State University of New York at Stony Brook
College of Engineering and Applied Sciences

Technical Report No. 744

Performance Analysis of Session Oriented Data Communications for Mobile Computing in Cellular Systems

by

Yunsang Park and Stephen S. Rappaport

Department of Electrical Engineering
State University of New York
Stony Brook, New York 11794-2350

e-mail: park@sbee.sunysb.edu , rappaport@sunysb.edu

Date: August 4, 1997

Performance Analysis of Session Oriented Data Communications for Mobile Computing in Cellular Systems

Yunsang Park and Stephen S. Rappaport

Department of Electrical Engineering
State University of New York
Stony Brook, New York 11794-2350

park@sbee.sunysb.edu, rappaport@sunysb.edu

ABSTRACT

Future cellular communication systems must seamlessly support services for a wide range of user needs, including voice, data, video and multi-media. It is envisioned that mobile users may do much computer processing in an off-line mode but must occasionally connect to a network in order to exchange data and/or files. For this purpose a communication *session* is initiated. During the session the user has access to network resources, although this access may be shared with others. Owing to the mobile environment, the user's connection to the network during a session may be severed. Since the mobile user can act semi-autonomously, such disconnections can be transparent. That is the mobile user can continue to function in an off-line mode while the system will begin transparent automatic reconnection attempts to reestablish a link to the network. Only after a fixed (given) number of such attempts to reconnect have failed, is the *session* deemed to have failed. The issue is complicated by the hostile mobile radio environment and by user mobility. We consider session-oriented communications and develop a tractable analytical model for traffic performance based on multi-dimensional birth-death processes. The approach allows consideration of various platform types, such as pedestrians, automobiles, and buses, which may have very different mobility characteristics. Performance characteristics, such as: blocking, forced session termination, carried traffic, the average time per suspension, and the average number of suspensions per session are calculated.

The research reported in this paper was supported in part by the U.S. National Science Foundation under Grant no. NCR 94-15530 and in part by BMDO/IST under grant no. N00014-9511217 administered by the U.S. Office of Naval Research. General research support from Hughes Network Systems is gratefully acknowledged.

INTRODUCTION

With increasing demand for a variety of wireless services, ATM /B-ISDN offers a solution with flexible bandwidth allocation, high speed, and quality of service (QoS) selection [1]. The use of ATM in the *wireless* context (WATM) is discussed in [2] and [3]. The technology can support an array of services including multi-media, voice, and mobile computing. ATM network architecture was initially designed for bandwidth-rich wired network environments. In order to adapt the technology to *mobile* and *wireless* environments, in which bandwidth is limited, resource management strategies such as connection oriented approaches can be useful. Special problems such as multipath fading, co-channel interference, and mobility support require special attention beyond the issues that usually arise in fixed and/or wired user networks. In addition, mobile issues such as implementation of hand-offs and the *implications* of unsuccessful hand-off should be reconsidered in the new context. Rerouting of packets in the fixed network to accommodate ATM type communications with mobile users is considered in [3].

In circuit-switched cellular communication systems, when the radio link between mobile user and base is disconnected the call is forced to terminate and cleared from system. In contrast, for some services that are supportable by WATM, mobile users can operate semi-autonomously. So with appropriate design a disconnection from the network may be transparent to the user. That is such disconnections need not result in a failed session with discarded information and wasted resource usage.

The issue "how to maintain connectivity of a mobile user to the network" can be very important for implementation of new mobile wireless services. In [10] multiple links from a mobile user to network base stations are considered for the purpose accommodating users with different bandwidth needs. In this paper, we consider systems that attempt to maintain mobile user connectivity to the network by automatically and transparently attempting to reconnect disrupted links. Disconnection of a radio link does not cause the session to be cleared from the system until all reconnection efforts are fail. A session that is interrupted due to a premature termination caused by a failure of hand-off attempt is treated as a *suspended* session. Reconnection attempts for suspended sessions will be initiated while the mobile user application (operating at a higher protocol layer) continues.

The model for traffic performance is cast in the analytical framework that has been developing in recent years [4], [5], [6], [7], [8], [11], [12]. The approach, which uses multidimensional birth-death processes, allows numerical computation of relevant state probabilities and traffic performance measures [4], [5], [9]. Consideration of mixed platform types that have different mobility characteristics is included in the model -- and cut-off priority [4], [6] is used to reserve some channels for hand-offs. The global balance equations [4] for the system are formulated and solved for the state probabilities. These state probabilities are used to compute important traffic performance measures for the proposed system.

RECONNECTION ATTEMPTS

In our system configuration, when the physical connection between a mobile terminal and the network fails, the session is suspended and the mobile terminal will attempt to reconnect by successive reconnection requests made at random time intervals. A maximum number of reconnection attempts, N, are allowed for each suspended session. If a reconnection has not been secured after this maximum is reached the session is considered to have failed and the call is cleared from the system. The number of reconnection attempts for suspended sessions is counted and updated in the counter in the mobile terminal. It is assumed that there is a maximum number of suspended sessions that the system will allow in each cell. This maximum is denoted, H. It is possible that a platform with a suspended session on board leaves its current cell. Upon entering the target cell a hand-off attempt will be made to access resources in the target cell. This hand-off attempt counts towards the limit, N. If, in the target cell, there are no channels available to accommodate the call, and if there are also already H suspended sessions in the target cell, the call cannot be admitted in the target cell. So, even if a suspended session has not exhausted the allowable number of reconnection attempts, it will be forced into termination if it fails its hand-off attempt and the system already has H suspended sessions in the target cell.

When a mobile platform with an *active session* moves to a target cell that has no channels available, the session will be suspended and the mobile's reconnection attempt counter will be set to 1 if there are fewer than H suspended sessions already in the target cell. In the target cell, additional reconnection attempts may be made. The mobile's counter will be incremented for each unsuccessful attempt. It may happen that the supporting platform moves to yet another cell.

At that time a hand-off attempt will be made. This hand-off attempt will count as an attempted reconnection. Specifically, if the hand-off attempt succeeds in getting a channel, the session will be continued and the reconnection counter will be set to 0. If there are no channels in the target cell but there are fewer than H suspended sessions the *reconnection effort* will continue (that is, the counter will be incremented to 2 in this example case). If there are no channels available and there are H suspended sessions in the new target cell, the session will be terminated. For non-terminated sessions, the process will continue in this way as long as there has not been N consecutive failed reconnection (hand-off or retry) attempts. When this limit is reached the session will be forced to terminate. The mobile terminal may confirm this termination to the network via the control channel (A timeout in the network can also be used as a backstop). A detailed description of mobile terminal's reconnection counter is given in the APPENDIX.

In the following we will let g be an index that defines the platform type and mobility. Consider a suspended session that has already failed k-1 reconnection attempts. The next reconnection attempt is called the "k-reconnection attempt" where $1 \le k \le N$. It is important to emphasize that there are two driving processes that generate reconnection attempts. One is the *retry process*, which consists of successive statistically independent realizations of a random variable, $T_r(k, g)$, to generate epochs for *retry* attempt times for a suspended session. The other is *the hand-off departure process* -- because hand-off attempts always try to establish a link and therefore count as *reconnection attempts*. The random variable gives the time from the previous reconnection event (either hand-off or retry) to the next *anticipated* retry attempt. The random variable, $T_r(k, g)$, can in general depend on k. Thus, the *minimum rate* of reconnection attempts depends on the number of attempts that have already been made. Of course, if the supporting platform leaves its current cell before the anticipated retry epoch, a hand-off attempt (to establish a link) will be made at that time and the value of k will be adjusted. If the session is in a suspended state after this attempt, a new random variable (for a retry epoch) will be generated. The random variable, $T_r(k, g)$, generated after the k-1 reconnection attempt, which represents the maximum time to the next anticipated retry attempt is called the "*k-trial time*". A suspended session that has not reestablished a link after k-1 reconnection trials and is waiting for the next (k^{th}) reconnection attempt, is called a "*k-suspended session*" .

MODEL DESCRIPTION

We consider a large geographical area tessellated by cells that are defined by proximity to specific network gateways (base stations). Large numbers of mobile platforms of several types traverse the region. The platform types differ primarily in the mobility characteristics and each platform can support at most one connection at any given time. The maximum number of simultaneous connections that each base station can support is C.

A hand-off is needed when a platform with either an active or suspended session moves to another cell. A hand-off attempt will only gain access to a connection in the target cell if there are less than C connections in progress in that cell. An active session that fails to gain access to a connection will lose its wireless link. In conventional cellular systems, this call (session) will be cleared from system. In our system configuration, however, if an active session loses its wireless link (i.e., is disconnected), the call may be suspended and reconnection attempts will be initiated.

As was done in previous work [4], [5], we use the concept of *dwell time* to characterize platform mobility. This is a random variable defined as the duration of time that a two-way link of satisfactory quality can be maintained between a platform and its current base, for whatever reason. The dwell time of platform in a cell depends on many factors including; mobility, signal power, propagation conditions, fading, etc. Although generalizations are possible, [4], [5], [11], [12], here we take the probability density function (p.d.f.) of dwell time to be a negative exponential distribution (with a parameter depending on mobility of the platform type). Similarly, the unencumbered session duration and the k-trial time were taken to be n.e.d. random variables (with parameters depending on intended session duration and the value of k, respectively).

EXAMPLE PROBLEM STATEMENT

The system has G types of platforms, indexed by $\{g=1,2,3.....G\}$. The call origination rate from a non-communicating g-type platform is denoted $\Lambda(g)$. We define $\alpha(g) = \Lambda(g)/\Lambda(1)$. The number of non-communicating g-type platforms in any cell is denoted $\nu(g,0)$. Therefore, the total call origination rate for g-type platforms in a cell is $\Lambda_n(g) = \Lambda(g) \cdot \nu(g,0)$. It is assumed that the number of non-communicating platforms is much larger than the maximum available connections in a cell so that the call generation rate does not depend on the number of sessions in progress (this is called an infinite population model).

Generally the bandwidth and other resources needed for connection of a call may depend on call type. A model that considers resource use based on call (connection) type is developed in [8]. However, in this paper, we wish to focus on the issue of maintaining connectivity, so for convenience, it is assumed that each active connection requires the same resources. Each cell or gateway can support a maximum of C connections. There are no quotas for specific mobility platform type. Cut-off priority for hand-offs and reconnection attempts is included in the present discussion. Thus, C_h connections in each cell are reserved for hand-off attempts (from platforms entering a cell) and for reconnection attempts from suspended sessions in the cell. A connection will be established for a new call only if there are fewer than $C - C_h$ active sessions in the cell. Hand-off attempts will fail to get a connection if there are C active sessions in the cell. An active session (attempting a hand-off to a target cell) will be *suspended* if it fails to get a connection but there are less than H suspended sessions in the target cell. It will be *terminated* if there are C active sessions and H suspended sessions in the target cell.

The platform is considered to "leave" the cell at the expiration of its current (random) dwell time. A *communicating* platform that leaves a cell generates a hand-off arrival to some other cell. Here the dwell time in a cell for g-type platform is taken as an n.e.d. random variable, $T_D(g)$, having a mean $\overline{T}_D(g) = 1/\mu_D(g)$. The k-trial time of a suspended session on g-type platform is a n.e.d. random variable, $T_r(g)$, having a mean $\overline{T}_r(k,g) = 1/\mu_r(k,g)$, where $1 \le k \le N$, and $\mu_r(k,g)$ ($k = 1,\ldots N; g = 1,..G$) is a the parameter that determines the reconnection attempt rate for a *k-suspended session* on a *g-type platform.*

STATE DESCRIPTION

Consider a single cell. We define the ***cell state*** by a sequence of non-negative integers. When a maximum of N reconnection attempts are permitted for a suspended session, the state of the cell can be written as G n-tuples as follows

$$
\begin{array}{cccccc}
u_1 & v_{1,1} & v_{1,2} & v_{1,3} & \cdots\cdots & v_{1,N} \\
u_2 & v_{2,1} & v_{2,2} & v_{2,3} & \cdots\cdots & v_{2,N} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
u_G & v_{G,1} & v_{G,2} & v_{G,3} & \cdots\cdots & v_{G,N}
\end{array}
$$

where $u_g\,\{g=1,2,3,\ldots,G\}$ is the number of active sessions on g-type platforms and $v_{g,k}\,\{g=1,2,\ldots,G;k=1,2,\ldots,N\}$ is the number of k-suspended sessions on g-type platforms. For convenience we order the states using an index $s = 0,1,\ldots,S_{\max}$. Thereafter, u_g and $v_{g,k}$ can be written explicitly dependent on the state. That is $u_g = u(s,g)$ and $v_{g,k} = v(s,g,k)$.

When the cell is in state s, the following characteristics can be determined. The number of active sessions is

$$u(s) = \sum_{g=1}^{G} u(s,g). \tag{1}$$

The number of suspended sessions on g-type platforms is

$$v(s,g) = \sum_{k=1}^{N} v(s,g,k). \tag{2}$$

The total number of suspended sessions in cell, regardless of platform type, is

$$v(s) = \sum_{g=1}^{G} v(s,g). \tag{3}$$

The number of sessions either active or suspended on g-type platforms in the cell is

$$J(s,g) = u(s,g) + v(s,g). \tag{4}$$

The total number of sessions in progress that are either active or suspended is

$$J(s) = u(s) + v(s). \tag{5}$$

There are constraints on permissible cell states. These include the total number of active sessions in a cell must be less than or equal to maximum supportable connections, $u(s) \le C$; and the total number of suspended sessions in a cell must be less than or equal to the maximum number of suspended sessions allowed in a cell, $v(s) \le H$.

DRIVING PROCESS

There are five major driving processes for this problem. We use Markovian assumptions for driving processes to allow solution within the multidimensional birth-death process framework [4], [5], [9]. Each process is listed below

- $\{n\}$: generation of new calls
- $\{c\}$: completion of calls
- $\{h\}$: hand-off arrival of calls

  $\{h_0\}$: active session hand-off arrivals

  $\{h_1\}$: 1-suspended session hand-off arrivals

  $\vdots$ $\vdots$

  $\{h_N\}$: N-suspended session hand-off arrivals
- $\{d\}$: hand-off departure of calls

  $\{d_0\}$: active session hand-off departures

  $\{d_1\}$: 1-suspended session hand-off departures

  $\vdots$ $\vdots$

  $\{d_N\}$: N-suspended session hand-off departures
- $\{r\}$: retry attempt

The dimension of the new call generation process, $\{n\}$, is G, since there are G different types of mobility platforms. Similarly, the dimensions of the hand-off arrival and departure processes, $\{h\}$ and $\{d\}$, are $G\times(N+1)$, since N-times of reconnection attempts are permitted for a suspended session besides of existence of active sessions. The dimension of the retry attempt process $\{r\}$ is $G\times N$, since N types of suspended sessions can arise from G different types of platforms.

Generation Of New Calls

A transition into state $\vec{s}$, due to a new call arrival on a g-type platform when the cell is in state x_n, will cause the state variable $u(x_n, g)$ to be increased by 1. Because of cut-off priority, C_h connections are held for arrivals of hand-off attempts (of active or suspended sessions) and for retry attempts of suspended sessions. Thus a permissible state x_n is a predecessor state of s for a new call arrives on g-type platform, if $u(x_n) < C - C_h$ and the state variables are related by

$$\begin{aligned} &u(x_n, g) = u(s, g) - 1 \\ &v(x_n, g, j) = v(s, g, j). \end{aligned} \tag{6}$$

Let $\Lambda_n(g)$ denote the average arrival rate per a cell of new calls from a g-type platform. Then, the corresponding transition flow is given by

$$\gamma_n(s, x_n) = \Lambda_n(g). \tag{7}$$

Completion Of Calls

A transition into state s, due to successful completion of a session on a g-type platform when the cell is in state x_c, will cause the state variable $u(x_c, g)$ to be decreased by 1. Thus a permissible state x_c is a predecessor state of s for call completion on a g-type platform if the state are related by

$$\begin{aligned} &u(x_c, g) = u(s, g) + 1 \\ &v(x_c, g, j) = v(s, g, j). \end{aligned} \tag{8}$$

The unencumbered session duration on a g-type platform is a n.e.d. random variable, $T(g)$, having a mean $\overline{T}(g) = 1/\mu(g)$. Then the transition flow into state s from x_c due to a session completion is given by

$$\gamma_c(s, x_c) = \mu(g) \cdot v(x_c, g). \tag{9}$$

Hand-Off Arrival Of Calls

ACTIVE SESSION HAND-OFF ARRIVALS

When a cell has less than C active sessions, arriving hand-off calls can obtain a connection. Thus a permissible state x_{h0} is a predecessor state of s for an active session hand-off arrival on a g-type platform when $u(x_{h0}) < C$ and the state variables are related by

$$\begin{aligned} &u(x_{h0}, g) = u(s, g) - 1 \\ &v(x_{h0}, g, j) = v(s, g, j). \end{aligned} \tag{10}$$

When an active session needs a hand-off, it will become a 1-suspended session if the target cell has C active sessions but the total number of suspended sessions in that cell is less than H. A transition into state s, due to an hand-off arrival of active session on a g-type platform when the cell is in state x_{h0} (in which the cell has C active sessions but the total number of suspended sessions in that cell is less than H), will cause the state variable $v(x_{h0}, g, 1)$ to increase by 1. Thus a permissible state x_{h0} is a predecessor state of s for hand-off arrival of active session on a g-type platforms when $u(x_{h0}) = C$ and $v(x_{h0}) < H$ if the state variables are related by

$$\begin{aligned} u(x_{h0}, g) &= u(s, g) \\ v(x_{h0}, g, 1) &= v(s, g, 1) - 1 \\ v(x_{h0}, g, j) &= v(s, g, j), \; j \neq 1. \end{aligned} \tag{11}$$

We let Λ_h be the average rate at which hand-off arrivals of active session impinge on the cell and F_g denote the fraction of hand-off arrival of active session that are from g-type platforms. Initially we guess Λ_h and F_g but these values are determined by the dynamics of process [5], [7]. Using the approach put forth in [5] we will subsequently determine these values using an iterative approach. The corresponding transition flow for active session hand-off attempts is given by

$$\gamma_{h0}(s, x_{h0}) = \Lambda_h \cdot F_g. \tag{12}$$

HAND-OFF ARRIVALS OF SUSPENDED SESSIONS

If the target cell has less than maximum number of simultaneously supportable connections, C, when a k-suspended session hand-off attempt arrives, ($1 \leq k \leq N$), the session will be accommodated and will become an active session in the target cell. Thus a permissible state x_{hk} is a predecessor state of s for hand-off arrivals of k-suspended sessions on g-type platform when $u(x_{hk}) < C$, if the state variables are related by

$$\begin{aligned} u(x_{hk}, g) &= u(s, g) - 1 \\ v(x_{hk}, g, j) &= v(s, g, j) \end{aligned} \tag{13}$$

where $1 \leq k \leq N$.

If the target cell has no available channel resources for connection of arriving hand-off call but the total number of suspended sessions in that cell is less than maximum supportable suspended sessions, H, a hand-off attempt will fail and a k-suspended session becomes a $(k+1)$-suspended session in the target cell unless the mobile's counter indicate $k=N$. Thus a permissible state x_{hk} is a predecessor state of s for hand-off arrival of k-suspended session on a g-type platform when $u(x_{hk}) = C$ and $v(x_{hk}) < H$, if the state variable are related by

$$\begin{aligned} &u(x_{hk}, g) = u(s, g) \\ &v(x_{hk}, g, k+1) = v(s, g, k+1) - 1 \\ &v(x_{hk,}, g, j) = v(s, g, j),\ j \neq (k+1) \end{aligned} \tag{14}$$

where $1 \leq k < N$.

It should be reminded that when the counter of the terminal indicate $N+1$, then the terminal has been attempted maximum allowable reconnection attempts. This call will be forced into termination. So, a failure of hand-off attempt when the counter of the terminal is N causes a call to be forced into termination. This type of termination (a call is terminated with $k=N$) is called *maximum termination* and described with detail in section of PERFORMANCE MEASURES.

Let $\Lambda_r(k)$ be the average rate at which hand-off arrivals of k-suspended session on a g-type platform impinge on the cell, $F_{rg}(k)$ denote the fraction of hand-off arrival of k-suspended sessions that are from g-type platform. Initially we guess $\Lambda_r(k)$ and $F_{rg}(k)$ but using an iterative method, we will subsequently determine the values required by the dynamics of the process [5]. Then the corresponding transition flow is given by

$$\gamma_{hk}(s, x_k) = \Lambda_r(k) \cdot F_{rg}(k) \tag{15}$$

where $1 \leq k \leq N$.

Departure Of Hand-Off Calls

HAND-OFF DEPARTURE OF ACTIVE SESSIONS

A transition into state s, due to a hand-off departure of active session on a g-type platform when the cell is in state x_{d0}, will cause the state variables $u(x_{d0}, g)$ to be decreased by 1. Thus a permissible state x_{d0} is a predecessor state of s for a hand-off departure of active session on a g-type platforms, if the state variables are related by

$$\begin{aligned} u(x_{d0}, g) &= u(s, g) + 1 \\ v(x_{d0}, j) &= v(s, g, j). \end{aligned} \tag{16}$$

The corresponding transition flow is given by

$$\gamma_{d0}(s, x_{d0}) = \mu_D(g) \cdot u(x_{d0}, g). \tag{17}$$

HAND-OFF DEPARTURES OF K-SUSPENDED SESSIONS

A transition into state s, due to a hand-off departure of k-suspended session on a g-type platform when the cell is state x_{dk}, will cause the state variable $v(x_{dk}, g, k)$ to be decreased by 1. Thus a permissible state x_{dk} is a predecessor state of s for *a hand-off departures of k-suspended session on g-type platforms*, if the state variables are related by

$$\begin{aligned} u(x_{dk}, g) &= u(s, g) \\ v(x_{dk}, g, k) &= v(s, g, k) + 1 \\ v(x_{dk}, g, j) &= v(s, g, j), \; j \neq k \end{aligned} \tag{18}$$

where $1 \leq k \leq N$.

The corresponding transition flow is given by

$$\gamma_{dk}(s, x_{dk}) = \mu_D(g) \cdot v(x_{dk}, g, k) \tag{19}$$

where $1 \leq k \leq N$.

Retry Attempt

If the cell has less than the maximum number of active sessions in progress when a terminal attempts a reconnection trial (either hand-off or retry) for a k-suspended session, the reconnection attempt will succeed and the session will become active. Thus, a transition into state s, due to a retry attempt for k-suspended session on a g-type platform, when $u(x_r) < C$ and the cell is state x_r, will cause the state variable $u(x_r,g)$ increased by 1 and $v(x_r,g,k)$ to be decreased by 1. So, a permissible state x_r is a predecessor state of s for a retry attempt of k-suspended session on a g-type platform, if $u(x_r) < C$ and the state variables are related by

$$\begin{aligned} &u(x_r, g) = u(s, g) - 1 \\ &v(x_r, g, k) = v(s, g, k) + 1 \\ &v(x_r, g, j) = v(s, g, j), \, j \neq k \end{aligned} \tag{20}$$

where $1 \leq k \leq N$.

If the system is supporting C active sessions when the terminal makes a retry attempt for a k-suspended session, the retry attempt will fail. When the counter of terminal indicate less number than N, the terminal will wait for next reconnection attempt. A transition into state s, due to a failure of a retry attempt for k-suspended session, if mobile's counter indicate less than maximum allowable reconnection attempts ($k< N$), on a g-type platform when a cell is state x_r, will cause the state variable $v(x_r,g,k)$ decreased by 1 and $v(x_r,g,k+1)$ increased by 1. Thus a permissible state x_r is a predecessor state of s for the failure of k-repeated trial ($k< N$) on a g-type platform when $u(x_r) = C$ and $v(x_r) < H$, if the state variables are related by

$$\begin{aligned} &u(x_r, g) = u(s, g) \\ &v(x_r, g, k) = v(s, g, k) + 1 \\ &v(x_r, g, k+1) = v(s, g, k+1) - 1 \\ &v(x_r, g, j) = v(s, g, j), \, j \neq k, \, j \neq k+1 \end{aligned} \tag{21}$$

where $1 \leq k < N$.

When the counter of the terminal indicate N+1, then the terminal has been attempted maximum allowable reconnection attempts. This call will be forced into termination. So, a failure of reconnection attempt when the counter of the terminal is N caused a call to be forced into termination. This type of termination (a call is terminated with k=N) is called *maximum termination* and described with detail in section of PERFORMANCE MEASURES. Thus a permissible state x_r is a predecessor state of s for the failure of N-reconnection attempt on a g-type platform when $u(x_r) = C$ and $v(x_r) < H$, if the state variables are related by

$$u(x_r, g) = u(s, g)$$
$$v(x_r, g, N) = v(s, g, N) + 1 \tag{22}$$
$$v(x_r, g, j) = v(s, g, j),\ j \neq N.$$

The corresponding transition flow is given by

$$\gamma_r(s, x_r, k) = \mu_r(k, g) \cdot v(x_r, g, k). \tag{23}$$

FLOW BALANCE EQUATIONS

From the above equations, the total transition flow into state s from any permissible predecessor state x can be written as

$$q(s,x) = \gamma_n(s,x) + \gamma_c(s,x) + \gamma_h(s,x) + \gamma_r(s,x) + \gamma_d(s,x), \tag{24}$$

where

$$\gamma_h(s,x) = \gamma_{h0}(s,x) + \gamma_{h1}(s,x) + \cdots + \gamma_{hN}(s,x), \tag{25}$$

$$\gamma_d(s,x) = \gamma_{d0}(s,x) + \gamma_{d1}(s,x) + \cdots + \gamma_{dN}(s,x), \tag{26}$$

and $\gamma_r(s,x) = \gamma_r(s,x,1) + \gamma_r(s,x,2) + \cdots + \gamma_r(s,x,N)$, (27)

in which $s \neq x$, and flow into a state has been taken as a positive quantity.

The total flow out of state s is denoted $q(s,s)$ and is given by

$$q(s,s) = -\sum_{\substack{k=0 \\ k \neq s}}^{S_{max}} q(k,s). \tag{28}$$

The statistical equilibrium state probabilities can be found by solving flow balance equations that are a set of $S_{max}+1$ simultaneous equations [5].

$$\sum_{j=0}^{S_{max}} q(i,j) \cdot p(j) = 0, \qquad i = 0, 1, \cdots\cdots, S_{max} - 1 \tag{29}$$

$$\sum_{j=0}^{S_{max}} p(j) = 1, \tag{30}$$

in which, for $i \neq j$, $q(i,j)$ is the net transition flow into state i from state j, and $q(i,i)$ is the total transition flow out of state i.

HAND-OFF ARRIVAL PARAMETERS

The average hand-off arrival rate of active session, Λ_h, the average hand-off arrival rate of k-suspended session, $\Lambda_r(k)$, where $1 \le k \le N$, the fraction of hand-off arrivals of active session that are g-type platform, F_g, and the fraction of hand-off arrivals of k-suspended sessions that are g-type platform, $F_{rg}(k)$, where $1 \le k \le N$, can be determined from the dynamics of the process itself. An iterative method can be used [5]. The average hand-off departure rate of active sessions on g-type platforms can be expressed as

$$\Delta_h(g) = \sum_{s=0}^{S_{\max}} \mu_D(g) \cdot u(s,g) \cdot p(s). \tag{31}$$

Thereafter, the overall average hand-off departure rate of active sessions can be written as

$$\Delta_h = \sum_{g=1}^{G} \Delta_h(g). \tag{32}$$

The average hand-off departure rates of k-suspended sessions on g-type platform can be expressed as

$$\Delta_r(g,k) = \sum_{s=0}^{S_{\max}} \mu_r(g) \cdot v(s,g,k) \cdot p(s). \tag{33}$$

Also, the overall average hand-off departure rates of k-suspended sessions can be written as

$$\Delta_r(k) = \sum_{g=1}^{G} \Delta_r(g,k). \tag{34}$$

From these equations, we find that the fraction of hand-off departures of active sessions that are g-type platforms is

$$F'_g = \Delta_h(g) / \Delta_h \tag{35}$$

and, the fraction of hand-off departures of k-suspended sessions on g-type mobility platform is

$$F'_{rg}(k) = \Delta_r(g,k) / \Delta_r(k). \tag{36}$$

Since the maximum allowable reconnection attempts is N, there are N *average hand-off departure rates* and fractions, each corresponding to a value of k. Any hand-off departure of an active session of a g-type platform a cell corresponding to a hand-off arrival of active session of a g-type platform to another cell. Also, a *hand-off departure of a k-suspended session of a g-type platform* from a cell, corresponds a *hand-off arrival of a k-suspended session of a g-type* *platform* to another cell. Therefore, for a homogeneous system in statistical equilibrium, the hand-off arrival and departure rates per cell must be equal and the component hand-off arrival rates of k-suspended sessions and hand-off departure rates of k-suspended sessions, where $1 \leq k \leq N$, must also equal one another. That is we must have

$$\begin{aligned} F_g &= F'_g \\ F_{rg}(k) &= F'_{rg}(k) \\ \Lambda_h &= \Delta_h \\ \Lambda_r(k) &= \Delta_r(k) \end{aligned} \tag{37}$$

where $1 \leq k \leq N$.

PERFORMANCE MEASURES

When the statistical equilibrium state probabilities and transition flow are found, the require performance measures can be calculated.

Carried Traffic

An important performance measure from a system point of view is the carried traffic. For given resources, larger carried traffic implies more efficient use and more revenue for the service provider [4]. The carried traffic for g-type platform, $A_c(g)$, is

$$A_c(g) = \sum_{s=0}^{S_{max}} u(s,g) \cdot p(s). \tag{38}$$

The total carried traffic, A_c, is

$$A_c = \sum_{g=1}^{G} A_c(g). \tag{39}$$

Average Number of K-Suspended Sessions

The average number of k-suspended sessions on g-type platforms, $A_w(k,g)$, is

$$A_w(k,g) = \sum_{s=0}^{S_{max}} v(s,g,k) \cdot p(s). \tag{40}$$

The average number of k-suspended sessions regardless of platform type, $A_w(k)$, is

$$A_w(k) = \sum_{g=1}^{G} A_w(k, g). \tag{41}$$

Then, the average number of suspended sessions, A_w, is

$$A_w = \sum_{k=1}^{N} A_w(k). \tag{42}$$

Blocking Probability

The blocking probability, P_B, is the average fraction of new call arrivals that are failed acquire a connection [4], [5]. Blocking events occurs when the cell is in one of state of following *disjoint* subsets of states

$$L_B = \{s: u(s) \geq C - C_h\}. \tag{43}$$

And, blocking probability is expressed as

$$P_B = \sum_{s \in L_B} p(s). \tag{44}$$

Hand-Off Failure Probability

The hand-off failure probability, P_H, is the average fraction of hand-off attempts that are denied in the target cell because the system in target cell supports maximum supportable connections and maximum supportable suspended sessions. A session, either active or suspended, that are denied in the target cell due to the lack of system capacity will be forced into termination and cleared from system database. We define following disjoint set of states, in which hand-off attempts will fail

$$L_H = \{s: u(s) = C, v(s) = H\}. \tag{45}$$

Then, the hand-off failure probability is expressed as

$$P_H = \sum_{s \in L_H} p(s). \tag{46}$$

Forced Termination Probability

The forced termination probability, $P_{FT}(g)$, is defined as the probability that a g-type of call that is not blocked is forced into termination due to hand-off failure during its lifetime. The terminated call will be cleared from system. There are two possible scenarios that a call is forced into termination due to a failure of hand-off attempt during its lifetime. Firstly, a call, whether active or suspended, attempts its hand-off to a cell in which the system already supports maximum number of connections and maximum number of suspended sessions. In this case, a call will be forced terminated even if it haven't finished its maximum allowable reconnection attempts. A premature termination of a session, either an active or suspended, that have not finished its maximum allowable reconnection attempts is called a ***non-maximum termination***. Secondly, a suspended session is forced terminated due to the failures of maximum allowable number of reconnection attempts. It should be recalled that a suspended session will attempt its hand-off when it move to another cell even though the radio link between mobile unit and base was disconnected. The hand-off attempt generated by a suspended session attempts will be counted as one reconnection attempt in the sense that the reconnection counter will be incremented. Therefore, the last attempt before termination can be either a retry attempt or a hand-off attempt. The termination of an N-suspended session due to the maximum number of allowable reconnection attempts being met is called a ***maximum termination***.

For the purpose of calculating forced termination probability, we defined following disjoint set of states.

$$L_D = \{s: u(s) = C, v(s) \neq H\} \tag{47}$$

$$L_A = \{s: u(s) \neq C\}. \tag{48}$$

When the cell is one of states in L_D, the cell already supports the maximum number of active sessions but the number of suspended sessions in the cell is less than the maximum supportable suspended sessions. If a call, either active or suspended call, arrives due to a hand-off attempt when a cell is one of these states, it will stay as a suspended session unless the reconnection counter indicates maximum allowable reconnection attempts, $(N+1)$. When the counter of call indicate $(N+1)$, a call will be forced into termination.

When the cell is one of the states L_A, the system supports fewer connections than the limit. Therefore, an arriving hand-off, either active or suspended call, and retry attempt will succeed in gaining access to network resources. The corresponding system probability is expressed as

$$P_D = \sum_{s \in L_D} p(s) \tag{49}$$

$$P_A = \sum_{s \in L_A} p(s). \tag{50}$$

It should be noted that a non-maximum termination can be caused only by the failure of a hand-off attempt. No retry attempt, can result in a non-maximum termination since the suspended session already occupies one of the H spaces allowed for suspended sessions in each cell. Succinctly, a retry attempt cannot cause a forced termination unless it is the last allowable attempt for that session and it fails to reconnect.

Figure 1 is a flow graph depicting events in the lifetime of a call. We can see that some suspended or active sessions will be forced to terminate because of the failure of their hand-off attempts or retry attempts. The probability that an active session on a g-type platform attempts a hand-off as an upcoming event before its session completion can be written as $\frac{\mu_D(g)}{\mu(g)+\mu_D(g)}$. And, the probability that an active session on a g-type platform successfully completes its session before a trial of hand-off attempt can be written as $\frac{\mu\ (g)}{\mu(g)+\mu_D(g)}$.

The probability that an active session becomes a 1-suspended session due the lack of system capacity in target cell is P_D. For suspended sessions, there are two possible upcoming events, hand-off or retry attempt. When the dwell time of a suspended session is less than the next trial time, the session will attempt hand-off. Otherwise, the session will attempt retry. The probability that a g-type k-suspended session attempts hand-off as a next upcoming event can be written as $\frac{\mu_D(g)}{\mu_r(k,g)+\mu_D(g)}$. Also, the probability that a g-type k-suspended session attempts retry as the next upcoming event can be written as $\frac{\mu_r(k,g)}{\mu_r(k,g)+\mu_D(g)}$.

A hand-off event of a suspended session, it can result in resumption of an active session, a forced termination, or continued suspension (with an incremented reconnection attempt counter). These events occur with respective P_A, P_H, and P_D.

A retry attempt generated by a suspended session, can result in: 1) the suspended session becoming active (with probability of P_A); 2) the continued suspension with incremented reconnection counter if the maximum number of allowed reconnection attempts has not been reached (with probability $P_H + P_D$).

There are two possible scenarios that a k-suspended session to become a (k+1)-suspended session. One is to attempt hand-off and be continued suspension. The other is to try retry attempts and be continued suspension. Therefore, the probability, $\theta_k(g)$, that a k-suspended session on a g-type platform becomes a (k+1)-suspended session through either a retry or a hand-off attempt can be written as

$$\theta_k(g) = \frac{\mu_r(k,g)\cdot(P_H + P_D)}{\mu_r(k,g)+\mu_D(g)} + \frac{\mu_D(g)\cdot P_D}{\mu_r(k,g)+\mu_D(g)} \tag{51}$$

$$= \frac{\mu_r(k,g)\cdot(P_H + P_D) + \mu_D(g)\cdot P_D}{\mu_r(k,g)+\mu_D(g)} \tag{52}$$

where $1 \le k < N$.

Even though a call has been experienced suspension ,the counter will be reset to zero if a suspended call to be activated again. A call may have those suspension experience during its life time before it successfully finish its session. Those suspension and reconnection process is initiated after an active call is fail its hand-off attempt. Therefore, the probability that an active call have a hand-off attempt as an upcoming event and acquire connection through either successful hand-off or retry attempt should be calculated for measuring the system performance. The probability, $\eta(g)$, that a g-type active call, having a hand-off attempt, will become an active session through some reconnection or hand-off processes is the sum of the probabilities that it succeeds an upcoming hand-off attempt, becomes a suspended session but succeeds its 1-retry attempt or next hand-off attempt, becomes a suspended session and fails its 1-retry attempt but succeeds its 2-retry attempt or next hand-off attempt, so on. Therefore, it can be written as

$$\eta(g) = \frac{\mu_D(g)\cdot P_A}{\mu(g)+\mu_D(g)} + \frac{\mu_D(g)\cdot P_D\cdot P_A}{\mu(g)+\mu_D(g)} + \frac{\mu_D(g)\cdot P_D\cdot P_A\cdot\theta_1(g)}{\mu(g)+\mu_D(g)}$$

$$+ \frac{\mu_D(g)\cdot P_D\cdot P_A\cdot\theta_1(g)\cdot\theta_2(g)}{\mu(g)+\mu_D(g)} + \dots\dots\dots + \frac{\mu_D(g)\cdot P_D\cdot P_A\cdot\theta_1(g)\cdot\theta_2(g)\cdots\theta_{N-1}(g)}{\mu(g)+\mu_D(g)} \tag{53}$$

$$= \frac{\mu_D(g) \cdot P_A}{\mu(g) + \mu_D(g)} \cdot \left(1 + P_D\left(1 + \sum_{k=1}^{N-1} \prod_{m=1}^{k} \theta_m(g)\right)\right) \quad (54)$$

where $1 \le k \le N$.

NON-MAXIMUM TERMINATION PROBABILITY

In previous, we define a premature termination of a session, either an active or suspended, that have not finished its maximum allowable reconnection as a *non-maximum termination.* Therefore, if the counter of terminal does ***not*** indicate (N+1) when a session is forced into termination, it is called non-maximum termination. The non-maximum termination probability, $P_{NT}(g)$, is defined as the probability that a g-type call, either active or suspended, that is forced into termination during its lifetime due to a failure of a hand-off attempt even if it have not finished its maximum allowable reconnection attempts. This probability will be the sum of probabilities that a session is forced into termination while in process, while in 1-suspended session,...., up to while in N-1 suspended session. The probability, $\psi_0(g)$, that a g-type call is forced terminated while its session is in progress due to a failure of hand-off during its lifetime can be written as

$$\psi_0(g) = \frac{\mu_D(g) \cdot P_H}{\mu_D(g) + \mu(g)} + \frac{\mu_D(g) \cdot P_H \cdot \eta(g)}{\mu_D(g) + \mu(g)} + \frac{\mu_D(g) \cdot P_H \cdot \eta(g)^2}{\mu_D(g) + \mu(g)} + \cdots\cdots \quad (55)$$

$$= \frac{\mu_D(g) \cdot P_H \cdot \sum_{i=0}^{\infty} \eta(g)^i}{\mu_D(g) + \mu(g)}. \quad (56)$$

With the same fashion, the probability that a g-type call is forced terminated while it is a k-suspended session, where $0 < k < N$, due to a failure of hand-off during its lifetime, $\psi_k(g)$, can be written as

$$\psi_k(g) = \frac{\bar{\mu}_D(g) \cdot P_D \cdot \sum_{i=0}^{\infty} \eta(g)^i}{\mu_D(g) + \mu(g)} \cdot \left(\frac{\mu_D(g) \cdot P_H}{\mu_D(g) + \mu_r(k, g)}\right) \cdot \prod_{m=1}^{k-1} \theta_m(g). \quad (57)$$

Then, the overall non-maximum termination probability, $P_{NT}(g)$, can be written as

$$P_{NT}(g) = \psi_0(g) + \psi_1(g) + \cdots \psi_{N-1}(g) \tag{58}$$

$$= \frac{\mu_D(g) \cdot P_H \cdot \sum_{i=0}^{\infty} \eta(g)^i}{\mu_D(g) + \mu(g)} \cdot \left(1 + \frac{\mu_D(g) \cdot P_D}{\mu_D(g) + \mu_r(k,g)} \sum_{k=1}^{N-1} \prod_{m=1}^{k-1} \theta_m(g)\right) \tag{59}$$

$$= \frac{\mu_D(g) \cdot P_H}{\mu_D(g) + \mu(g)} \cdot \frac{1}{1 - \eta(g)} \cdot \left(1 + \frac{\mu_D(g) \cdot P_D}{\mu_D(g) + \mu_r(k,g)} \sum_{k=1}^{N-1} \prod_{m=1}^{k-1} \theta_m(g)\right). \tag{60}$$

PROBABILITY OF MAXIMUM TERMINATION

The probability of maximum termination, $P_{MT}(g)$, is defined as the probability that a g-type call that is forced to terminate during its lifetime because the maximum allowable number of reconnection attempts have been reached. The last attempt of maximum termination can be either a hand-off or retry. The probability of maximum termination, $P_{MT}(g)$, can be written as

$$P_{MT}(g) = \frac{\mu_D(g) \cdot (P_H + P_D) \cdot P_D \cdot \prod_{i=1}^{N-1} \theta_i(g)}{(\mu(g) + \mu_D(g)) \cdot (1 - \eta(g))}. \tag{61}$$

The forced termination probability, $P_{FT}(g)$, is defined as the probability that a g-type of call that is not blocked is forced into termination due to hand-off failure or failure of reconnection attempts during its lifetime. It can be written as

$$P_{FT}(g) = P_{NT}(g) + P_{MT}(g). \tag{62}$$

Average Time Per Suspension

A call may experience a suspension or some suspensions during its lifetime. When a call becomes a suspended session, the terminal on that call will start its reconnection process. This reconnection effort can succeed or fail. The average time per suspension on a g-type platform, $W(g)$, is the expected time that reconnection process will carry on. Therefore, it is the average time frame from the point that an active session becomes a suspended session to the point that a suspended session becomes an active session or be forced into termination.

To calculate the average time per suspension, determine the average rate of suspension and the number of calls in suspension from the state probabilities. Little's law will then be applied to find the average amount of time in suspension [7].

The average rate of call suspension for a g-type platform, $H(g)$, is given by

$$H(g) = \sum_{s \in L_D} p(s) \cdot \left[\Lambda_h \cdot F_g \right] \tag{63}$$

$$= P_D \cdot \Lambda_h \cdot F_g . \tag{64}$$

And we can determine the average number of suspended calls on g-type platform, $A_w(g)$, from equation (40) as follows

$$A_w(g) = \sum_{k=1}^{N} A_W(k, g). \tag{65}$$

Thereafter, we can find the average time per suspension of a g-type call, $W(g)$. Using Little's law, this is

$$W(g) = A_w(g) / H(g). \tag{66}$$

Average Number of Suspensions Per Session

The average number of suspensions per session on g-type platform, $S(g)$, is the expected number of suspensions for a call on g-type platform during its lifetime. Firstly, we will determined the average rate of accommodating g-type calls, $E(g)$. Since some part of calls will be blocked due to lack of system capacity, the average rate of accommodating g-type calls , $E(g)$, can be expressed as

$$E(g) = \Lambda_n(g) \cdot (1 - P_B). \tag{67}$$

And, then the average number of suspensions per session on g-type platform, $S(g)$, can be written as

$$S(g) = H(g) / E(g). \tag{68}$$

Average Times Per Session

AVERAGE TIME IN SUSPENSION PER SESSION

The average time in suspension per session on g-type platform, $M_s(g)$, is the expected waiting time for a call on g-type platform during its lifetime. By Little's law, this can be written as

$$M_s(g) = S(g) \cdot W(g) = \frac{H(g)}{E(g)} \cdot \frac{A_w(g)}{H(g)} = \frac{A_w(g)}{E(g)}. \tag{69}$$

AVERAGE TOTAL LIFETIME OF A SESSION

The average total lifetime, regardless how it ends (call completion or forced termination), of a g-type call, $M_t(g)$, is the expected time that a call spends in the system. It should be noted that a call can finish its session successfully or unsuccessfully. By Little's law, this can be written as

$$M_t(g) = \frac{A_c(g) + A_w(g)}{E(g)}. \tag{70}$$

Suspension Time

The fraction of a call's lifetime that it is suspended is an important performance metric. For a call on a g-type platform, this is denoted by $L(g)$. This can be written as

$$L(g) = \frac{M_s(g)}{M_t(g)} = \frac{A_w(g)}{A_c(g) + A_w(g)}. \tag{71}$$

DISCUSSION OF RESULTS

Numerical results were generated using the approach described in this paper. For all figures, an unencumbered session duration of 100s was assumed. Two platform types, low mobility and high mobility, were considered. A mean dwell time of 500s was assumed for low mobility platforms and 100s for high mobility platform. A homogeneous system was assumed. The mean k-trial time of a g-type k-suspended session was chosen to be 20s for $1 \le k \le N$ (That is, $\overline{T}_r(k, g) = 20s$).

The abscissas for Figures 3 - 11 show call demand with the assumptions stated above. In these, the abscissas are the new call origination rate for platform type 1 (denoted as $\Lambda(1)$). The ratio of new call origination rates of other platform types to that of type 1 platforms were fixed by the parameters $\alpha(g)$. For all calculations in this paper, $\alpha(g) = 1$ was assumed.

When an active or suspended session requires a hand-off, the session is terminated if there are C connections in progress in the target cell *and* no waiting spaces are available. This probability is denoted P_H and is a calculated result. Figure 3 shows hand-off failure probability, P_H, as a function of new call origination rate on type 1 platform. While many parameters affect forced termination probability, the influence of a P_H is very strong. There are two important parameters that can control hand-off failure probability in our system configuration. One is the number cut-off priority channels, C_h. Clearly, it is seen that if more channels are reserved for hand-off or retry attempts, a smaller hand-off failure probability is obtained. However, as we can see in Figure 4, with increasing the arriving C_h, arriving calls (new calls) will be more likely to fail to acquire a connection. The other important parameter is the number of maximum allowable of reconnection attempts, *N*. As this parameter increased, a session is less likely to be terminated.

Figure 5 shows the dependence of the forced termination probability on the number of maximum allowable reconnection attempts, *N*. As we can see, increasing *N* results in fewer sessions being forced to terminate during their lifetimes. It is also seen in Figure 9 that this is mainly due to a reduction in maximum termination probability. We see that increase of maximum allowable reconnection attempt, *N*, has a dramatic affect on reducing of maximum termination probability.

Figure 6 shows the dependence of forced termination probability on the number of maximum supportable suspended sessions, *H*. The more calls that can be supported as suspended sessions, the less the forced termination probability. It is seen in Figure 8 that this is mainly because of reduction of non-maximum termination probability.

Figure 7 shows the forced termination probability for various value of C_h used for cut-off priority. When we increase the number of connections that are reserved for hand-off or reconnection attempts, forced termination probability is decreased. But (as we see in Figure 4) blocking of new calls is increased. It is also seen that calls on slow mobiles have smaller forced termination probability. That is because a call on slow mobile can finish its session with relatively fewer hand-offs during its lifetime.

Figure 10 shows the dependence of average time per suspension, $W(g)$, on call demand. The number of maximum allowable reconnection attempts, *N*, and the platform mobility are parameters. Consider suspended sessions on fast mobiles in comparison with those on slow platforms. Since reconnection attempts are made at hand-off events, the former would generally have earlier opportunities to resume active status. Thus calls on slow mobiles have greater average waiting time per suspension than calls on fast mobiles. This is shown in the figure. It is seen that increasing *N* increases $W(g)$. With increasing call demand this trend is more clear. This is because, with higher *N*, fewer calls will undergo forced termination and the queue of suspended calls will be increased.

Figure 11 shows the fractional suspension time with *N* as a parameter. It is seen that with increasing call demand the fractional suspension time increases. This is because with increasing call demand the system is increasingly crowded, so a call is more likely to be suspended. It is also seen that calls on fast mobiles have a greater fractional suspension time than calls on slow mobiles. Consider calls on fast mobiles in comparison with those on slow mobile. Recall from Figure 5 that increasing *N* reduces the forced termination probability for calls on both fast and slow platforms. Fast mobiles especially benefit from increasing *N*. That is, more are served to completion. However, we see from Figure 11 that these calls will spend more of their lifetime in suspension. Thus, both the likelihood of successful completion and the fraction of time spent in suspension increase with increasing *N*. The effect is more pronounced for calls on fast mobiles.

CONCLUSIONS

In order to support mixed user types including multimedia users and mobile computing in a cellular communication system, we consider a scheme in which there are automated attempts to maintain network connectivity for users. In the case of a link failure, this allows a user to continue in a temporary *off-line* mode while awaiting an active network connection in the background. The multi-dimensional birth-death process framework that we are developing can be used to compute theoretical traffic performance characteristics for the proposed scheme. The model captures mixed platform mobilities and call session types, hand-off issues and priority, as well as reconnection attempts. Traffic performance depends on, traffic demand and mix, the amount of priority given for hand-off calls and the limit on the number of allowable reconnection attempts. Forced termination of a call during its lifetime is studied. Two kinds of forced termination are considered. Those are non-maximum termination and maximum termination. Relevant traffic performance measures can be calculated for example system parameters.

REFERENCES

[1] CCITT Recommendation I.212, "Broadband Aspects of ISDN," CCITT IXth Plenary Assembly, Melbourne, Australia, vol. 3, fasc. III. 7, 1988.

[2] D. Raychaudhuri, "Wireless ATM: An Enabling Technology for Multimedia Personal Communication," Wireless Networks 2, pp. 163-171, J.C. Baltzer AG, Science Publishers.

[3] A. Acampora and M. Naghishineh, "An Architecture and Methodology for Mobile Executed Hand-off in Cellular ATM Networks," IEEE JSAC, Oct. 1994. pp. 1365-75.

[4] S. S. Rappaport, "The Multiple-Call Hand-off Problem in High-Capacity Cellular Communication System," IEEE Trans. On Vehic. Technol., Aug. 1991, vol. VT-40, no.3 pp. 546-557.

[5] S. S. Rappaport, " Blocking Hand-off and Traffic Performance for Cellular Communication System with Mixed Platforms," IEE (British) Proceedings, Part I, Communications, Speech and Vision, October 1993, vol. 40, no. 5, pp. 389-401.

[6] D. Hong and S. S. Rappaport, " Traffic Model and Performance Analysis for Cellular Mobile Radio Telephone Systems with Prioritized and Non-Prioritized Hand-off Procedures," IEEE Trans. On Vehicular Technology, Aug. 1986, vol. VT-35, no. 3, pp. 77-92.

[7] Y. S. Park and S. S. Rappaport, " Cellular Communication Systems with Voice and Background Data," Proc. MoMuC-3, 3rd Int'l. Workshop on Mobile Multimedia Communications, Princeton, N.J., Sept. 25-27, 1996, pp. A.2.2.2:1-6. (Also to appear in forthcoming book).

[8] C. Purzynski and S. S. Rappaport, " Prioritized Resource Assignment for Mobile Cellular Communication Systems with Mixed Services and Platform Types," IEEE Trans. On Vehicular Technology, August 1996, vol. 45, No.3, pp. 443-458.

[9] R. B. Cooper, Introduction to Queuing Theory, 2nd ed., New York: Elsevier-North Holland, 1981.

[10] A. M. Chen and R. R. Rao, "Resource Management for Third Generation Cellular Communication Systems," Proc. IEEE Vehicular Technology Conference VTC '97, Phoenix, May 4-7, 1997.

[11] P. V. Orlik and S. S. Rappaport , "Traffic Performance and Mobility Modeling of Cellular Communications with Mixed Platforms and Highly Variable Mobilities," Proc. IEEE Vehicular Technology Conference VTC '97, Phoenix, May 4-7, 1997, pp. 587-591.

[12] P. V. Orlik and S. S. Rappaport, "A Model for Teletraffic Performance and Channel Holding Time Characterization in Wireless Cellular Communications with General Session and Dwell Time Distributions," CEAS Technical Report No.742, July 31, 1997, College of Engineering and Applied Sciences, State University of New York, Stony Brook, NY 11794.

APPENDIX: *Operation of The Reconnection Counter at Mobile Terminal.*

Figure 2 shows the detailed operation of the reconnection counter and reconnection process. The reconnection attempt counter at the mobile operates in the following way. If the session is active, the counter is set at $k=0$. If then the session is suspended the counter is immediately incremented to $k=1$ (waiting for the first reconnection attempt). A random time is generated for a possible first retry attempt epoch. If a hand-off attempt arises before this retry attempt time, a retry attempt will be made at the hand-off epoch. Otherwise the retry attempt will be made at the (randomly) scheduled reconnection epoch. In either case, if the attempt fails, the counter will be incremented by 1. If the attempt succeeds, the session becomes active and the counter is reset to $k=0$. This will continue until one of the following three events occurs.

1) The session is successfully completed.
2) The session is forced to terminate due to the lack of waiting space in the target cell when a hand-off attempt is made (This is called *non-maximum termination*).
3) There are N successive failed reconnection attempts (This is called *maximum termination*).

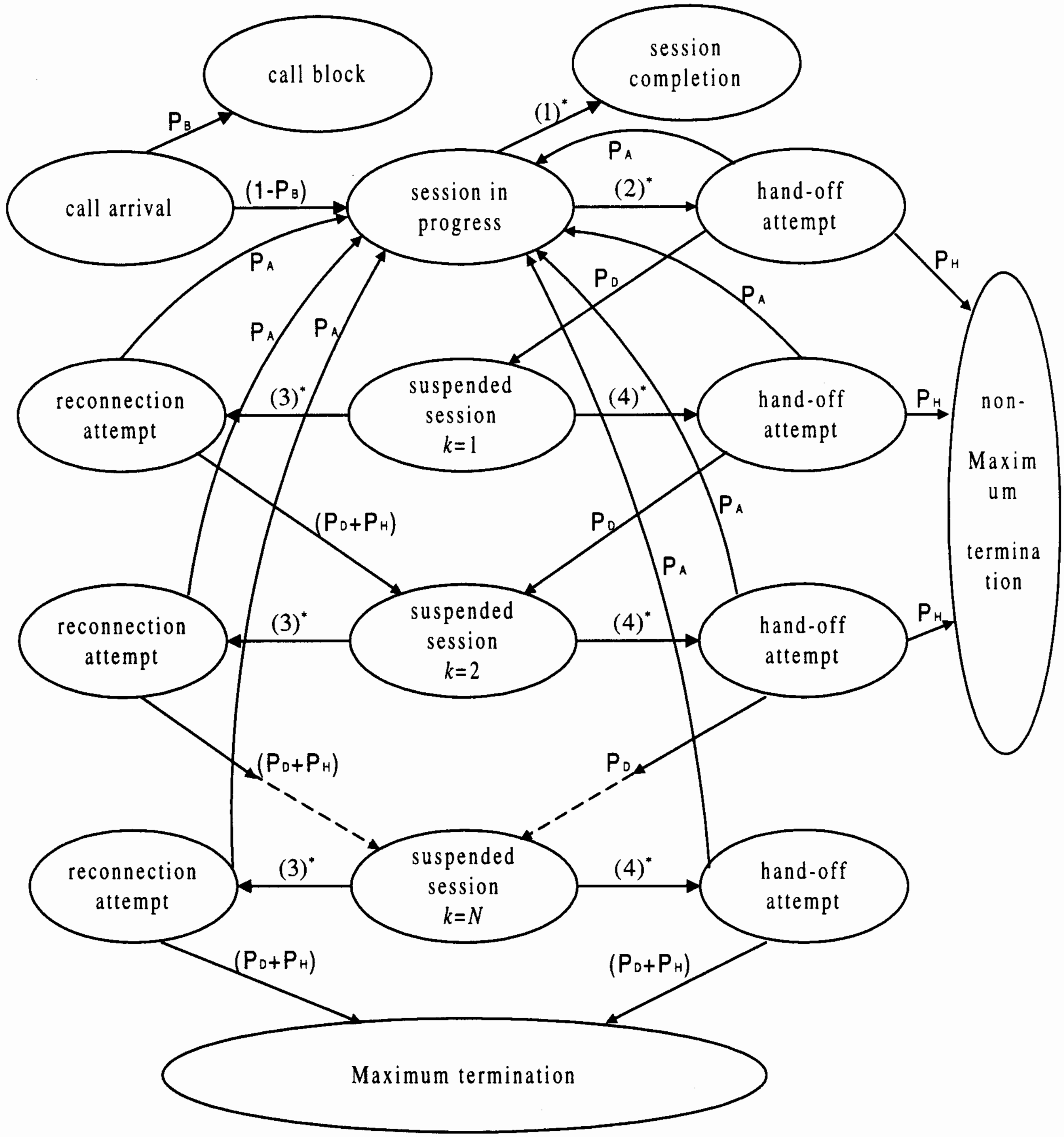

Figure 1. Flow graph of call event where $(1)^* = \dfrac{\mu(g)}{\mu_D(g)+\mu(g)}$, $(2)^* = \dfrac{\mu_D(g)}{\mu_D(g)+\mu(g)}$, $(3)^* = \dfrac{\mu_r(k,g)}{\mu_D(g)+\mu_r(k,g)}$, and $(4)^* = \dfrac{\mu_D(g)}{\mu_D(g)+\mu_r(k,g)}$.

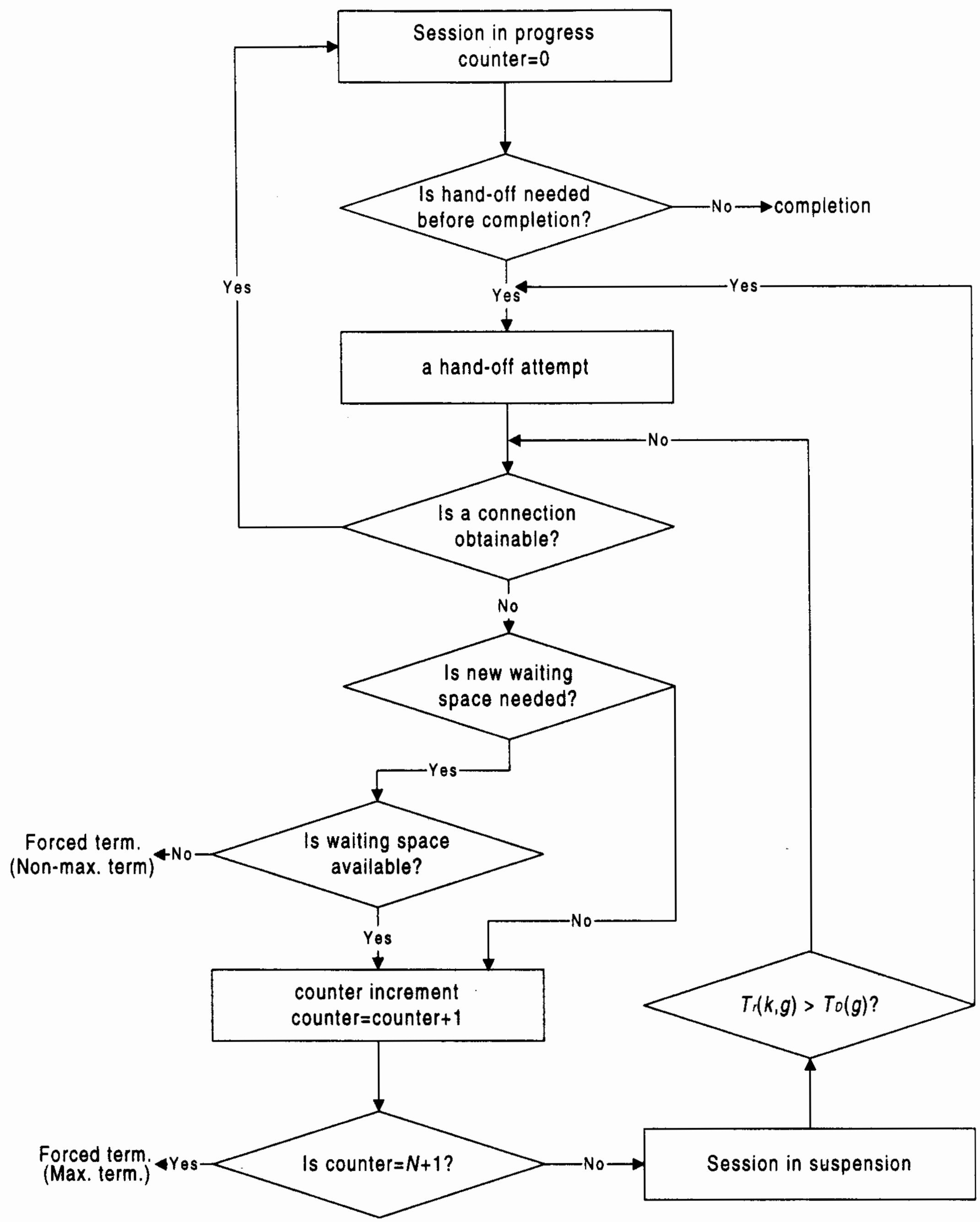

Figure 2: Flow chart of operation of the reconnection counter.

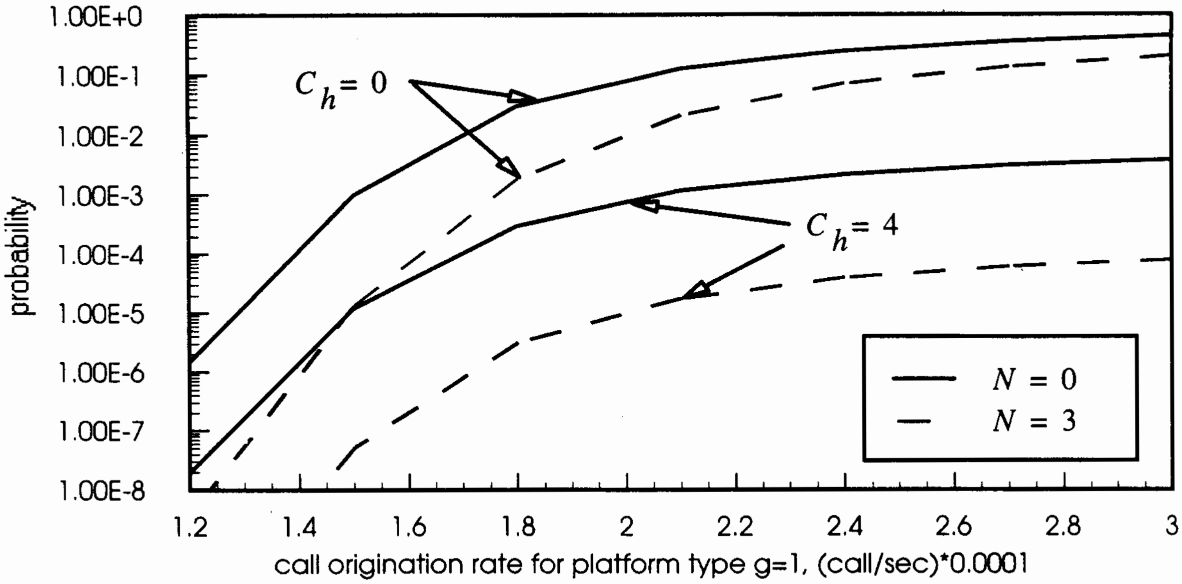

Figure 3: Hand-off failure probability: C=25, G=2, $\nu(1,0) = \nu(2,0) = 300$, $\alpha(2) = \Lambda(2) / \Lambda(1) = 1.0$, $\overline{T}(1) = \overline{T}(2) = 100$s, $\overline{T}_D(1) = 100s$, $\overline{T}_D(2) = 500s$, H=3, $\overline{T}_r(1,1) = \overline{T}_r(2,1) = \overline{T}_r(3,1) = \overline{T}_r(2,1) = \overline{T}_r(2,2) = \overline{T}_r(2,3) = 20$s .

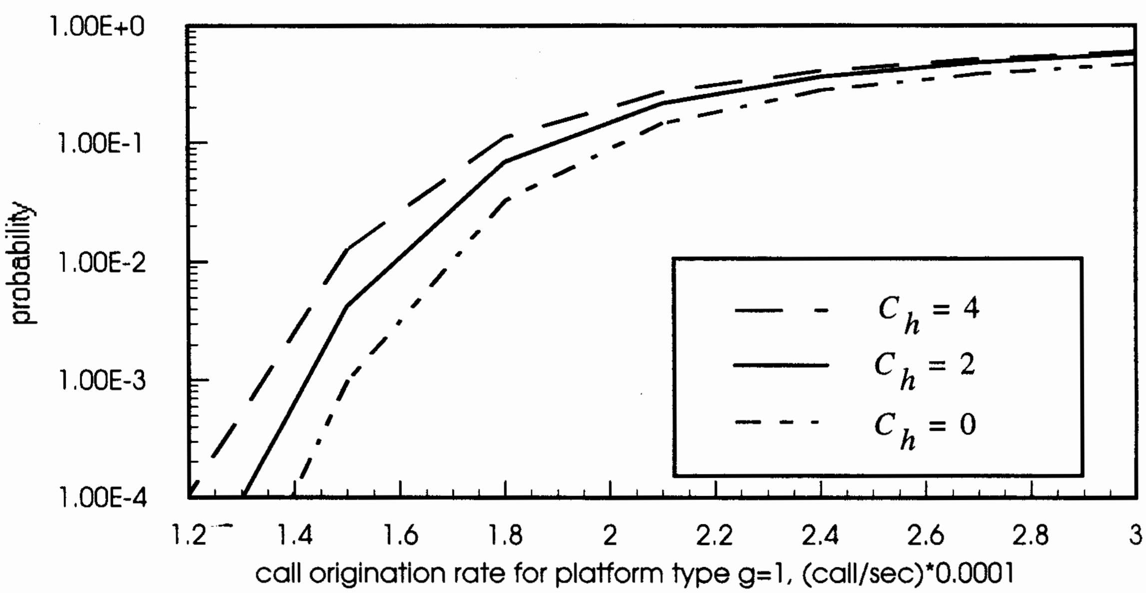

Figure 4: Blocking probability: C=25, G=2, $\nu(1,0) = \nu(2,0) = 300$, $\alpha(2) = \Lambda(2) / \Lambda(1) = 1.0$, $\overline{T}(1) = \overline{T}(2) = 100$s, $\overline{T}_D(1) = 100s$, and $\overline{T}_D(2) = 500s$.

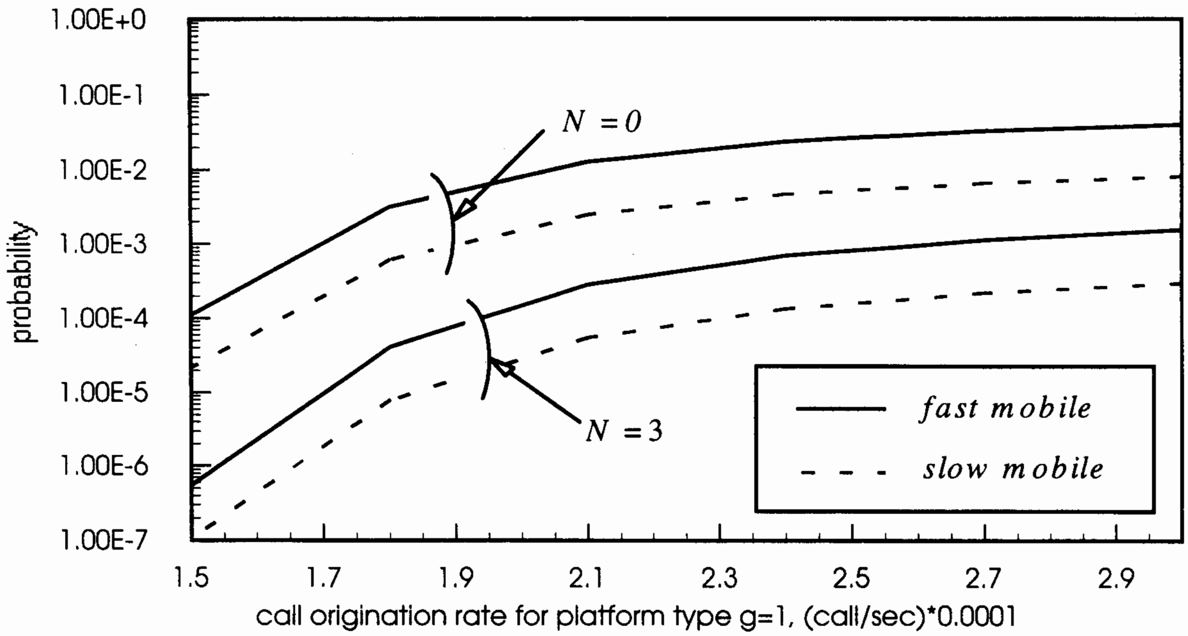

Figure 5: Forced termination probability: C=25, G=2, $C_h = 2, \nu(1,0) = \nu(2,0) = 300$, $\alpha(2) = \Lambda(2)/\Lambda(1) = 1.0$, $\overline{T}(1) = \overline{T}(2) = 100\text{s}$, $\overline{T}_D(1) = 100s$, $\overline{T}_D(2) = 500s$, H=3 and $\overline{T}_r(1,1) = \overline{T}_r(2,1) = \overline{T}_r(3,1) = \overline{T}_r(2,1) = \overline{T}_r(2,2) = \overline{T}_r(2,3) = 20\text{s}$.

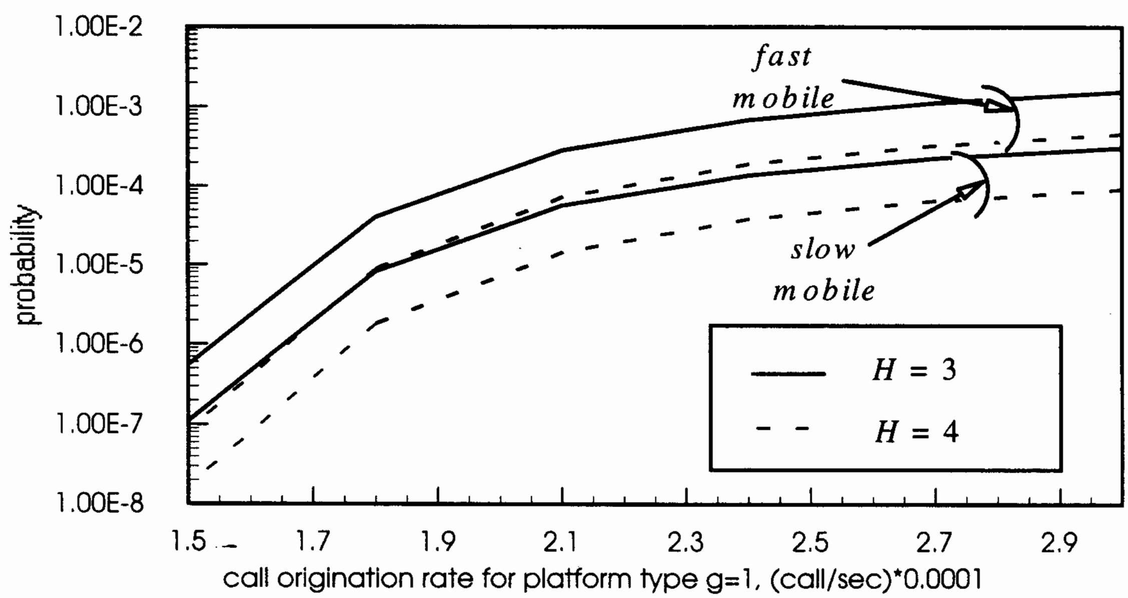

Figure 6: Forced termination probability: C=25, G=2, $C_h = 2, \nu(1,0) = \nu(2,0) = 300$, $\alpha(2) = \Lambda(2)/\Lambda(1) = 1.0$, $\overline{T}(1) = \overline{T}(2) = 100\text{s}$, $\overline{T}_D(1) = 100s$, $\overline{T}_D(2) = 500s$, N=3 and $\overline{T}_r(1,1) = \overline{T}_r(2,1) = \overline{T}_r(3,1) = \overline{T}_r(2,1) = \overline{T}_r(2,2) = \overline{T}_r(2,3) = 20\text{s}$.

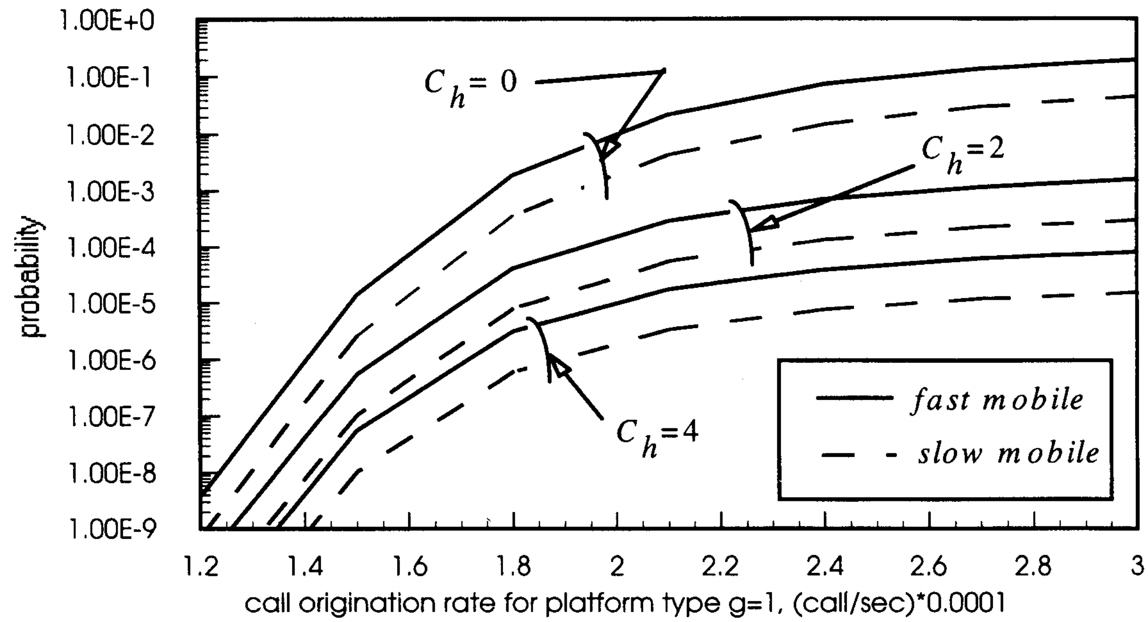

Figure 7: Forced termination probability: C=25, G=2, $\nu(1,0) = \nu(2,0) = 300$, $\alpha(2) = \Lambda(2) / \Lambda(1) = 1.0$, $\overline{T}(1) = \overline{T}(2) = 100$s, $\overline{T}_D(1) = 100s$, $\overline{T}_D(2) = 500s$, H=3, N=3 and $\overline{T}_r(1,1) = \overline{T}_r(2,1) = \overline{T}_r(3,1) = \overline{T}_r(2,1) = \overline{T}_r(2,2) = \overline{T}_r(2,3) = 20$s .

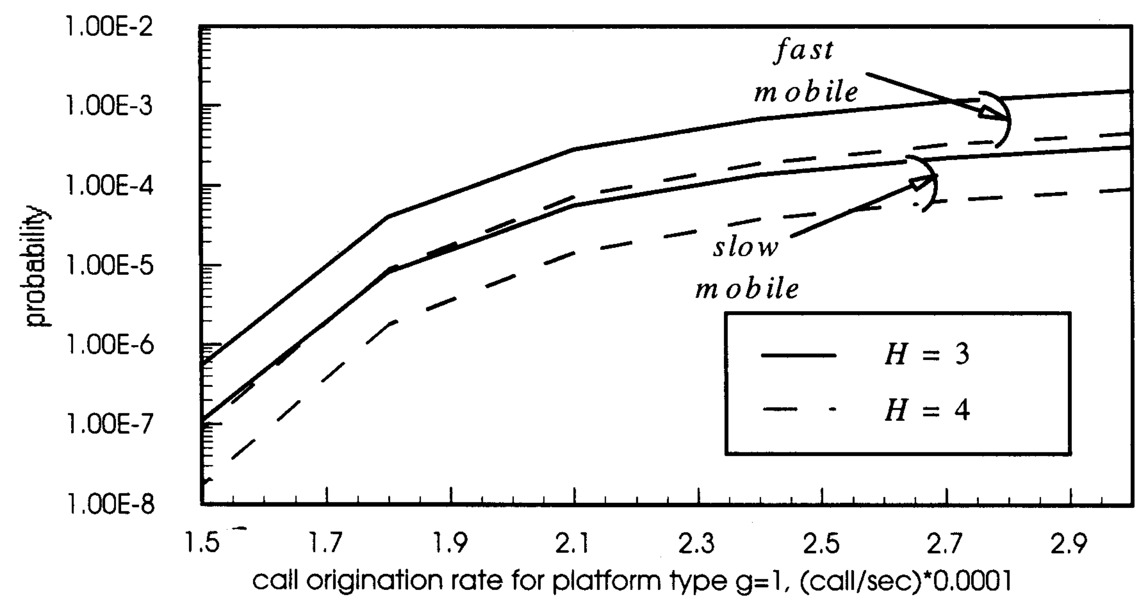

Figure 8: Non-maximum termination probability: C=25, G=2, $C_h = 2$, $\nu(1,0) = \nu(2,0) = 300$, $\alpha(2) = \Lambda(2) / \Lambda(1) = 1.0$, $\overline{T}(1) = \overline{T}(2) = 100$s, $\overline{T}_D(1) = 100s$, $\overline{T}_D(2) = 500s$, N=3 and $\overline{T}_r(1,1) = \overline{T}_r(2,1) = \overline{T}_r(3,1) = \overline{T}_r(2,1) = \overline{T}_r(2,2) = \overline{T}_r(2,3) = 20$s.

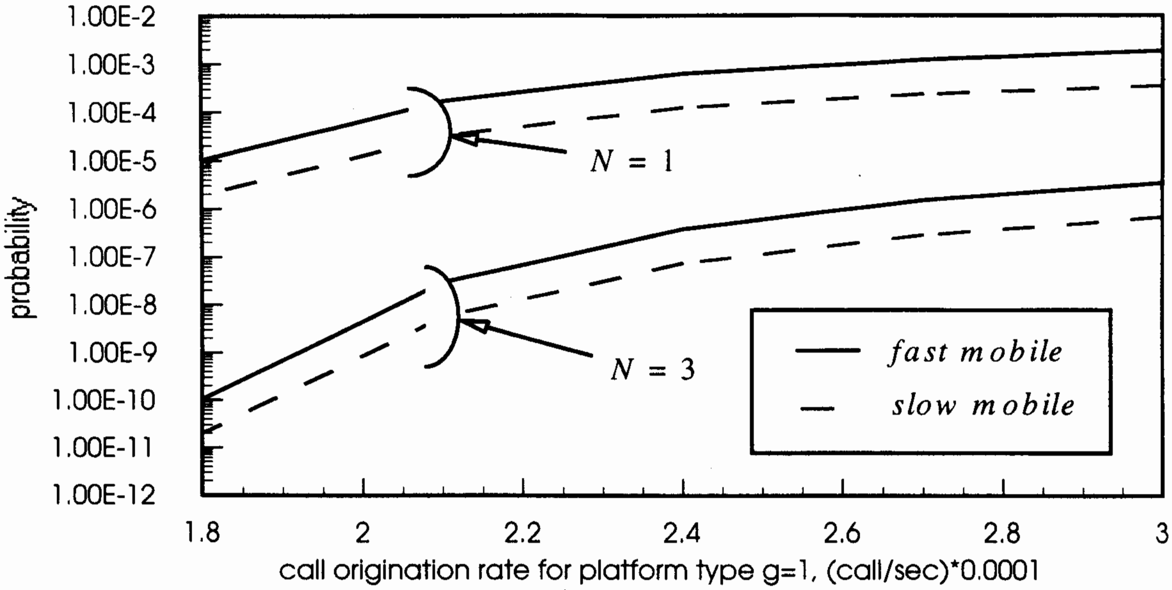

Figure 9: Maximum termination probability: C=25, G=2, $C_h = 2, \nu(1,0) = \nu(2,0) = 300$, $\alpha(2) = \Lambda(2)/\Lambda(1) = 1.0$, $\overline{T}(1) = \overline{T}(2) = 100$s, $\overline{T}_D(1) = 100s$, $\overline{T}_D(2) = 500s$, H=3 and $\overline{T}_r(1,1) = \overline{T}_r(2,1) = \overline{T}_r(3,1) = \overline{T}_r(2,1) = \overline{T}_r(2,2) = \overline{T}_r(2,3) = 20$s.

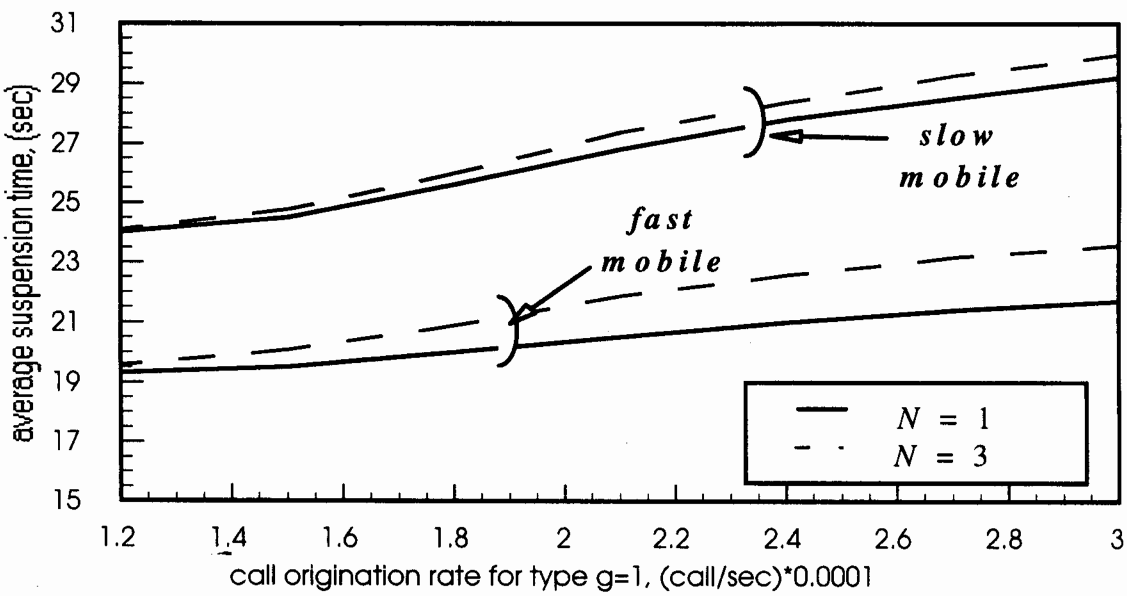

Figure 10: Average time per suspension: C=25, G=2, $C_h = 2$, $\nu(1,0) = \nu(2,0) = 300$, $\alpha(2) = \Lambda(2)/\Lambda(1) = 1.0$, $\overline{T}(1) = \overline{T}(2) = 100$s, $\overline{T}_D(1) = 100s$, $\overline{T}_D(2) = 500s$, H=3 and $\overline{T}_r(1,1) = \overline{T}_r(2,1) = \overline{T}_r(3,1) = \overline{T}_r(2,1) = \overline{T}_r(2,2) = \overline{T}_r(2,3) = 20$s.

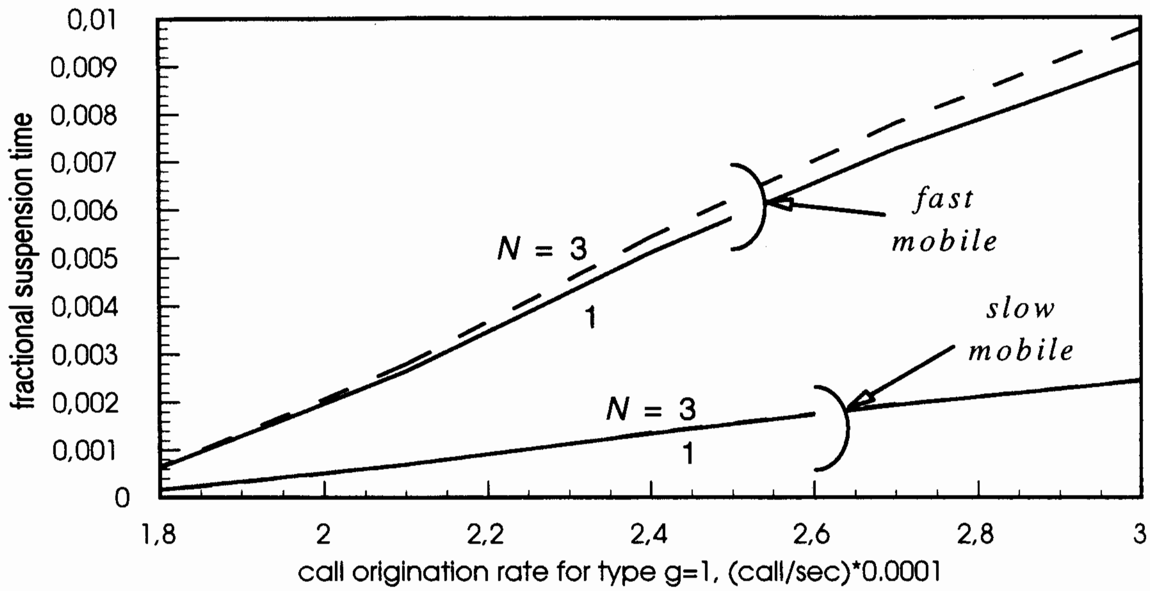

Figure 11: Fractional suspension time: C=25, G=2, $C_h = 2$, $\nu(1,0) = \nu(2,0) = 300$, $\alpha(2) = \Lambda(2) / \Lambda(1) = 1.0$, $\overline{T}(1) = \overline{T}(2) = 100s$, $\overline{T}_D(1) = 100s$, $\overline{T}_D(2) = 500s$, H=3 and $\overline{T}_r(1,1) = \overline{T}_r(2,1) = \overline{T}_r(3,1) = \overline{T}_r(2,1) = \overline{T}_r(2,2) = \overline{T}_r(2,3) = 20s$.