Handoff Analysis of the Hierarchical Cellular System

Sung-Hong Wie, Jae-Shin Jang, Byung-Cheol Shin, and Dong-Ho Cho

Abstract—As mobile telephone traffic demands increase in the future, interest in cellular systems with a hierarchical structure will grow. Hierarchical cellular system can provide high system capacity, efficient channel utilization, and an inherent load-balancing capability. In this type of system, calls blocked in a microcell flow over to a macrocell which is superimposed on a number of microcells. When overflow calls which are accommodated in a macrocell cannot be returned to a microcell, the macrocell channels become saturated. Thus, we propose a hierarchical cellular system with an underflow scheme which permits overflow calls to return to microcells. We model this underflow scheme and analyze system performance.

Index Terms—Handoff, hierarchical, overflow, underflow.

I. INTRODUCTION

SMALL cell systems allow greater spectral reuse, larger capacity, and use of low power hand-held user devices. However, small cell systems induce an increase in the number of cell boundary crossings by mobile users. The required exchange of supervisory messages can become burdensome. As cell size becomes smaller, the probability that a user crosses a cell boundary during a call increases, and the forced termination probability that a call is interrupted due to handoff failure becomes larger.

The hierarchical cellular system can accommodate users with different mobilities in one personal communication system [1]–[4] and reduce the forced termination probability. Reference [1] introduced a handoff priority scheme using priority queueing. Reference [2] investigated an overflow scheme in a hierarchical cellular system and modeled overflow traffic as IPP traffic. References [3] and [4] also described an overflow scheme in a hierarchical system and provided an analysis with different user mobilities. Reference [3] showed the load-balancing capability of a hierarchical system with an overflow scheme and modeled overflow traffic as a Poisson phenomenon. Reference [4] modeled overflow traffic as MMPP traffic. References [1]–[4] classified mobile users into high and low mobility groups in which few low mobility users would cross a cell boundary while a call was in progress. A small cell size increases the probability of cell boundary crossings by a high mobility user. Such cell boundary crossings require handoff’s where the continuation of a call depends on the availability of channels in the cell to which the vehicle moves. Thus, some fraction of handoff attempts fail and the corresponding calls are forced to terminate. Forced termination is considerably less desirable from the user’s viewpoint than the occurrence of new call blocking. Therefore, large macrocells are dedicated to high mobility users while smaller microcells service low mobility users in a hierarchical system. High mobility users experience fewer cell boundary crossings. Therefore, their forced termination probability is lower than when all users are served in a cell of the same size.

To reduce the forced termination probability of low mobility users, these users are directed to a larger macrocell which is superimposed on a group of microcells, if there are no idle channels in a microcell. The systems modeled by [1]–[4] will cause low mobility users to occupy some channels in macrocells. In this situation, because channels cannot be released during the duration of a call, some high mobility users will not be served in the macrocell. Therefore, the forced termination probability of high mobility users is larger than low mobility users. To alleviate these problems we propose a hierarchical cellular system with an underflow scheme. Reference [5] considered a reverse hierarchical scheme but failed to include a regular cell structure and an underflow traffic model where overflow calls can be returned to the microcell. Because our proposed scheme allows overflow low mobility users to return to a microcell, most macrocell channels will be available for high mobility users. Thus, the scheme reduces both the new call blocking probability and the forced termination probability of high mobility users. This is especially important in a nonuniform traffic situation where some microcells have a large traffic load while others have a small load. In this situation, the underflow scheme provides a considerable decrease in the call blocking probability and the forced termination probability in the macrocell. In a cellular system with an underflow scheme, the handoff scenarios of low mobility users can be described as follows: If the system finds idle channels in the next microcell, low mobility user overflow calls into a macrocell would be directed to the next microcell when users cross the microcell boundary. Therefore, low mobility users would temporarily occupy idle channels in the macrocell prior to transfer to idle channels in an adjoining microcell.

The hierarchical system is described and modeled in Section II. A performance analysis is presented in Section III. Numerical results are presented in Section IV. Finally, conclusions are given in Section V.

II. SYSTEM DESCRIPTION AND MODEL

We consider two-level duplicated coverage of a particular area which is continuously covered by a set of microcells, as shown in Fig. 1. One macrocell is superimposed on each group of M microcells. We assume that both microcells and macrocells are regular hexagons and that mobile users are uniformly distributed over this area. Mobile users are classified into low mobility pedestrians and high mobility vehicles. A microcell accepts new call and handoff requests from low mobility users. If
there are no idle channels in a microcell, low mobility user requests overflow into a macrocell. Although macrocells are designed for high mobility vehicles, overflow requests can be serviced. If an adjacent microcell has idle channels, overflow calls are directed to that microcell by the underflow scheme when the user crosses the microcell boundary. We assume that both the overflow and underflow traffic are distributed as Markov modulated Poisson process (MMPP) processes. The modeling procedures for the overflow and underflow traffic are presented in Sections II-B and D.

The thick line in Fig. 2 denotes a low mobility call trajectory. If a low mobility user call overflows into the macrocell and an idle channel is available in the next microcell when the user crosses the microcell boundary, the call is directed to the next microcell by the underflow scheme. As shown in Fig. 2, a macrocell “M” is superimposed on microcells “i, j, k, and h” in which a low mobility user is served. In the situation shown in Fig. 2(a), microcells “i, k, and h” have idle channels. However, microcell “j” has no idle channels. Thus, a handoff request into “j” is rejected and the call overflows to the macrocell. An underflow request is sent to microcell “k” when the user crosses the boundary between “j” and “k.” The request is successful and the call is transferred to the microcell. In the situation shown in Fig. 2(b), microcells “j” and “k” have no idle channels. Therefore, the underflow request is rejected when the user crosses the boundary between “j” and “k.” However, the next underflow request is successful. Therefore, the interarrival time of the underflow request is the same as the channel holding time of the low mobility user in a microcell.

We assume that the overall hierarchical system is homogeneous, i.e., all cells of the same hierarchical level are statistically identical. Therefore, we can analyze the overall system by focusing on only one cell in each level and consider the statistical behavior of this focused cell under the condition that all neighboring cells exhibit typical random independent behavior. To render the problem amenable to solution using the notion of the multidimensional birth-death processes, some additional assumptions are needed. We assume that the new call arrival and the handoff call arrival processes follow a Poisson distribution and that both the channel holding time in a cell and the call duration time are distributed exponentially.

In the following section, we model the microcell, the macrocell, the overflow traffic, and the underflow traffic. Fig. 3 shows a block diagram of the hierarchical cellular system. Performance of the microcell is analyzed and the overflow traffic is modeled as MMPP. Performance of the macrocell is analyzed and the underflow traffic is modeled as MMPP. Analysis of the hierarchical system is performed with initial parameters of the underflow traffic \((Q_0, A_0)\) until convergence of the initial parameters.

A. Model of Microcells

Basic notations to be used in the analysis are defined as follows:

- \(C_m\) number of channels allocated to a microcell;
- \(\lambda_0\) new call arrival rate per user;
- \(n_0\) number of noncommunicating users in a microcell;
- \(\lambda_{in}\) handoff call arrival rate per user;
- \(\lambda_{on}\) new call arrival rate per user;
- \(A_{in}\) handoff call holding time distribution;
- \(A_{on}\) new call holding time distribution;
- \(\sigma_{in}\) handoff call routing probability;
- \(\sigma_{on}\) new call routing probability;
- \(\alpha\) fraction of users in a microcell that are communicating;
- \(\beta\) fraction of users in a microcell that are not communicating.

Fig. 1. Hierarchical cellular structure.

Fig. 2. Underflow request of low mobility users: (a) case 1 and (b) case 2.

Fig. 3. Block diagram of the hierarchical system.
WIE et al.: HANDOFF ANALYSIS OF THE HIERARCHICAL CELLULAR SYSTEM

new call arrival rate of low mobility users ($\lambda_n = \lambda_0$);

handoff call arrival rate of low mobility users;

arrival rate matrix and the infinitesimal generator of the underflow traffic;

call duration time and channel holding time;

average call duration time and average channel holding time ($\mu = T_c^{-1}$, $U_I = T_I^{-1}$).

We assume that the number of noncommunicating users, $\lambda_0$, is constant. This number is valid for ($\lambda_0 \gg C_m$) because the number of noncommunicating users is larger than the number of communicating users. That is, the new call arrival rate for low mobility users, $\lambda_n$, is also constant. We assume that a nonpriority scheme is applied in microcells. Microcells do not need a priority scheme because macrocells provide reservation channels. Therefore, new calls and handoff calls can be accepted whenever there are idle channels in microcells. We assume that underflow traffic is a two-state MMPP and both new calls and handoff calls are Poisson processes. We can model the total input traffic toward the microcell as MMPP because a Poisson process is a special case of MMPP [6]. Therefore, we can analyze the microcell by applying an MMPP/M/C_m/C_m system [2]. The infinitesimal generator of the microcell is given by (1), shown at the bottom of the page, where $\Lambda_T = (\lambda_n + \lambda_h)I + \Lambda_d$ and $U_T = (\mu + U_I)I$. Matrix $Q_m$ is $2(C_m + 1) \times 2(C_m + 1)$ matrix. Many quantities of interest may be calculated by studying the Markov chain represented by $Q_m$. All of these require the computation of the probability vector $\mathbf{P}_m = (P_m(0), P_m(1), \ldots, P_m(C_m))$, and the stationary vector of $Q_m$, which satisfies $\mathbf{P}_m Q_m = 0$ and $\mathbf{P}_m e = 1$.

With the steady state probability $\mathbf{P}_m$ we can calculate the blocking probabilities of new calls and handoff calls. First, the blocking probability of new calls and handoff calls can be written as

$$P_{b1} = \mathbf{P}_m(C_m) e_2$$

where $e_2$ is the unit column vector of order 2 ($e_2^T = (1, 1)$). The blocking probability of the underflow traffic, $P_{b2}$, can be given by

$$P_{b2} = \frac{P_m(C_m) \Lambda_d e_2}{\sum_{\lambda=0}^{C_m} P_m(i) \Lambda_d e_2}$$

B. Model of the Overflow Traffic

We use the overflow traffic model described in [2]. When all channels in a microcell are busy (state $C_m$), arrival calls overflow to the macrocell. The instantaneous arrival rate is, therefore, a Poisson process with parameter $(\lambda_n + \lambda_h)$. In all other states ($0, 1, \ldots, C_m - 1$), there are no requests to the macrocell.

The overflow traffic generated by a microcell is modeled by an MMPP which can be expressed by two matrices. These are the infinitesimal generator $Q_m$ of the microcell and the arrival rate matrix, $\Lambda_o$, which has $(2(C_m + 1))$ states, such as

$$\Lambda_o = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & (\lambda_n + \lambda_h)I \end{pmatrix}$$

We can divide microcell states into two groups. One is “all channels are busy” (“on” state), and another is “at least one channel is idle” (“off” state). Overflow traffic from a microcell can be approximated into an interrupted Poisson process (IPP) [2], as shown in Fig. 4. This approximation is proposed in [6]. The following matrix defines an approximated IPP:

$$Q_{IPP} = \begin{pmatrix} -a & a \\ b & -b \end{pmatrix}, \quad \Lambda_{IPP} = \begin{pmatrix} 0 & 0 \\ 0 & (\lambda_n + \lambda_h)I \end{pmatrix}$$

The parameters $a, b$ can be calculated using an approximation in [6]. First, we must calculate the first and second moments of the instantaneous rate of the MMPP $(Q_m, \Lambda_o)$ using the following formula:

$$\alpha^{(i)} = \mathbf{P}_m \Lambda_o^i e.$$ 

We can calculate the variance using $v = \alpha^{(2)} - (\alpha^{(1)})^2$. The time constant is defined as

$$\tau_v = v^{-1} \int_0^\infty \mathbf{P}_m \Lambda_o e^{Q_m t} - e \mathbf{P}_m \Lambda_o e dt$$

$$= v^{-1} \left[ \mathbf{P}_m \Lambda_o (e^{Q_m t} - Q_m)^{-1} \Lambda_o e - (\alpha^{(1)})^2 \right].$$

Because we know the arrival rate matrix of the approximated IPP, we can simply evaluate the parameters $a, b$ as follows:

$$\pi_1 = \sum_i P_m(i) e_2$$

$$\pi_2 = P_m(C_m) e_2$$

$$a = \pi_2 / \tau_v$$

$$b = \pi_1 / \tau_v.$$ 

The arrival rate in the macrocell is a combination of overflow from $M$ microcells. The combination of these $M$ IPPs gives an
Fig. 4. Overflow traffic approximation by interrupted Poisson process (IPP).

MMPP associated with \((2^M \times 2^M)\) matrices. Under the assumption that all microcells are identical, the MMPP may be reduced to \((M + 1)\) states defined by matrix \(Q_U\) and \(A_U\). In state \(i\), \(i\) microcells are busy and the other cells have at least one idle channel. Therefore, the instantaneous arrival rate is \(i(\lambda_n + \lambda_h)\). Therefore, we can write \(Q_U\) and \(A_U\) as (9) and (10), shown at the bottom of the page.

C. Model of Macrocells

Some basic notations necessary to describe macrocell behavior are as follows:

- \(C_M\) number of channels allocated to a macrocell;
- \(C_{M1}, C_{M2}\) number of guard channels for handoff calls and high mobility new calls;
- \(\lambda_p\) new call arrival rate per vehicle;
- \(v_{H0}\) number of noncommunicating vehicles in a macrocell;
- \(v_H, v_L\) number of high mobility vehicles and low mobility users in communication;
- \(\lambda_{Hn}\) new call arrival rate of the high mobility vehicle \((\lambda_{Hn} = v_{H0}\lambda_p)\);
- \(\lambda_{Hh}, \lambda_{Lh}\) handoff call arrival rate of high mobility vehicles and low mobility users;
- \(T_H, T_L\) channel holding time of high mobility vehicles and low mobility users in a macrocell;
- \(\bar{T}_H, \bar{T}_L\) average channel holding time of high mobility vehicles and low mobility users in a macrocell \((U_H = \bar{T}_H^2, U_L = \bar{T}_L^2)\).

We assume that the macrocell is applied according to the guard channel scheme in which first priority is given to handoff calls, second priority is assigned to new calls of high mobility users, and third priority is given to new calls of low mobility users. That is, if the number of idle channels is less than \(C_{M1}\), then handoff calls will be accommodated and other calls will be blocked. Also, if the number of idle channels is greater than \(C_{M1}\) and less than \(C_{M1} + C_{M2}\), then both the new calls of high mobility users and handoff calls will be accommodated and overflow new calls will be blocked. If the number of idle channels is greater than \(C_{M1} + C_{M2}\), then handoff calls, new calls of high mobility users, and overflow new calls of low mobility users will be accommodated.

Over time, the macrocell randomly changes state due to underlying processes and system dynamics. The underlying processes that affect the system are new call generation of high mobility users, handoff call arrivals of high and low mobility users from neighboring macrocells, handoff call departures of high and low mobility users to neighboring macrocells, call completions of high and low mobility users, overflow traffic arrivals of low mobility users from microcells, underflow traffic departures of low mobility users, and state transitions of the MMPP. We assume that the above processes are statistically independent. It should be noted that some physical events that are pertinent to system performance, such as handoff call arrivals when all channels are busy, do not cause state transitions.

1) Flow Balance Equation: We use the methodology of [7] to construct macrocell flow balance equations. The state of a macrocell is characterized by a sequence of nonnegative integers, such as \(v_H, v_L\), and \(X\), where \(X\) is the state of the overflow traffic model represented by MMPP. We give each state an integer index, \(s\), ranging from 0 to \(s_{\text{max}}\) which is determined so that all possible states are accommodated, as shown in [7]. Thus, state \(s\) corresponds to a distinct sequence of nonnegative integers as follows: \((v_H(s), v_L(s), X(s))\). The total number of calls \(J(s)\) in progress for a macrocell in state \(s\) is given by:
\[
J(s) = v_H(s) + v_L(s).
\]

It is convenient to order the states as follows: Primary (first) according to increasing values of the number of occupied macrocell channels \(J(s)\); Secondary (second), according to increasing values of \(v_H(s)\); Tertiary (third), according to increasing values of \(v_L(s)\); Quaternary (fourth), according to increasing values of \(X(s)\).

To find the statistical equilibrium state probabilities for a macrocell, we write the flow balance equations among the states. These are a set of \(s_{\text{max}} + 1\) simultaneous linear equations among the unknown state probabilities, \(P_M = (p_M(0), p_M(1), \ldots, p_M(s_{\text{max}}))\). We require the infinitesimal generator \(Q_M = (q_{ij})\) where \(q_{ij}\) is the transition

\[
Q_U = \begin{pmatrix}
-Ma & Ma \\
b & -(M - 1)a - b & (M - 1)a \\
& & \ddots & \ddots & \ddots \\
& & & Mb & -Mb
\end{pmatrix}
\]

and

\[
A_U = \begin{pmatrix}
0 & (\lambda_n + \lambda_h) \\
(\lambda_n + \lambda_h) & 2(\lambda_n + \lambda_h) \\
& \ddots & \ddots \\
& & & M(\lambda_n + \lambda_h)
\end{pmatrix}
\]
rate from state $i$ to state $j$. The probabilities can be given by solving the flow balance equations of the form: $P_M Q_M = 0$, and $P_M e = 1$.

Now, we can define the state transition rates from state $s$ to $x_{\{1, \ldots, 6\}}$ where $s$ is the current state and $x_{\{1, \ldots, 6\}}$ is the next state $(s, x_{\{1, \ldots, 6\}} \in \{0, 1, 2, \ldots, S_{\text{max}}\})$.

1) Due to call arrivals: Let $r_n(s, x)$ denote the rate component from the current state $s$ to the next state $x$ due to call arrivals. If the state, $x_1$, corresponds to the sequence of \{vt(s) + 1, vl(s), X(s)\}, then the rate is given by

\[ r_n(s, x_1) = \begin{cases} 
    \lambda_{Hn} + \lambda_{Hs}, & \text{if } J(s) < C_M - C_M1 \\
    \lambda_{Hn}, & \text{if } C_M - C_M1 \leq J(s) < C_M.
\end{cases} \]  

Similarly, if the state, $x_2$, corresponds to the sequence of \{vt(s) + 1, vl(s), X(s)\}, then the rate is given by

\[ r_n(s, x_2) = \begin{cases} 
    \lambda_L + X(s)(\lambda_n + \lambda_h), & \text{if } J(s) < C_M - C_M1 \\
    \lambda_L + X(s)\lambda_h, & \text{if } C_M - C_M1 \leq J(s) < C_M.
\end{cases} \]  

(11)

2) Due to call departures: Let $r_c(s, x)$ denote the rate component from the current state $s$ to the next state $x$ due to call departures. If the state, $x_3$, corresponds to the sequence of \{vt(s) - 1, vl(s), X(s)\}, then the rate is given by

\[ r_c(s, x_3) = (\mu + U_H)vt(s). \]  

Similarly, if the state, $x_4$, corresponds to the sequence of \{vt(s) - 1, vl(s), X(s)\}, then the rate is given by

\[ r_c(s, x_4) = (\mu + U_L + U_H(1 - \frac{X(s)}{M})) vl(s). \]  

(13)

In the above equation, a third term occurs due to underflow traffic departures.

3) Due to MMPP state transitions: Let $r_t(s, x)$ denote the rate component from the current state $s$ to the next state $x$ due to MMPP state transitions. If the state, $x_5$, corresponds to the sequence of \{vt(s), vl(s), X(s) - 1\}, then the rate is given by

\[ r_t(s, x_5) = q_U(X(s), X(s) - 1) \]  

where $q_U(i, j)$ is the $(i, j)$ element of matrix $Q_U$ in (4). Similarly if the state, $x_6$, corresponds to the sequence of \{vt(s), vl(s), X(s) + 1\}, then the rate is given by

\[ r_t(s, x_6) = q_U(X(s), X(s) + 1). \]  

(15)

2) Blocking Probability in the Macrocell: There are four types of calls in a macrocell. These are, low mobility new calls, handoff calls, and high mobility new calls, and handoff calls. Each has a blocking probability as follows.

- Low mobility new calls can be served when there are more than $C_{M1} + C_{M2}$ idle channels in a macrocell. We can define the set $B_0$ as: $B_0 = \{s | J(s) \geq C_M - C_{M1} - C_{M2}\}$. If a macrocell is in a state $s$ that belongs to $B_0$, then low mobility new calls are blocked. Thus

\[ P_{BL, n} = \frac{\sum_{s \in B_0} X(s)\lambda_n p_M(s)}{\sum_{s = 0}^{\text{max}} X(s) p_M(s)} = \sum_{s \in B_0} p_M(s). \]  

(17)

- High mobility new calls can be served if there are more than $C_{M1}$ idle channels in a macrocell. Therefore, we can define the blocking set of high mobility new calls as follows: $B_1 = \{s | J(s) \geq C_M - C_{M1}\}$. Thus, the blocking probability of high mobility new calls is given by

\[ P_{BH, n} = \frac{\sum_{s \in B_1} \lambda_n p_M(s)}{\sum_{s = 0}^{\text{max}} \lambda_n p_M(s)} = \sum_{s \in B_1} \lambda_n p_M(s). \]  

(18)

- In a macrocell, there are two scenarios for handoff of low mobility users. These are handoff calls that overflow from a microcell, and handoff calls from neighboring macrocells. The handoff calls are blocked if there are no idle channels in the macrocell. We can define the set $H_0$ as: $H_0 = \{s | J(s) = C_M\}$. Thus, the handoff failure probability of low mobility calls that overflow from a microcell is given by

\[ P_{BL, h1} = \frac{\sum_{s \in H_0} X(s)\lambda_n p_M(s)}{\sum_{s = 0}^{\text{max}} X(s) p_M(s)} = \sum_{s \in H_0} p_M(s). \]  

(19)

and the handoff failure probability of low mobility calls from neighboring macrocells is

\[ P_{BL, h2} = \frac{\sum_{s \in H_0} \lambda_n p_M(s)}{\sum_{s = 0}^{\text{max}} \lambda_n p_M(s)} = \sum_{s \in H_0} \lambda_n p_M(s). \]  

(20)

- Similarly, the handoff failure probability of high mobility calls from neighboring macrocells can be written as

\[ P_{BH, h} = \frac{\sum_{s \in H_0} \lambda_H p_M(s)}{\sum_{s = 0}^{\text{max}} \lambda_H p_M(s)} = \sum_{s \in H_0} \lambda_H p_M(s). \]  

(21)

D. Model of the Underflow Traffic

The underflow arrival matrices, $A_U$ and $Q_U$, must be calculated from the underflow departure matrices. The state transitions of the underflow traffic are based on the state transitions of the macrocell. That is, the infinitesimal generator of the underflow traffic is the same as the infinitesimal generator $Q_M$ of
the macrocell. The total underflow traffic departure rate \( \delta_{d}(s) \) from a macrocell is dependent on the macrocell state \( s \) and the number of low mobility users. That is, \( \delta_{d}(s) \) is given by \( \delta_{d}(s) = U_{FL}(s) \). We assume that a macrocell is superimposed on \( M \) microcells and that microcells are statistically identical. Therefore, the departure rate of the underflow traffic to a microcell, \( \delta_{d}(s) \), is given as: \( \delta_{d}(s) = \delta_{d}(s)/M \). Underflow traffic is a Poisson process with a rate dependent on a macrocell state. In other words, we can model the underflow traffic as MMPP. The matrix \( \Delta_{d} \) showing the departure rate of the underflow traffic to a microcell is denoted as

\[
\Delta_{d} = (d_{i,j}), \quad \text{where} \quad d_{i,j} = \begin{cases} \delta_{d}(i), & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}
\]

(22)

The underflow departure traffic to a microcell is an MMPP process, \((Q_{M}, \Delta_{d})\). The \(Q_{M}\) and \(\Delta_{d}\) are \((s_{\text{max}}+1)\times(s_{\text{max}}+1)\) matrices. If \(s_{\text{max}}\) is large, then analysis of the hierarchical cellular system is difficult. Therefore, we approximate a \((s_{\text{max}}+1)\)-state MMPP into a two-state MMPP with \((Q_{d}, \Delta_{d})\), as shown in [6]. With these results we can model the underflow traffic as two-state MMPP with \((Q_{d}, \Delta_{d})\).

### III. Performance Analysis

#### A. Blocking Probabilities

Blocking of a low mobility new call occurs when the new call that has been rejected in a microcell is also rejected in an overlaying macrocell. Thus, the blocking probability of a low mobility new call is as follows:

\[
P_{L,n} = P_{bL} P_{BL,n}.
\]

(23)

Also, the blocking probability of a high mobility new call is as follows:

\[
P_{H,n} = P_{BH,n}.
\]

(24)

#### B. Forced Termination Probability

The forced termination probability is the probability that a call which is not initially blocked will be interrupted due to handoff failure during the call. First, we consider low mobility users. If a new call is accommodated by the cellular system, the probabilities that this new call is accommodated by a microcell and a macrocell are denoted by \( S_{1} \) and \( S_{2} \), respectively. The probability that a new call is accommodated by the system is \((P_{L,n})\). The probabilities that a new call of a low mobility user will be served by a microcell and macrocell are \((1 - P_{bL})\) and \((P_{bL}(1 - P_{BL,n}))\), respectively. Therefore, \( S_{1} \) and \( S_{2} \) can be written as

\[
S_{1} = \frac{1 - P_{bL}}{1 - P_{L,n}}
\]

(25)

\[
S_{2} = \frac{P_{bL}(1 - P_{BL,n})}{1 - P_{L,n}}.
\]

(26)

Note that \( S_{1} + S_{2} = 1 \).

Let \( a_{1} \) be the probability that a low mobility call currently served by a microcell requires another handoff request which is accommodated by neighboring microcells. Similarly, let \( a_{2} \) denote the probability that the call requires another handoff request which fails. Then, using the Markovian properties of the model, we determine

\[
a_{1} = \left(1 - P_{bL}\right) \frac{U_{L}}{\mu + U_{T}}
\]

(27)

\[
a_{2} = P_{bL} \frac{U_{L}}{\mu + U_{T}}.
\]

(28)

A low mobility call currently served in a macrocell will require either an underflow request to a microcell, or a handoff to neighboring microcells, or the call will be completed. Let \( b_{1} \) be the probability that a low mobility call currently served by a macrocell requires an underflow request to a microcell which will be accommodated. Let \( b_{2} \) be the probability that the call requires an underflow request which fails. These are then given as

\[
b_{1} = \left(1 - P_{bL}\right) \frac{U_{L}}{\mu + U_{T} + U_{L}}
\]

(29)

\[
b_{2} = P_{bL} \frac{U_{L}}{\mu + U_{T} + U_{L}}.
\]

(30)

Let \( c_{1} \) be the probability that a low mobility call currently served by a macrocell requires a handoff request to a neighboring macrocell which will be accommodated. Let \( c_{2} \) be the probability that the call requires a handoff request which fails. These are given as

\[
c_{1} = \left(1 - P_{bL, h2}\right) \frac{U_{L}}{\mu + U_{T} + U_{L}}
\]

(31)

\[
c_{2} = P_{bL, h2} \frac{U_{L}}{\mu + U_{T} + U_{L}}.
\]

(32)

Now let us focus on a new call that is initially accommodated. When this call leaves the service of a microcell, it can be handed off to the macrocell or forced into termination. After a call that is currently accommodated by a macrocell is handed off several times to neighboring macrocells, the call requires an underflow request to a microcell. The call can be accommodated by a microcell on the underflow request or be forced into termination by a handoff failure. The probability that a low mobility call currently accommodated by a microcell is forced into termination is \( P_{ft1} \), and \( P_{ft2} \) is the probability that a low mobility call currently accommodated by a macrocell is forced into termination. Then, \( P_{ft1} \) can be written as

\[
P_{ft1} = \sum_{i=0}^{\infty} a_{1}^{i} a_{2} P_{BL, h1} + \sum_{i=0}^{\infty} a_{1}^{i} a_{2}(1 - P_{BL, h1}) P_{ft2}.
\]

(33)

In the above equation, the first term is the probability that the \( (i + 1) \)th handoff request fails and there are no idle channels in the macrocell for the blocked handoff call. The second term is the forced termination probability that the handoff call which is rejected by a microcell is accommodated by the macrocell and
is terminated by some macrocell event. Similarly, \( P_{jt2} \) can be written as
\[
P_{jt2} = \sum_{j,k=0}^{\infty} \frac{(j+k)}{j} b_1^j b_2^k c_2
\]
\[
+ \sum_{j,k=0}^{\infty} \frac{(j+k)}{j} b_1^j b_2^k b_1 P_{ft1}.
\]
(34)

In (34) the first term is the probability that the underflow request of the overflow handoff call fails \( j \) times, while the macrocell handoff succeeds \( k \) times and is terminated by a macrocell handoff failure. The second term is the probability that the macrocell handoff request succeeds \( k \) times, while the \((j+1)\)th underflow request succeeds and is terminated by some microcell event. Here, \( \Theta_1 \) and \( \Theta_2 \) are defined as follows:
\[
\Theta_1 = \sum_{j,k=0}^{\infty} \frac{(j+k)}{j} b_1^j b_2^k
\]
(35)
\[
\Theta_2 = \sum_{i=0}^{\infty} a_1^i \Theta_2.
\]
(36)

Then, we can write \( P_{ft1} \) and \( P_{jt2} \) as
\[
P_{ft1} = P_{BL,H} \Theta_2 + (1 - P_{BL,H}) \Theta_2 P_{jt2}
\]
(37)
\[
P_{jt2} = \Theta_1 \Theta_2 + \Theta_1 b_1 P_{ft1}.
\]
(38)

Solving (37) and (38) with respect to \( P_{ft1} \) and \( P_{jt2} \), we have
\[
P_{ft1} = \frac{P_{BL,H} \Theta_2 + (1 - P_{BL,H}) \Theta_2 \Theta_2 P_{jt2}}{1 - (1 - P_{BL,H}) \Theta_1 \Theta_2 b_1}
\]
(39)
\[
P_{jt2} = \frac{\Theta_1 \Theta_2 + P_{BL,H} \Theta_2 + (1 - P_{BL,H}) \Theta_2 (1 - \Theta_2 b_1)}{1 - (1 - P_{BL,H}) \Theta_1 \Theta_2 b_1}.
\]
(40)

Using these results, we find the expression for the forced termination probability of the low mobility user as follows:
\[
P_{FT,L} = S_1 P_{ft1} + S_2 P_{jt2}
\]
\[
= S_2 \Theta_1 \Theta_2 + (S_1 + S_2 \Theta_1 b_1)
\]
\[
\times \frac{P_{BL,H} \Theta_2 + (1 - P_{BL,H}) \Theta_2 \Theta_2 P_{jt2}}{1 - (1 - P_{BL,H}) \Theta_1 \Theta_2 b_1}.
\]
(41)

Because high mobility users are served only by a macrocell, we can simply calculate the forced termination probability \( P_{FT,H} \). Let \( d_1 \) be the probability that a high mobility call currently served by a macrocell requires another handoff request which is accommodated by a neighboring macrocell. Similarly, let \( d_2 \) denote the probability that the call requires another handoff request which fails. Then, we calculate
\[
d_1 = (1 - P_{BH,H}) \frac{U_H}{\mu + U_H}
\]
(42)
\[
d_2 = P_{BH,H} \frac{U_H}{\mu + U_H}.
\]
(43)

From these results, we find the expression for the forced termination probability of the high mobility users as follows:
\[
P_{FT,H} = \sum_{i=0}^{\infty} d_1 d_2 = \frac{d_2}{1 - d_1}.
\]
(44)

C. Carried Traffic

Another important system performance measure is carried traffic. For a given number of channels, a large carried traffic value implies efficient use of bandwidth. Carried traffic per microcell, \( E_m \), and carried traffic per macrocell, \( E_M \), can be easily calculated once the state probabilities are determined. They are simply the average number of occupied channels per cell, and can be as given by
\[
E_m = \sum_{i=0}^{C_m} \sum_{j=0}^{\infty} j \rho_m(i) \epsilon_2
\]
(45)
\[
E_M = \sum_{i=0}^{\infty} \rho_m(i) PM(i).
\]
(46)

Because a macrocell is superimposed on \( M \) microcells, the total carried traffic in the area covered by one macrocell is given by
\[
E_{total} = ME_m + EM
\]
\[
= M \sum_{i=0}^{C_m} \sum_{j=0}^{\infty} j \rho_m(i) \epsilon_2 + \sum_{i=0}^{\infty} \rho_m(i) PM(i).
\]
(47)

D. System Event Rate

It is useful to define an overall system parameter where an event may be a new call attempt, a handoff request, a call termination, or an underflow request. A larger event rate implies a larger control load on the network. The event rate per microcell \( R_m \) is written as
\[
R_m = \sum_{i=0}^{C_m} \rho_m(i) [(\lambda_n + \lambda_h + i \mu + i U_i)I + \Delta_j] \epsilon_2
\]
(48)

where \( \epsilon_2 \) is the unit column vector of order 2. Similarly, the event rate per macrocell, \( R_M \), can be described as
\[
R_M = \sum_{i=0}^{\infty} [(\lambda_H + \lambda_H + \lambda_H + J(s) \mu + U_L(s)I + X(s)(\lambda_n + \lambda_h)]PM(s).
\]
(49)

Because a macrocell is superimposed on \( M \) microcells, the total system event rate, \( R_{total} \), in the area covered by one macrocell is given by
\[
R_{total} = MR_m + R_M.
\]
(50)

IV. NUMERICAL RESULTS

A system of 49 microcells is considered where a macrocell is superimposed on seven microcells \( (M = 7) \). The number of channels allocated to each microcell and macrocell are 6 and 10,
respectively. In a macrocell, the number of channels which are reserved for handoff calls and new calls of high mobility users are 3 and 5, respectively. That is, $C_{MH} = 6$, $C_M = 10$, $C_{ML} = 3$, and $C_{M2} = 5$. In this situation, a macrocell will accommodate an overflow new call of a low mobility user only if the number of channels in use is less than 2 ($\leq 10 - (3 + 5)$). A macrocell will accommodate a new call of a high mobility user only if the number of channels in use is less than 7 ($\leq 10 - 3$). The macrocell will accommodate a handoff call if there are idle channels available in the macrocell. We developed a computer simulation model to validate the analytical results. To avoid the edge effect, it is assumed that microcells located at the edge of a macrocell block are folded so that every microcell has neighbors. In simple terms, the simulation environment has a spherical geometry.

The call arrival rate per user, $\lambda_0$, varies from 0.00015 to 0.00035 (calls/s). The call arrival rate per vehicle, $\lambda_v$, is assumed to be the same as $\lambda_0$. The number of noncommunicating low mobility users, $n_0$, is 100. The number of noncommunicating vehicles, $n_{v0}$, is 80. The average call duration of user, $\bar{t}_c$, is 100 s. The average channel holding time of low mobility users in a microcell, $\bar{T}_L$, is 40 s. The average channel holding time of high mobility users in a macrocell, $\bar{T}_H$, is 40 s. We can calculate the channel holding time of low mobility users in a macrocell, $\bar{T}_M$, as described in [1].

Fig. 5 shows the blocking probability of a new call. Because an increase of the departure rate in the macrocell due to the underflow scheme increases the average number of idle channels in the macrocell, the underflow scheme reduces the blocking probability for high mobility users. If the underflow scheme is applied, the first term for the blocking probability of low mobility users in (17) will increase and the second term will decrease. The major term is changed according to the input traffic load. When the input traffic is light, the second term is major and $P_{b,nL}$ is reduced. When traffic is heavy, the first term is major and $P_{b,nL}$ is increased. However, fluctuation of the blocking probability of low mobility users is negligible.

Fig. 6 shows the forced termination probability for low and high mobility users. Because the underflow scheme increases the average number of idle channels in a macrocell, the forced termination probability of high mobility users is reduced. When microcell-to-microcell call handoff fails, the macrocell can accommodate the call. The underflow scheme allows the overflow low mobility user to return to a microcell which is covered by the same macrocell. Thus, the underflow scheme reduces the forced termination probability for low mobility users.

Fig. 7 shows the total carried traffic in the area covered by one macrocell. The underflow scheme decreases both the blocking probability of high mobility users and the forced termination probability of all users. The decrease in the number of low mobility users due to an increase of the blocking probability for low mobility users reduces the carried traffic in the system. However, reduction of the forced termination probability increases the average communicating time, which is not interrupted by handoff failures. Reduction of the blocking and forced termination probabilities for high mobility users increases the carried traffic for high mobility users. In this environment, the underflow scheme slightly increases the total carried traffic in the area covered by a macrocell.
Fig. 8 shows the total system event rate in the area covered by one macrocell. The underflow scheme increases the total event rate of the overall system. This phenomenon occurs due to the fact that low mobility users currently served by a macrocell attempt to return to a microcell. It is the one disadvantage of the underflow scheme.

Table I shows a summary of the results illustrated in Figs. 5–8. This represents a performance increase, or a decrease in the performance of our proposed scheme. This table shows significant performance improvements for the new call blocking and forced termination probabilities of high mobility users, and for the forced termination probability of low mobility users.

In our simulation environment, the total number of channels covered by a macrocell region is \(52 = (\mathcal{M} \times C_m + C_M)\) for case 1. Case 2 can be defined as follows: \(C_m = 7, \mathcal{M} = 3, C_{M1} = 1\) and \(C_{M2} = 2\). That is, the total number of channels in case 2 is the same as in case 1. However, the number of channels in a microcell is increased by 1 and the number of channels in a macrocell is reduced. Figs. 9 and 10 show performance comparisons between cases 1 and 2. In case 2, because the number of channels in a macrocell is reduced, both the blocking probability and the forced termination probability of high mobility users are significantly increased to unacceptable levels. Although the number of channels in a microcell is increased, the forced termination probability of a low mobility user is significantly increased. This phenomenon results from a reduction of the number of channels in a macrocell. Reduction in the number of macrocell channels results in performance deterioration because macrocell channels are included in the common pool for low mobility users. In this case the increase in the number of microcell channels is only of minor importance. However, in view of the blocking probability of low mobility users, microcell channels are important terms. A decrease in the number of macrocell channels results in a minor effect and the blocking probability of low mobility users is decreased because the new calls of low mobility users have a low priority in a macrocell.

V. CONCLUSION

In a hierarchical cellular system lacking the underflow scheme, low mobility user calls that overflow to a macrocell cannot return to a microcell. Thus, both the blocking probability and the forced termination probability for high mobility users are large.

We have proposed a hierarchical cellular system with an underflow scheme and analyzed the performance of the system. Performance characteristics for users with different mobilities are evaluated. The proposed underflow scheme reduces the
blocking probability for high mobility users and the forced termination probability for high mobility users. It also reduces the forced termination probability for low mobility users. However, this scheme slightly increases the total carried traffic of the system.

Although not shown here due to complexities in the analysis, the underflow scheme has the property of helping to balance traffic patterns. Because macrocell channels are included in the channel pool, low mobility users in a hot spot area can be served by a macrocell. Disadvantages of this new system are that the system control load and the blocking probability for low mobility users are slightly increased due to the underflow scheme.

REFERENCES


Sung-Hong Wie received the B.S. and M.S. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Taejon, Korea, in 1995 and 1997, respectively. He is currently working toward the Ph.D. degree in electrical engineering at the same institute. His research interests include multiple access protocol, handoff, and resource management in wireless network.

Jae-Shin Jang received the B.S. degree in electrical engineering from the Dong-A University, Pusan, Korea in 1990. He received the M.S. and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Taejon, Korea, in 1992 and 1998, respectively. Since 1997, he has been working for Samsung Electronics Co., Ltd. in Korea. His research interests include multiple access protocol, handoff, and performance evaluation CDMA mobile communications network.

Byung-Cheol Shin received the B.S. degree in electrical engineering from the Seoul National University in 1975, and the M.S. and Ph.D. degrees, both in electrical and electronics engineering from the Korea Advanced Institute of Science and Technology (KAIST), in 1977 and 1984, respectively. He worked as a research staff member in Electronics and Telecommunications Research Institute, Korea, from 1977 to 1980. He visited SRI International, Menlo Park, CA, as an International Fellow in 1987. He also visited Rutgers University as a Visiting Scholar in 1995. From 1984 to 1998, he was with Korea Advanced Institute of Science and Technology. Since August 1998, he has been with Chungbuk National University as a Faculty Member in the School of Electrical and Electronics Engineering. His research interests include real-time/multimedia communication and performance analysis of ATM, high speed networks, and wireless networks.

Dong-Ho Cho received the B.S. degree in electrical engineering from the Seoul National University in 1979, and the M.S. and Ph.D. degrees, both in electrical and electronics engineering from the Korea Advanced Institute of Science and Technology (KAIST), in 1981 and 1985, respectively. From 1987 to 1997, he was Professor of Computer Engineering at the Kyunghee University. Since 1998, he has been an Associate Professor of Electrical Engineering at KAIST. His professional interests include wire/wireless communication network, protocol and services.