwidth unit is the minimum amount of (uplink) bandwidth that can be assigned to any mobile user. For example, while a voice call user may require the least bandwidth, i.e., a single BU, a video mobile terminal will likely require several BUs. The parameter Ct comprises the number of free bandwidth units Cf and the number Cu of bandwidth units that are currently being used. The number of free and "used" BUs varies over time. Their relation with the total number of BUs is

$$Ct(t) = Cu(t) + Cf(t) \qquad (1)$$

Here, we consider that the parameter Ct is constant over time. However, the approach that we propose next can be extended to include values of Ct that change over time. The number of bandwidth units being used in a cell changes over time. When a base station receives hand-off requests, it has to honor them as best as possible, trying to maintain the number of dropped calls to a minimum. Dropped calls are generally the result of *cell overloading*, a condition characterized by an excessive number of active users that results in an insuficient number of free resources left in a cell receiving a hand-off. Since a base station cannot control neither the number of hand-off requests nor the number of call disconnect requests, the best mechanism available to the base station to prevent/reduce overloading is to limit the number of newly accepted connections. Obviously, drastic measures could be taken, i.e., a wireless connection can be either terminated prematurely or dropped during a hand-off, but should be avoided as much as possible so as to keep a good quality of service.

Using information provided by shadow clusters, each base station can obtain estimates on the number of BUs that are going to be used in the near future, *considering only the current active mobiles in the wireless network*. Based on these estimates, base stations can then decide, using a distributed algorithm, whether or not to accept new connection requests.

Assuming that an active mobile cannot enter or exit a cell multiple times within the time interval being considered, we observe that an active mobile that is in a reference cell has three possible outcomes:

- It can remain active and stay within the reference cell.
- It can terminate the connection without leaving the cell.
- While active, it can move to neighbor cells.

Likewise, an active mobile currently within a neighbor cell but moving towards a reference cell has three possible outcomes:

- It can move to the reference cell and stay active in this cell.
- It can enter the reference cell briefly and then exit.
- It can enter the reference cell and end its connection.

Based on these observations, we can determine the components for an estimate on the number of BUs $Cu_j^*(t)$ to be used by base station j at times $t_1, t_2, t_3, \ldots, t$ (where $t_2 - t_1 = \tau, t_3 - t_2 = \tau, \ldots$) by using:

$$Cu_j^*(t) = Cu_j(t_o) - Cu_j^I(t) - Cu_j^{II}(t) + Cu_j^{III}(t) - Cu_j^{IV}(t) - Cu_j^V(t) \qquad (2)$$

Let $c(i)$ denote the number of BUs being used by active mobile i, and let also

$$P_{tot}(i,j,k,t) = \frac{1}{t - t_1 + 1} \sum_{t_1}^{t} P_{am}(i,j,k,t) \qquad (3)$$

denote the *total active mobile probability* that active mobile i, currently in cell j, will be active in cell k within the time interval $[t_1, t]$. Then, the terms of $Cu_j^*(t)$ can be determined according to:
$Cu_j(t_0)$, the initial number of busy BUs, is a known quantity.

$$Cu_j^I(t) = \sum_{i \in Ij} \left(1 - P_{tot}(i,j,j,t) - \textstyle\sum_{k \in K(i)} P_{tot}(i,j,k,t)\right) c(i) \qquad (4)$$

where $Cu_j^I(t)$ is the number of BUs that become free by active mobiles that end their calls within the time interval $[t_1, t]$. In Eq.(4), $P_{tot}(i,j,j,t)$ is the total active mobile probability that a mobile remains active and in cell j within the time period $[t_1, t]$.

$$Cu_j^{II}(t) = \sum_{i \in Ij} \sum_{k \in K(i)} P_{tot}(i,j,k,t) c(i) \qquad (5)$$

where $Cu_j^{II}(t)$ is the number of BUs that become free by active mobiles that emigrate from cell j to bordering cells within the time period $[t_1, t]$.

$$Cu_j^{III}(t) = \sum_{i \notin Ij} P_{tot}(i,k,j,t)c(i) \quad (6)$$

where $Cu_j^{III}(t)$ is the number of BUs that become busy due to hand-offs by external mobiles within cell j in the time period $[t_1, t]$, and $Ij \subset I$ is the set of all active mobiles within cell j.

$$Cu_j^{IV}(t) = \sum_{i \notin Ij, j \in K(i)} \sum_{k \in K'(i)} P_{tot}(i,j,k,t)c(i) \quad (7)$$

where $Cu_j^{IV}(t)$ is the number of BUs that become free by external active mobiles that enter and leave the reference cell within the time interval $[t_1, t]$, and $K'(i) \in J$ is the set of all cells where an external active mobile may be found at time t after briefly transiting through the reference cell.

$$Cu_j^{V}(t) = \sum_{i \notin Ij, j \in K(i)} \alpha_{tot}(P_{tot}(i,k,j,t), M(i), Et(i,t), t)\ c(i) \quad (8)$$

where $Cu_j^{V}(t)$ is the number of BUs that become free when an external mobile enters the reference cell and terminates its wireless connection within the time interval $[t_1, t]$, and

$$\alpha_{tot}(P_{tot}(i,j,k,t), M(i), Et(i,t), t) = \frac{1}{t-t_1+1} \sum_{t_1}^{t} \alpha(P_{tot}(i,j,k,t), M(i), Et(i,t), t) \quad (9)$$

where α is a function that determines the probability that an external active mobile, after entering a reference cell, will terminate its connection before the time interval $[t_1, t]$. Arguments of the α function include the total active mobile probability $P_{tot}(i,k,j,t)$ that an external active mobile i will move to the reference cell j, a connection class description $M(i)$ of active mobile i, e.g., video, audio, voice, and the total elapsed time $Et(i,t)$ of the call at time t.

From Eqs. (3)-(8), it is clear that base stations do not need to transmit individual values of active mobile probabilities $P_{am}(i,j,k,t)$ to their neighbors; rather, they only need to transmit values that correspond to sums of products, i.e., active mobile probabilities P_{am} multiplied by the number of BUs $c(i)$ being used.

Ideally, each base station recomputes these estimates at each time step, and determines the values for the expected BU occupancy values for the next $t_1, t_2, \ldots, t$ steps, i.e., $Cu_j^*(t_1)$, $Cu_j^*(t_2)$, ..., $Cu_j^*(t)$ by using recursion. When these estimates are obtained, a base station j can determine the estimates for the number of free BUs $Cf_j^*(t_1)$, $Cf_j^*(t_2)$, ..., $Cf_j^*(t)$ for the next $t_1, t_2, \ldots, t$ time steps simply by using

$$Cf_j^*(t) = Ct_j - Cu_j^*(t) \quad (10)$$

Note that these estimates are based on only the current active mobiles in the network, and considering that no new connection requests are accepted during the time interval $[t_1, t]$. If the estimates for $Cu^*(t)$ yield values smaller than Ct, then the base stations will be able to admit new active mobiles to the wireless network. In the following section, we use the estimates developed in Eqs. (2)-(10), and propose a decision algorithm for the acceptance of new mobiles to the wireless network.

5 Call Admission Algorithm Based on Shadow Clusters

The call admission algorithm is implemented in a distributed fashion, with every base station exchanging information with its neighbors continuously. As before, we assume that time is discretized in slots of length τ. Also, we assume that new call requests are reported at the beginning of each time slot, and that a decision regarding an admission request is made sometime before the end of the same time slot where the request was received.

The call admission algorithm is applicable to both mobile-initiated calls and calls initiated by users connected to the wireline network (assuming that the target mobile has already been located). Without loss of generality, we explain the call admission algorithm from the perspective of base station j. Let $c(i')$ denote the number of BUs that mobile i' is requesting for admission to the wireless network. The sequence of steps that base station j executes in every time slot is the following:

1. Base station j gathers call connection requests from mobiles within its cell. Each received request includes a description of the desired QoS parameters and the number of BUs being requested for the call. The base station checks if its current resources and current physical environmental conditions can support the requested parameters, e.g., for requesting mobile i', base station j checks if $c(i') \leq Cf_j(t_0)$. If it is evident that one or more of the requirements from a given mobile user cannot be satisfied, the request is immediately turned down. If $\sum_{i'} c(i') > Cf_j(t_0)$, the decision of which connection requests should be turned down is resolved later.

2. Base station j defines a shadow cluster $K(i')$ for each mobile i' making an admission request, and informs the base stations within the corresponding shadow clusters about these requests. The base station also informs its neighbors about the preliminary active mobile probabilities and the number of requested BUs for each of the mobiles making the call requests.
3. Base station j receives the preliminary active mobile probability estimates and number of requested BUs from neighbor base stations, which have previously received their own admission requests. Base station j receives this information only if it is within the shadow cluster of a new connection request currently in another cell. Based on the active mobile probabilities from its own new requests, and on the new requests in neighbor cells, base station j computes *occupancy estimates* $Oe_j(t)$ for the future $t_1, t_2, \dots t$ time steps by using the following expression:

$$Oe_j(t) = \frac{Cf_j^*(t)}{\sum_{i' \in I_j} P_{tot}(i',j,j,t)c(i) + \sum_{i' \notin I_j} P_{tot}(i',k,j,t)c(i)} \quad (11)$$

Eq. (11) gives an indication on the base station's free BU utilization expectancy for each of the next $t_1, t_2, \dots t$ steps if *all of the new call requests in the wireless network at time* t_0 *were accepted.* The higher the value of $Oe_j(t)$, the more likely a mobile in cell j will have bandwidth resources available at time t. Base station j distributes the occupancy estimates $Oe_j(t_1)$, $Oe_j(t_2), \dots Oe_j(t)$ among the base stations that sent earlier data about their own new connection requests. Conversely, base station j receives occupancy estimates from all base stations that fall within the shadow clusters of base station j's new call requests. Note that if a given base station k is part of two or more shadow clusters defined by connection requests in cell j, then base station j receives only a single set of occupancy estimates from base station k.
4. After having collected occupancy estimates from neighbor base stations, base station j computes *survivability estimates* $Se_j(i', t)$ for each new mobile i' making a connection request in cell j. The survability estimate of mobile i' for the next t time steps $t_1, t_2, \dots t$ can be calculated from

$$Se_j(i',t) = O_j(t)P_{tot}(i',j,j,t) \cdot \prod_{k \in K(i')} Oe_k(t)P_{tot}(i',j,k,t) \quad (12)$$

In Eq. (12), equal treatment is given for all connection requests regardless of the number of BUs being requested. However, Eq. (12) can be adjusted to give preference to new connection requests with multiple BUs. The higher the value of $Se(i',t)$, the more likely mobile i' will survive at times $t_1, t_2, \dots t$. Thus, mobiles with higher values for their survivability estimates should be accepted more often than mobiles with lower values. The preliminary decision on whether to accept or reject a connection request is handled by a decision function $\Omega(Se(i',t), Q_{dp}(i'))$, where $Q_{dp}(i')$ is the *dropping probability* QoS parameter requested by mobile i'. $\Omega()$ should be defined so as to favor the acceptance of mobiles with higher overall survivability estimates, provided that the requested dropping probability parameter is satisfied. $\Omega()$ should be optimized so that the network achieves maximum occupancy with an acceptable number of dropped calls. If in a given base station, and after the preliminary acceptance decisions, it is found that the total number of requested BUs exceeds the current number of free channels, then the request(s) with the smallest value(s) for the survivability estimate is(are) dropped. This procedure is followed provided that for each of the other accepted calls, a viable connection can be established through the wireline network.

6 Conclusions and Future Research

Shadow clusters are likely to be most useful in future wireless networks with small cells, with irregular and time varying traffic loads, e.g., time varying "hot spots", and an average high number of cell hand-offs per call.

We have shown algorithms that can be used to implement shadow clusters. These algorithms presume knowledge about the probability that a mobile will be active in a given cell and at particular times. The accuracy in determining these probabilities depends on the amount of knowledge that is available about the mobiles. In future work we plan to explore methods to determine active mobile probabilities *Pam* based on the information type available about a mobile. Since the effectiveness of the shadow cluster concept depends on the accuracy in the determination of the active mobile probabilities *Pam*, at present time it is not possible to quantitatively evaluate the performance of the shadow cluster concept.

More accurate results could be obtained if in the shadow cluster's equations other probabilistic moments, e.g., variances, were used. Unfortunately, as a tradeoff, the additional probabilistic information would also add complexity to the algorithms, resulting in higher communication burdens on the wireline network, and requiring higher processing capabilities on the base stations.

The call admission algorithm proposed in this paper usually favors the admission of mobiles that have a relaxed call dropping probability over those that require a stringent call dropping probability. In future work we also plan to improve the fairness of this algorithm.

Acknowledgement

The authors wish to thank Dr. Zygmunt Haas for his valuable suggestions and insightful discussions.

REFERENCES

[1] A.S. Acampora, M. Naghshineh, "An Architecture and Methodology for Mobile-Executed Handoff in Cellular ATM Networks," *IEEE Journal on Selected Areas in Comm.*, Vol. 12, No. 8, pp. 1365–1374, Oct. 1994.

[2] M.J. McTiffin, A.P. Hulbert, T.J. Ketseoglou, W. Heimsh, G. Crisp, "Mobile Access to an ATM Network Using a CDMA Air Interface," *IEEE Journal on Selected Areas in Comm.*, Vol. 12, No. 5, pp. 900–908, June 1994.

[3] M. Naghshineh, M. Schwartz, "Distributed Call Admission Control in Mobile/Wireless Networks," submitted, *PIRMC'95.*

[4] D. Raychaudhuri, N.D. Wilson, "ATM Based Transport Architecture for Multiservices Wireless Personal Communication Networks," *IEEE Journal on Selected Areas in Comm.*, Vol. 12, No. 8, pp. 1401–1414, Oct. 1994.

[5] J. Tajima, K. Imamura, "A Strategy for Flexible Channel Assignment in Mobile Communication Systems," *IEEE Transactions on Vehicular Technology*, Vol. 37, No. 2, pp. 92–103, May 1988.

[6] W.-B. Yang, E. Geraniotis, "Admission Policies for Integrated Voice and Data Traffic in CDMA Packet Radio Networks," *IEEE Journal on Selected Areas in Comm.*, Vol. 12, No. 4, pp. 654–664, May 1994.