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

Technical Report no. 732

Cellular Communication Systems with Voice and Background Data

Yunsang Park and Stephen .S. Rappaport

Department of Electrical Engineering
State University of New York at Stony Brook
Stony Brook, New York 11794-2350
e-mail:park@sbee.sunysb.edu, rappaport@sunysb.edu

Date : Dec. 10, 1997

CELLULAR COMMUNICATION SYSTEMS WITH VOICE AND BACKGROUND DATA †

Yunsang Park and Stephen S. Rappaport

Department of Electrical Engineering

State University of New York at Stony Brook

Stony Brook, New York 11794-2350

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

ABSTRACT

Cellular communication systems that support both voice calls and background data calls arising from various platform types having different mobility characteristics are considered. Hand-off problems of voice and data calls are studied. A tractable analytical model for traffic performance is developed using multidimensional birth-death processes. Preemptive priority is used to assure transparency to voice calls. We examine two possible strategies for handling preempted data calls. One is a prioritized scheme and the other is non-prioritized scheme. Theoretical traffic performance characteristics are calculated. For voice calls, blocking probability and forced termination probability are calculated. The possible causes of data call termination are considered. These include forced termination probability of a data call and pushed-out failure probability. The average time that a data call spends in the system and the average waiting time of data calls are determined.

1. INTRODUCTION

Although cellular communication systems were primarily designed for voice communications, current advances include provisions for data transmission as well. It is desirable to support this demand with existing bandwidth and with minimal impact on the existing infrastructure.

One approach is to use the channels during idle times between voice calls to transmit data calls. This concept is based on different attributes of voice and data services. A delay of voice service in cellular communication network of 100 ms will be recognizable and irritating to the user, but a small amount of delay for the data users is not critical. An example of such "background data transmission" is CDPD, Cellular Digital Packet Data, which is designed to provide packet data services in an overlay to the existing analog cellular telephone network [1], [2].

The present paper is inspired by [1] and [3], but we use techniques put forth in [4] and [5] to enable a more complete analysis. In particular, our framework allows consideration of a variety of platforms having different mobility characteristics in the same system as well as cut-off priority for hand-off calls. We devise a model to compute fundamental traffic performance measures for a cellular communication system where voice calls as well as background data calls are accommodated. In this system, voice calls have preemptive priority over data calls so that the data transmissions are transparent to voice call users. Arriving voice calls can displace active data calls. These displaced data calls are held on queue for resources (channels) to become available. We will say that such queued calls are in a waiting room. We consider an upper limit on the number of waiting calls. For time-insensitive data calls the delay incurred because of voice call interruptions can be tolerated within broad limits.

We consider *mixed* platform types. That is platforms with differing mobility characteristics are present in the system. In addition to preemptive priority for voice calls with respect to data, a cut-off priority scheme is used to reserve some resources (channels) for voice call hand-offs. The problem is modeled using multidimensional birth-death processes [4], [5], [6]. We define a state space and identify state transitions according to several schemes that specify how voice calls and data calls are

managed. For each we formulate the global balance equations and solve for the state probabilities, using the framework developed in some of our earlier work [4], [5], [7].

When a voice call arrives, it may be necessary to suspend an active data call in order to accommodate it. This would be the case if all channels are occupied and some active data calls are in progress. The choice of which *active* data call to suspend is assumed to be random. If a voice call arrives when the waiting room is full, a data call must be terminated to accommodate the voice call. Two cases are considered here: (i) An active data call will always receive a waiting room space. The choice of which suspended data call to terminate is random. (ii) An active data call is not guaranteed waiting room space. One of the data calls (either in the waiting room or active) is chosen at random and is terminated. We call the first scheme a prioritized scheme, the second is non-prioritized.

Another issue of call management is which a data call that has been suspended should be activated when a channel becomes available. Again there are several possibilities. These include activating: the oldest call; the youngest call; or any call regardless of how long it has been waiting [3]. Here we assume that data calls held are activated randomly when a channel becomes available.

The problem is to compute the performance measures for the schemes described above for a given finite waiting room for data calls and a given number of channels in each cell. For numerical purposes we assume that there are two types of calls. One, typified by voice, requires immediate connections for service. The other, typified by time-insensitive data, can be queued. It is possible to consider (within the same categories) calls that can be distinguished by different intended session duration distributions. This is deferred to future work. A system that is spatially homogeneous is considered here although this too can be generalized.

In this paper, first voice traffic performance is considered. Data traffic has no influence on voice call traffic performance under the specified access schemes. So traffic performance for voice calls is the same as that of a system with no background data traffic. The blocking probability, (which is the fraction of new voice calls that are denied access to a channel), and, the forced termination probability, (which is the probability that a non-blocked call is interrupted due to hand-off failure

3

during its lifetime) are the important factors for voice call users. These probabilities are calculated from the state probabilities.

Data calls, of course, are profoundly affected by the voice traffic. The average rate of data call suspensions and the average number of waiting calls are determined from the state probabilities. Data traffic performance measures are considered. These include the average time that a data call spends in the system and the average waiting time of data calls.

The model determines blocking probability, average waiting time, pushed-out termination probability, and forced termination probability for data calls as function of overall offered call volume. Because suspended data calls must undergo hand-off as the supporting platforms move, forced termination of *suspended* data calls as well as active data are considered.

The result is a comprehensive computational model for cellular voice systems with background data. The model allows theoretical determination of the major pertinent traffic performance characteristics for voice and data calls supported on mixed platforms with various mobilities.

2. MODEL DESCRIPTION

We consider a large geographic region covered by cells that are defined by proximity to specific network gateways. The region is traversed by large numbers of mobile platforms that are of several types. The platform types differ primarily in the mobility characteristics. A platform can generate two types of calls which are voice type and data type. In the development presented here, a platform can support only one call at any given time, whether data type or voice type. These connection-oriented data calls use the channel during idle times between voice calls so that voice call traffic is not influenced by the background data calls.

As in previous work [4], [5], [7], we use the concept of dwell time, the amount of time that a mobile platform is within communications range of a given gateway, to characterize the mobility of

4

platforms. The dwell time of a platform depends on many factors such as mobility, signal power, propagation conditions, etc. We model the probability density function (pdf) of dwell time as a negative exponential distribution. Different mobilities are modeled by different parameters for this p.d.f.. Generalization to a broader class of mobilities can be accomplished within the same framework [4], [5].

The unencumbered call (session) duration of voice call can be modeled as negative exponential function (ned). Session duration of a data call, which depends on the size of data and speed of modem, can be modeled as negative exponential distribution (ned). It is possible to consider within the same formulation that there are different types of calls in each of these categories that can be distinguished by distinct intended session duration distributions. For instance, a relatively large data file transmission, such as an image file, can be distinguished from a small data file transmission using different session duration distributions. It is also possible to consider different data service types that are modeled with a different form session duration distribution. We assume that the arrival processes of both voice calls and data calls are Poisson point processes and are independent of one another.

3. EXAMPLE PROBLEM STATEMENT

Population : The system can support G types of mobile platforms, indexed by $\{g = 1, 2, 3, \dots G\}$. Only one call can be supported by a platform at any given time no matter what type of call. The voice call origination rate from a noncommunicating g-type platform is denoted $\Lambda_v(g)$. We define $\alpha(g) = \Lambda_v(g)/\Lambda_v(1)$. The data call generation rate from a noncommunicating g-type platform is denoted $\Lambda_w(g)$. It is, also, defined $\beta(g) = \Lambda_w(g)/\Lambda_v(g)$. Potentially, a platform can generate two types of calls, voice type and data type. The number of noncommunicating g-type platforms in any cell is denoted $v(g,0)$. Therefore, the total voice call generation rate for g-type platforms in a cell can be denoted $\Lambda_{nv}(g) = \Lambda_v(g) \times v(g,0)$ and the total data call generation rate for g-type platforms in a cell can be denoted $\Lambda_{nw}(g) = \Lambda_w(g) \times v(g,0)$. It is assumed that the number of noncommunicating platforms is much larger than the number of channels in a cell so that the call generation rate does not depend on the number of calls being served (this is called an infinite population model).

5

Resources : Each cell or gateway has W waiting spaces for data calls and a total of C channels to serve both voice and data. There are no channel quotas. C_h channels in each cell are reserved for voice hand-off calls but background data calls can access those channels if they are available. A new voice call will be served if there are fewer than $C - C_h$ channels occupied by voice calls. A hand-off attempt of voice call will fail only if all channels are occupied by voice calls. Arrival of a data call, whether it is a new or hand-off call, will be held if the number of channels in use in the target cell is C and waiting space is available. If no waiting space is available when a new data call or a data hand-off call arrives, the call will be blocked.

Session and Dwell Times : The unencumbered session duration of a voice call on a g-type platform is a ned random variable, $T_v(g)$, having a mean $\overline{T}_v(g)=1/\mu_v(g)$. The unencumbered session duration of data call on a g-type platform is a ned random variable, $T_w(g)$, having a mean $\overline{T}_w(g)=1/\mu_w(g)$. The dwell time is defined as the duration of time that a two-way link can be maintained between a platform and its current base, for whatever reason. These reasons include fading, propagation effects, and path loss. The platform is considered to "leave" the cell at the (random) dwell time. *Communicating* platforms that leave a cell generate a hand-off arrival to some other cell. The dwell time in a cell for a g-type platform is a ned random variable, $T_D(g)$, having a mean $\overline{T}_D(g)=1/\mu_D(g)$.

The problem is to calculate relevant performance characteristics of voice call traffics and data call traffics. These include *blocking probability, hand-off failure probability,* and *forced termination probability* for voice and data calls as defined in [5]. We also introduced *pushed-out failure probability, average time a data call spends in the system,* and *average waiting time for those data calls that wait* as additional performance measures for data calls. *Pushed-out failure probability* is defined as the average fraction of data calls on g-type of mobile platform that are terminated due to an arrival of a voice call. *Average time a data call spends in the system* is defined to be the average amount of time a data call on g-type platform spends in the system. *Average waiting time for those data calls that wait* is defined to be the average amount of time a data call on g-type platform spends in the waiting space before entering service, being forced terminate, or being pushed-out.

6

The mathematical analysis is similar to that used in [5].

4. STATE DESCRIPTION

Considering a single cell, we define the state of that cell by a sequence of nonnegative integers. This can be conveniently written as G 3-tuples.

$$v_1, z_1, w_1,$$
$$v_2, z_2, w_2,$$
$$\vdots, \vdots, \vdots,$$
$$v_G, z_G, w_G, \tag{1}$$

where $v_g\{g = 1, 2, \ldots\ldots, G\}$ is the number of voice calls on g-type platforms, $z_g\{g = 1, 2, \ldots\ldots, G\}$ is the number of data calls served on g-type platforms, and $w_g\{g = 1, 2, \ldots\ldots, G\}$ is the number of data calls on g-type platforms that are waiting for service. Then, we order the states using an index $s = 0, 1, \ldots\ldots, S_{max}$ for convenience. Thereafter, v_g, z_g, and w_g are explicitly dependent on the state. That is $v_g = v(s,g)$, $z_g = z(s,g)$, and $w_g = w(s,g)$.

If only one type of mobile platform is present in the system, the state can be described by 2 state variables, the number of voice calls (v) and the number of data calls ($w + z$) in the system. However, when multiple types of mobile platforms coexist in a cell, it is impossible to uniquely describe a state using only two state variables for each platform type. For example, consider a system that has 10 channels and 10 waiting spaces. Suppose that there are 6 voice calls (v=6) and 11 data calls ($z + w$=11) on the channels or in the waiting room at any given time. If only one type of mobile platform (say type-1 platform) is present in the system, we can deduce that 7 data calls are waiting for service ($w=w_1$=7) in the waiting spaces, since 10 channels are in the system. Therefore, we just need 2 state variables, the number of voice calls (v) and the number of data calls in the system ($z + w$), to describe a state. If, however, more than one type of mobile platform is in the system (suppose type-1 and type-2 platforms), we can not determine the number of data calls in the waiting room which is on type-1 mobile platform or on type-2 mobile platform. We may know the total

7

number of data calls in the waiting spaces ($w_1 + w_2 = 7$), but we do not know each value of w_1 or w_2. Therefore, we use three state variables for *each* platform type (v_g, z_g, w_g) to specify the cell state.

When the cell is in state s, the following characteristics can be determined. The number of channels that are occupied by voice calls is

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

(2)

The number of channels that are used by background data calls is

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

(3)

The number of data calls held by system is

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

(4)

Therefore, the number of channels being used by g-type platforms is

$$j(s,g) = v(s,g) + z(s,g).$$

(5)

The total number of channels in use can be expressed as

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

(6)

Constraints on permissible states include $j(s) \leq C$ and $w(s) \leq W$ where C is the number of channels in a cell and W is the number of waiting room spaces for the data call held in the system. The number of states needed to represent a system with two platform types and both voice as well as data calls is shown in Table 1. The number of possible states can be formidable, so it is important to use efficient algorithms to solve the resulting equations. A description of the computational algorithm is given in Appendix A.

5. STATE SPACE

The state of a cell corresponds to a point in the state space. The state can change due to several events such as call generation, call completion, hand-off call arrival, and hand-off call departure. The

state transition rates are state dependent. Figure 1 shows the state transition diagram for the case of $G = 1$ (single platform mobility type).

The state space is subdivided into 4 different regions and 7 different boundaries as follows. We define the set of states in which the number of voice calls is less than $C - C_h$ and the total number of calls is less than C as R_1

$$R_1 = \{s : \ j(s) < C, \ \ v(s) < C - C_h, \ \ w(s) = 0\}. \tag{7}$$

In this region, *all types of calls*, whether voice or data type, whether new or hand-off call, can access a channel and be served without any effect on calls in progress.

We define the set of states in which the number of voice calls is less than $C - C_h$ and the total number of calls in the system is greater than C but less than $C + W$ as R_2

$$R_2 = \{s : \ j(s) = C, \ \ v(s) < C - C_h, \ \ 0 < w(s) < W\}. \tag{8}$$

When the system is in region, R_2, an arriving data call, whether new or hand-off call, must wait for service to assure the transparency of data calls to the voice call users. On the other hand, *an arriving voice call*, whether new or hand-off call, can acquire a channel with the suspension of an active data call.

The set of states in which $C - C_h$ or more channels are occupied by voice calls, the number of voice calls is less than C, and the total number of calls is greater than C but less than $C + W$ is R_3

$$R_3 = \{s : \ j(s) = C, \ \ C - C_h \leq v(s) < C, \ \ 0 < w(s) < W\}. \tag{9}$$

In this region, *only voice hand-off calls* are allowed to access a channel. (To accommodate such a call an active data call will be suspended). An arriving data call will be held (forced to wait) and a new voice call will be blocked.

We define the set of states in which $C - C_h$ or more channels are used by voice calls, the number of voice calls in less than C, and the total number of calls is less than C as R_4

$$R_4 = \{s : \ j(s) < C, \ \ C - C_h \leq v(s) < C, \ \ w(s) = 0\}. \tag{10}$$

9

In this region, only *new voice calls* will be blocked. *A voice hand-off* call, *a new data* call, and *a hand-off data* call can access a channel.

We define a boundary as a set of states which does not belong to one of the regions defined previously. When the system is on the boundary between R_1 and R_2, the behavior is different from that in either R_1 or R_2. If a voice call, whether a new or hand-off call, arrives, an active data call must be suspended. This attribute is identical to that of R_2. However, on this boundary the effect of a voice call departure event is not same as that in R_2, rather it is identical to that in R_1. Since there are no data calls held by the system, no data call will be reactivated when a voice call completes or departs. Therefore, we treat this boundary as a distinct region.

We define the set of states that are on the boundary between R_1 and R_2 as B_1

$$B_1 = \{s: \quad j(s) = C, \quad v(s) < C - C_h, \quad w(s) = 0\}. \tag{11}$$

On this boundary, an arrival of a voice call, whether a new or hand-off call, causes a suspension of an active data call. An arriving data call, whether a new or hand-off call, will be held.

Similarly, we define the set of states as the boundary between R_3 and R_4 as B_6

$$B_6 = \{s: \quad j(s) = C, \quad C - C_h \leq v(s) < C, \quad w(s) = 0\}. \tag{12}$$

On this boundary, the effect of a hand-off voice call arrival event is the same as that of a state which is in R_3 (an active data call is suspended). But the effect of a voice call departure event is the same as that of a state which is in R_4 (no data call will be reactivated since there is no waiting data call for service). On this boundary, new voice call will be blocked and a data call, whether a new or hand-off call, will be held.

We define B_3 as the set of states in which all channels are occupied by voice calls and there is at least one held data call and at least one empty waiting space for a data call

$$B_3 = \{s: \quad j(s) = C, \quad v(s) = C, \quad 0 < w(s) < W\}. \tag{13}$$

10

On this boundary, an arriving voice call, whether a new or hand-off call, will be blocked. An arriving data call, whether a new or hand-off call, will be held.

We define the set of states for which a voice call arrival, whether new or hand-off call, causes the termination of a data call as B_2

$$B_2 = \{s: \ j(s) = C, \ v(s) < C - C_h, \ w(s) = W\}. \tag{14}$$

When all waiting spaces for suspended data calls are full, an arriving data call, whether a new or hand-off call, will be blocked due to the lack of waiting spaces. An arrival of a voice call, whether a new or hand-off call, causes the termination of a data call.

In the set of states, B_4 (defined below), only the arrival of a voice hand-off call causes a data call termination.

$$B_4 = \{s: \ j(s) = C, \ C - C_h \leq v(s) < C, \ w(s) = W\}. \tag{15}$$

New voice calls however will be blocked if $s \in B_4$.

When all channels are occupied by voice calls and all waiting spaces are full, an arrival of a call, no matter what type, will be blocked. We define the sets of states in which C channels are serving voice calls and W waiting spaces are full as B_5

$$B_5 = \{s: \ j(s) = C, \ v(s) = C, \ w(s) = W\}. \tag{16}$$

In this region, only call completions and hand-off departures cause a change in cell state.

Finally, we define the sets of states in which all channels are serving voice calls but there is no data call in the system as B_7

$$B_7 = \{s: \ j(s) = C, \ v(s) = C, \ w(s) = 0\}. \tag{17}$$

In this region, a voice call arrived, whether new or hand-off call, will be blocked and an arrival of a data call, whether new or hand-off call, make the cell state change to region B_3.

11

We summarize the attributes of the regions and boundaries in Table 2.

6. STRATEGY FOR MANAGEMENT OF DATA CALLS

Since data calls can access channels only during idle times between voice calls, the system must manage data call traffic. The issues are when, how and which data calls are suspended, terminated and reactivated. Many different strategies are possible. Some were considered in [3]. In this paper, we assume that the system randomly chooses a data call for reactivation, suspension, or termination, if it is necessary.

If a new or hand-off voice call arrives when the system is in B_2, or if a hand-off voice call arrives when the system is in B_4, a data call suspension and a data call termination may occur simultaneously. There are (at least) two ways to manage this situation. One strategy is to assume that an active data call is always guaranteed a waiting space. This is done by choosing a data call in the waiting room for termination (strategy I) and then placing the previously active data call in the vacated space. Another strategy is to randomly choose a data call in the cell for termination, regardless of whether the call is active or waiting (strategy II). The relation between predecessor and successor states are different for each strategy.

Figure 2 shows an example of the state transition with different strategies. The resources (channels and waiting spaces) of a cell are fully occupied by calls, whether voice, active data, or waiting data calls. If a voice call arrives when a cell is in the state indexed k (shown in the Figure 2), an active data call must be suspended. Let us suppose that a $g = 1$ type of data call is suspended. When the system uses strategy I (a prioritized scheme), the data call just suspended will receive a waiting space, so a *waiting* data call must be terminated. The state indexed k is the only possible predecessor state of the state indexed l for a new voice call arrival on a $g = 1$ type platform. However, for strategy II (a non-prioritized scheme), the data call that is ejected to accommodate a voice call arrival can be either an active or a suspended data call. Therefore, for this same arrival event the state indexed k can be a possible predecessor state not only of the state indexed l, but also of the

12

state indexed m. These two cases are analyzed in Discussion of Results.

Given that the system is in state s, the conditional probability, $P(s,g)$, that a data call held *on a g-type platform* will be reactivated when a channel becomes available is given by

$$P(s,g) = \begin{cases} w(s,g)/w(s) & \text{if } s \in R_2 \cup R_3 \cup B_2 \cup B_3 \cup B_4 \cup B_5 \\ 0 & \text{otherwise.} \end{cases} \tag{18}$$

If there are no waiting spaces for data calls held, $P(s,g)$ is always zero regardless of cell state.

When the system is in state, s, the probability, $Q_s(s,g)$, that an active data call *on a g-type platform* will be suspended when a voice call arrives is given by

$$Q_s(s,g) = \begin{cases} 0 & \text{if } s \in R_1 \cup R_4 \\ z(s,g)/z(s) & \text{otherwise.} \end{cases} \tag{19}$$

When the waiting spaces for data calls are full, an arriving voice call, whether new or hand-off, causes a termination of a data call. Which data call is terminated is determined by the strategy.

For strategy I (a prioritized scheme), only a suspended data call will be chosen for termination when the waiting spaces are full and a voice call arrives. Therefore, for strategy I, the probability that a data call which is held *on a g-type platform* will be terminated is given by

$$Q_t(s,g) = \begin{cases} w(s,g)/w(s) & \text{if } s \in B_2 \cup B_4 \\ 0 & \text{otherwise.} \end{cases} \tag{20}$$

In strategy II (a non-prioritized scheme), not only a *suspended* data call but also an *active* data call can be terminated when the waiting spaces are full and a voice call arrives. There are two possible scenarios. The first one is that the system chooses a held data call and a terminated data call from the same type of mobile platform. For instance, an active data call on g-type of mobile platform is chosen for suspension and a data call on g-type of mobile platform is, also, chosen for termination. For this scenario, the probability that a data call *on a g-type platform* will be terminated is given by

$$Q_t(s,g) = \begin{cases} \{w(s,g)+1\}/\{w(s)+1\} & \text{if } s \in B_2 \cup B_4 \\ 0 & \text{otherwise.} \end{cases} \tag{21}$$

13

The other is that the system chooses a held data call and a terminated data call from different types of mobile platforms. For instance, an active data call on l-type of mobile platform is chosen for suspension and a data call on g-type ($g \neq l$) of mobile platform is chosen for termination. For this scenario, the probability that a data call *on a g-type platform* will be terminated is given by

$$Q_t(s,g) = \begin{cases} w(s,g)/\{w(s)+1\} & \text{if } s \in B_2 \cup B_4 \\ 0 & \text{otherwise.} \end{cases} \tag{22}$$

If there are no waiting spaces for data calls held, $Q_t(s,g) = Q_s(s,g)$, given in equation (19).

7. DRIVING PROCESSES

There are eight possible driving processes. Each process has G-dimensions because G different types of mobile platforms coexist in a cell. We use Markovian assumptions for the driving processes to render the problem amenable to solution using multidimensional birth-death processes. Each process is listed below

- $\{nv\}$: generation of new voice calls

- $\{nw\}$: generation of new data calls

- $\{cv\}$: completion of voice calls

- $\{cw\}$: completion of data calls

- $\{hv\}$: hand-off arrival of voice calls

- $\{hw\}$: hand-off arrival of data calls

- $\{dv\}$: hand-off departure of voice calls

- $\{dw\}$: hand-off departure of data calls

For convenience, dummy variable k is used as follows: $k = 1, 2, \ldots \ldots, G$.

14

7.1. new voice call arrivals

In the region R_1, a transition into state s, due to a new voice call arrival on a g-type platform when the cell is in state x_{nv}, will cause the state variable $v(x_{nv}, g)$ to be incremented by 1. Thus a permissible state x_{nv} is a predecessor state of s for a new voice call arrival on g-type platform, if the state variables are related by

$$
\begin{aligned}
v(x_{nv}, g) &= v(s, g) - 1 \\
v(x_{nv}, k) &= v(s, k) \quad , \quad k \neq g \\
z(x_{nv}, k) &= z(s, k) \\
w(x_{nv}, k) &= w(s, k)
\end{aligned}
\tag{23}
$$

and $s \in R_1$.

If a new voice call arrives on a g-type platform when the system is in region R_2 or B_1, an active data call must be suspended in order to serve the higher priority voice call. Therefore, the system will choose an active data call for suspension with probability $Q_s(s, g)$. Suppose a data call which is on l-type platform is suspended due to an arrival of voice call on g-type platform, then the state variables are related by

$$
\begin{aligned}
v(x_{nv}, g) &= v(s, g) - 1 \\
v(x_{nv}, k) &= v(s, k) \quad , \quad k \neq g \\
z(x_{nv}, l) &= z(s, l) + 1 \\
z(x_{nv}, k) &= z(s, k) \quad , \quad k \neq l \\
w(x_{nv}, l) &= w(s, l) - 1 \\
w(x_{nv}, k) &= w(s, k) \quad , \quad k \neq l
\end{aligned}
\tag{24}
$$

and $s \in R_2 \cup B_1$.

When the system is state on the boundary B_2, a new voice call will get a channel and a data call will be terminated. Suppose a data call which is on l-type platform is suspended and a data call on

15

m-type platform is terminated due to an arrival of voice call on g-type mobility platform. then the state variables are related by

$(i) m \neq l$

$$
\begin{aligned}
v(x_{nv}, g) &= v(s, g) - 1 \\
v(x_{nv}, k) &= v(s, k) \quad , \quad k \neq g \\
z(x_{nv}, l) &= z(s, l) + 1 \\
z(x_{nv}, k) &= z(s, k) \quad , \quad k \neq l \\
w(x_{nv}, l) &= w(s, l) - 1 \\
w(x_{nv}, m) &= w(s, m) + 1 \\
w(x_{nv}, k) &= w(s, k) \quad , \quad k \neq l, m
\end{aligned}
$$

$(ii) m = l$

$$
\begin{aligned}
v(x_{nv}, g) &= v(s, g) - 1 \\
v(x_{nv}, k) &= v(s, k) \quad , \quad k \neq g \\
z(x_{nv}, l) &= z(s, l) + 1 \\
z(x_{nv}, k) &= z(s, k) \quad , \quad k \neq l \\
w(x_{nv}, k) &= w(s, k)
\end{aligned}
\tag{25}
$$

and $s \in B_2$.

Let $\Lambda_{nv}(g)$ denote the average arrival rate per cell of new voice calls from g-type platforms. Then, the corresponding transition flow is given by

$$
\gamma_{nv}(s, x_{nv}, g) = \begin{cases}
\Lambda_{nv}(g) & s \in R_1 \\
Q_s(s, g) \times \Lambda_{nv}(g) & s \in R_2 \cup B_1 \\
Q_s(s, g) \times Q_t(s, g) \times \Lambda_{nv}(g) & s \in B_2.
\end{cases}
\tag{26}
$$

16

7.2. new data call arrivals

In the region R_1 or R_4, a transition into state s, due to a data call arrival on a g-type platform when the cell is state x_{nw}, will cause the state variable $z(x_{nw}, g)$ to incremented by 1. Thus a permissible state x_{nw} is a predecessor state of s for a data call arrival on g-type platform, if the state variables are related by

$$v(x_{nw}, k) = v(s, g)$$
$$z(x_{nw}, g) = z(s, k) - 1$$
$$z(x_{nw}, k) = z(s, k) \quad , \quad k \neq g$$
$$w(x_{nw}, k) = w(s, k) \tag{27}$$

and $s \in R_1 \cup R_4$.

Even though all channels in the system are occupied by voice or data calls, an arriving data call can wait for its transmission opportunity if a waiting space in the system is available. That event happens when the system is in R_2, R_3, B_1, B_3, B_6, or B_7, and a new data call arrives on a g-type platform. Therefore, x_{nw} is a predecessor state of s if

$$v(x_{nw}, k) = v(s, k)$$
$$z(x_{nw}, k) = z(s, k)$$
$$w(x_{nw}, g) = w(s, g) - 1$$
$$w(x_{nw}, k) = w(s, k) \quad , \quad k \neq g \tag{28}$$

and $s \in R_2 \cup R_3 \cup B_1 \cup B_3 \cup B_6 \cup B_7$.

When the system is in B_2, B_4, or B_5, an arriving data call will be blocked due to the lack of waiting spaces. Let $\Lambda_{nw}(g)$ denoted the average arrival rate per cell of new data calls from g-type platform. Then the transition flow into state s from x_{nw} due to a new data call arrival is given by

$$\gamma_{nw}(s, x_{nw}, g) = \Lambda_{nw}(g) \qquad s \in R_1 \cup R_2 \cup R_3 \cup R_4 \cup B_1 \cup B_3 \cup B_6 \cup B_7. \tag{29}$$

Similar descriptions of successor-predecessor relationships among the states and of the related state transition rates are given for each of the driving processes in Appendix B (Sections 7.3-7.8).

17

8. FLOW BALANCE EQUATIONS

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

$$q(s,x) = \gamma_{nv}(s,x) + \gamma_{nw}(s,x) + \gamma_{cv}(s,x) + \gamma_{cw}(s,x) + \gamma_{hv}(s,x) + \gamma_{hw}(s,x) + \gamma_{dv}(s,x) + \gamma_{dw}(s,x) \quad (30)$$

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

$$q(s,s) = - \sum_{\substack{k=0 \\ k\neq s}}^{S_{max}} q(k,s). \quad (31)$$

The statistical equilibrium can be found using the flow balance equations that are set of $S_{max}+1$ simultaneous equations for the unknown state probabilities,

$$\sum_{j=0}^{S_{max}} q(i,j) \times p(j) \;=\; 0, i = 0, 1, \ldots\ldots, S_{max} - 1 \quad (32)$$

$$\sum_{j=0}^{S_{max}} p(j) \;=\; 1 \quad (33)$$

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.

9. THE HAND-OFF ARRIVAL PARAMETERS

The average hand-off arrival rate of voice call, Λ_{hv}, that of data call, Λ_{hw}, the fraction of voice call hand-off arrivals that are g-type platform, F_{vg}, and the fraction of data call hand-off arrivals that are g-type platform, F_{wg}, can be determined from the dynamics of the process itself. An iterative method can be used [5]. The average hand-off departure rate of voice calls on g-type platforms can be expressed as

$$\Delta_{hv}(g) = \sum_{s=0}^{S_{max}} \mu_D(g) \times v(s,g) \times p(s). \quad (34)$$

Thereafter, the overall average hand-off departure rate of voice call can be written as

$$\Delta_{hv} = \sum_{g=1}^{G} \Delta_{hv}(g). \quad (35)$$

18

The average hand-off departure rate of data call on a g-type platform can be expressed as

$$\Delta_{hw}(g) = \sum_{s=0}^{Smax} \mu_D(g) \times [z(s,g) + w(s,g)] \times p(s). \tag{36}$$

Also, the overall average hand-off departure rate of data call can be written as

$$\Delta_{hw} = \sum_{g=1}^{G} \Delta_{hw}(g). \tag{37}$$

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

$$F'_{vg} = \Delta_{hv}(g)/\Delta_{hv}, \tag{38}$$

and, the fraction of data call hand-off departures on g-type mobility platform is

$$F'_{wg} = \Delta_{hw}(g)/\Delta_{hw}. \tag{39}$$

It should be noted that any hand-off departure of a g-type platform from a cell corresponds to a hand-off arrival of a g-type call to another cell. Therefore, the hand-off arrival and departure rates per cell for a homogeneous system in statistical equilibrium must be equal. That is

$$F_{vg} = F'_{vg} \tag{40}$$

$$F_{wg} = F'_{wg} \tag{41}$$

$$\Lambda_{hv} = \Delta_{hv} \tag{42}$$

$$\Lambda_{hw} = \Delta_{hw}. \tag{43}$$

10. PERFORMANCE MEASURES

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

10.1. voice call performance measures

10.1.1. blocking probability

The blocking probability for voice calls regardless of the platform type on which it generated is the average fraction of new voice calls that are denied access to a channel. A new voice call will be

19

blocked when the state is in region $R_3, R_4, B_3, B_4, B_5, B_6$, or B_7. Therefore. the blocking probability for voice calls is given by

$$P_{Bv} = \sum_{s \in R_3 \cup R_4 \cup B_3 \cup B_4 \cup B_5 \cup B_6 \cup B_7} p(s). \tag{44}$$

10.1.2. hand-off failure probability

The hand-off failure probability for voice calls regardless of the platform type is the average fraction of hand-off attempts that are denied a channel in the target cell. An arriving hand-off call is blocked when the state of target cell is in B_3, B_5, or B_7. Therefore, the hand-off failure probability of voice calls is expressed as

$$P_{Hv} = \sum_{s \in B_3 \cup B_5 \cup B_7} p(s). \tag{45}$$

10.1.3. forced termination probability

Perhaps, the more significant factor than the blocking probability from user's point of view is the forced termination probability, $P_{FTv}(g)$. This is defined as the probability that g-type voice call that is not blocked is interrupted due to hand-off failure during its life time as in [4], [5], [7]. It can be shown that the forced termination probability of g-type voice call is given by

$$P_{FTv}(g) = \frac{\mu_D(g) \times P_{Hv}}{\mu_v(g) + \mu_D(g) \times P_{Hv}}. \tag{46}$$

10.2. data call performance measures

10.2.1. blocking probability and hand-off failure probability

The blocking probability for data calls regardless of the platform type on which it generated is the average fraction of new data calls that are denied a channel or a waiting space in the cell due to the lack of capacity. The event of data call blocking occurs when the system is in region B_2, B_4. or B_5. And, it can be written as

$$P_{Bw} = \sum_{s \in B_2 \cup B_4 \cup B_5} p(s). \tag{47}$$

When the state of target cell is in B_2, B_4, or B_5, a hand-off attempt of a data call regardless of the platform type is denied a channel or waiting room space in the target cell. Therefore, the

20

hand-off failure probability of data call is the same as the blocking probability of data call since there is no channel or waiting spaces quotas for hand-off data calls. And it can be written as

$$P_{Hw} = \sum_{s \in B_2 \cup B_4 \cup B_5} p(s). \tag{48}$$

10.2.2. forced termination probability

The forced termination probability, $P_{FTw}(g)$, is defined as the probability that a g-type data call that is not blocked is interrupted due to hand-off failure during its life time. It can be shown that this is given by

$$P_{FTw}(g) = \frac{\mu_D(g) \times P_{Hw}}{\mu_w(g) + \mu_D(g) \times P_{Hw}}. \tag{49}$$

10.2.3. pushed-out failure probability

It should be noted that there is another cause of data call terminations. This is the termination of a data call due to *a voice call arrival* when the waiting spaces are full. We call such an event a pushed-out because the data call that is terminated in this case is pushed out of a finite length queue. For the sake of consistency of previous nomenclature of voice call forced termination probability, we limit the meaning of forced termination probability of data calls as the termination due *hand-off failure* during its life time. The pushed-out failure probability, $P_{po}(g)$, is defined as the average fraction of data calls on g-type of mobile platform that are terminated due to an arrival of a voice call. The event of a data call pushed-out failure occurs when the system is in region B_2 or B_4. This can be written as

$$P_{po}(g) = Q_t(s,g) \times \sum_{s \in B_2 \cup B_4} p(s). \tag{50}$$

10.2.4. average time a data call spends in the system

After a data call arrives, its transmission will be controlled by the system. To calculate the time that a data call spends in the system, $S_D(g)$, Little's law is used. Firstly, we will determine the average rate of accomodating data calls, $A_D(g)$. Since some part of data calls will be blocked due to the lack of system capacity, the average rate of accomodating data calls, $A_D(g)$, can be expressed as

$$A_D(g) = \Lambda_w(g) \times (1 - P_{Bw}). \tag{51}$$

21

And we can determine the average number of data calls, whether it is active or waiting, from the state probabilities as follows

$$B_D(g) = \sum_{s=0}^{Smax} [z(s,g) + w(s,g)] \times p(s). \tag{52}$$

Based on the average rate of accomodating data calls, $A_D(g)$, and the average number of data calls, $B_D(g)$, we can calculate the average time a data call spends in the system as follows

$$S_D(g) = B_D(g)/A_D(g). \tag{53}$$

10.2.5. average waiting time for those data calls that wait

To calculate average waiting time, consider calls that *enter* the "waiting" room. We will determine the average rate of such entries and the average number of waiting calls from the state probabilities. Little's law will then be applied to find the average waiting time of those calls that wait.

Even though all channels are full, an arriving data call will be held by system if waiting space is available. This is the case when the system is in R_2, R_3, B_1, B_3, B_6, or B_7. When the system is in regions R_2 or B_1, an arriving voice call, whether new or hand-off, causes an active data call suspension. The system must choose one active data call to suspend randomly, the probability that an active data call from a g-type platform will be suspended due to this transition is $Q_s(s,g)$. When the cell is in R_3 or B_6, a new voice call will be blocked but an arriving hand-off call can occupy a channel with suspension of an active data call. The system chooses a data call to be suspended at random and the probability that an active data call from a g-type platform will be suspended due to this transition is $Q_s(s,g)$. On the other hand, when the cell is in B_3 or B_7, an arriving voice call, either new or hand-off, will be blocked. So only arriving data calls will be held in the waiting room. Therefore, the average rate of data call held for a g-type platform, $H_D(g)$, is given by

$$
\begin{aligned}
H_D(g) &= \sum_{s \in R_2 \cup B_1} \{\{\Lambda_{nw}(g) + [\Lambda_{hw} \times F_{wg}]\} + \{\Lambda_{nv} + \Lambda_{hv}\} \times Q_s(s,g)\} \times p(s) \\
&+ \sum_{s \in R_3 \cup B_6} \{\{\Lambda_{nw}(g) + [\Lambda_{hw} \times F_{wg}]\} + \Lambda_{hv} \times Q_s(s,g)\} \times p(s)
\end{aligned}
$$

22

$$+ \sum_{s \in B_3 \cup B_7} \{\Lambda_{nw}(g) + [\Lambda_{hw} \times F_{wg}]\} \times p(s) \qquad (54)$$

$$(55)$$

where $\Lambda_{nv} = \sum_{g=1}^{G} \Lambda_{nv}(g)$.

Secondly, we can simply calculate the average number of data call held for a g-type platform, $N_D(g)$. It is written by

$$N_D(g) = \sum_{s \in B_1 \cup R_2 \cup B_2 \cup R_3 \cup B_3 \cup B_4 \cup B_5 \cup B_6} w(s,g) \times p(s) \qquad (56)$$

Finally, we can find the average waiting time for data calls on a g-type platform that must wait. $W_D(g)$. Using Little's law, this is

$$W_D(g) = N_D(g)/H_D(g). \qquad (57)$$

The average waiting time of a data call for a g-type platform, $W_D(g)$, is defined to be the average amount of time a data call on g-type platform spends in the waiting room space before entering service, being forced to terminate, or being pushed out.

11. DISCUSSION OF RESULTS

Numerical results were generated using the approach described in this paper. For all figures, an unencumbered voice call duration of 100s was assumed and an unencumbered data call duration of 20s was assumed. Two platform types, low mobility and high mobility, were considered. The mean dwell time of 500s was assumed for a low mobility platform and that of 100s was assumed for a high mobility platform. A homogeneous system was assumed. It is, also, assumed that there are 15 channels in the system and 300 noncommunicating g-type platforms exist in a cell.

The abscissas for Figures [3]-[9] reflect the call demands with the assumptions stated above. In these, the abscissa is the new voice call origination rate for platform type 1 (denoted $\Lambda_v(1)$). The ratio of new voice call generation rates from other platform types were held with respect to type 1 with parameters $\alpha(g)$. Also, the new data call generation rate for platform type g is determined with

23

respect to new voice call origination rate using parameters, $\beta(g)$. For all calculations in this paper. $\alpha(g)=\beta(g)=1$ is assumed. Figures [3]-[8] were generated with the assumption that the system uses strategy I for data call management.

Figure 3 shows blocking and forced termination probability of voice call as a function of new voice call arrival rate on type 1 platform. Since the data call traffic is transparent to voice call users. the traffic performances of voice call with data traffic is identical to those of without data traffic. As the number of reserved channels, C_h, increase, fewer voice hand-off calls are forced to terminate but more new voice call will be blocked. Clearly, calls on fast mobile platforms have higher forced termination probability than that on slow mobile platform. This is because the calls on fast mobile platforms are likely to experience more hand-offs during a call lifetime.

Figure 4 shows the blocking probability of data calls. It is seen that the more channels are reserved for voice hand-off calls, the less data calls are blocked. This is because the more channels are reserved for voice hand-off calls, the more new voice call arrived are blocked, and, consequently. the opportunity to transmit a data call is increased.

Figure 5 shows the forced termination probability of data calls. It is seen that if more channels are reserved for voice hand-off calls, fewer data calls are forced to terminate. This is because more new voice calls will be blocked, and more *data hand-off calls* can be accommodated by channels. It can also be seen that calls on fast mobile platforms have higher forced termination probability than that on slow mobile platform. This is because the calls on fast mobile platforms are likely to experience more hand-offs during a call lifetime.

Figure 6 shows the pushed-out failure probability of data calls. As more channels are reserved for *voice hand-off calls*, more *new voice calls* will be blocked, and fewer data calls will be pushed out due to the arrival of voice call. It is also seen that the pushed-out failure probability for slow mobile platform is always higher than that for fast mobile platform. This is because the data calls from fast mobile platforms tend to spend less time in the waiting room than those from slow mobile platforms. The waiting room tends to be populated by data calls on slow mobile platforms. On the

24

other hand, more data calls from fast mobile platform will be forced to terminate due to the failure of hand-off during its lifetime.

Figure 7 shows the average waiting time of a data call that must wait. It is seen that the expected waiting time is relatively insensitive to the platform mobility. As more channels are reserved for voice hand-off call, the less waiting time of data call is expected. Figure 8 shows the expected waiting time of a data call as the number of waiting spaces is changed. It is shown that there is no recognizable difference of system performance when the call demands are relatively low. However, as call demands increase, we see a clear difference of system performance between relatively small and large size of waiting rooms. Figure 9 shows the comparison of the expected waiting time using two different data call management strategies. As we see, the performance of data call traffic is insensitive to those strategies.

Figure 10 shows the average time a data call spends in the system for fast and slow mobiles. Since calls on board fast mobiles experience relatively more hand-offs than those on board slow mobiles, they are more likely to terminate due to hand-off failure. This is reflected in the shorter average time in the system for fast mobiles. It is also seen that when the call demands are relatively low, reserving channels for voice hand-off calls has little effect on the average time spent in the system by calls on either slow or fast mobiles. However, as call demand increases, the number of channels (C_h) reserved for voice hand-off calls has a greater influence on the average time spent in the system for calls on both fast and slow mobiles. As more channels are reserved for *voice hand-off calls* (C_h increases), more *new* voice calls will be blocked and voice call traffic decreases. Since data calls have access to those reserved channels, the amount of time a data call spends in waiting room will decrease. So, the average time a data call spends in the system decreases.

12. CONCLUSION

The framework that we are developing using a state description and multidimensional birth-death processes can be used to compute theoretical traffic performance characteristics for cellular commu-

25

nication systems that support voice calls and background data calls with mixed platform types. The different mobility is modeled as the different dwell time characteristics and the different call type, voice call and background data call, is modeled as the different call session time characteristics. Traffic performance of data call are dependent on the number of reserved channels for voice hand-off calls. The forced-termination probability of data call is dependent on the mean of the dwell time that a mobile platform spends within communications range of a wireless gateway. Two possible strategies for the management of data calls are considered. The traffic performances according to each strategy are calculated and shown to be insensitive to the strategy. Cut-off priority for voice hand-off calls in cellular systems allows a very favorable exchange between blocking and forced termination of voice call performances. Furthermore, it provides improved data call traffic performances such as reduced blocking probability, forced-termination probability, pushed-out termination probability, and average waiting time.

13. REFERENCES

[1] W. D. Grover, W. A. Krzymien, J. C. Chin, "Background Data Service for Cellular Mobile Radio," in *Proc. 43rd IEEE Vehic. Technol. Conf.*, pp. 766-769, 1993.

[2] "CDPD Specification version 1.0" Pittiglio Rabin Todd & McGrath Co., Costa Mesa, CA, July 1993.

[3] W. D. Grover, W. A. Krzymien, J. C. Chin, "Strategies for Management of Connection-Oriented Data Sessions Employing the idle Times in Cellular Telephony," *IEEE Trans. on Vehic. Technol.*, vol. 44, no. 2, May 1995.

[4] S. S. Rappaport, "The Multiple-Cell Hand-off Problem in High-Capacity Cellular Communications System" in *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 Systems with Mixed Platforms", in *IEE(British) Proceedings, Part I, Communications, Speech and Vision*, October 1993, vol. 140, no. 5, pp 389-401.

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

[7] 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" in *IEEE Trans. on Vehicular Technology*, August 1986, vol. VT-35, no 3, pp. 77-92.

14. APPENDIX

A. COMPUTATIONAL PROCEDURE

The quantities, $F'_{vg}, F'_{wg}, \Delta_{hv}$, and Δ_{hw} are initially guessed. After solving the flow balance equations, the iterative method[4] is used for obtaining the unknown state probabilities $p(s)$. It should be noted that a modified Gauss-Seidel algorithm was used to solve the flow balance equations. The scheme is similar to that given in [4].

B. DRIVING PROCESSES [SECTION 7.3 - 7.8]

7.3 voice call completions

In the region R_1, R_4, B_1, B_6, or B_7, a transition into state s due to a voice call completion on a g-type platform when the cell is in state x_{cv}, will simply cause the state variable $v(x_{cv}, g)$ to be decreased by 1. Thus a permissible state x_{cv} is a predecessor state of s for a voice call completion on g-type platforms, if the state variables are related by

$$
\begin{aligned}
v(x_{cv}, g) &= v(s, g) + 1 \\
v(x_{cv}, k) &= v(s, k) \quad , \quad k \neq g \\
z(x_{cv}, k) &= z(s, k) \\
w(x_{cv}, k) &= w(s, k)
\end{aligned}
\tag{58}
$$

and $s \in R_1 \cup R_4 \cup B_1 \cup B_6 \cup B_7$.

When the system is in R_2, R_3, B_2, B_3, B_4, or B_5, a data call held will be reactivated if a channel becomes available due to a completion of a voice call. The system chooses a data call for reactivation with probability $P(s, g)$. Also, there exist G permissible predecessor states, x_{cv}, based upon the platform mobility type of the reactivated data call. Suppose a data call on l-type platform is reactivated due to a completion of voice call on a g-type of mobile platform, then the state variables are related by

$$
v(x_{cv}, g) = v(s, g) + 1
$$

27

$$v(x_{cv}, k) = v(s, k) \quad , \quad k \neq g$$

$$z(x_{cv}, l) = z(s, l) - 1$$

$$z(x_{cv}, k) = z(s, k) \quad , \quad k \neq l$$

$$w(x_{cv}, l) = w(s, l) + 1$$

$$w(x_{cv}, k) = w(s, k) \quad , \quad k \neq l \tag{59}$$

and $s \in R_2 \cup R_3 \cup B_2 \cup B_3 \cup B_4 \cup B_5$.

The unencumbered voice call duration on a g-type platform is a ned random variable. $T_v(g)$. having a mean $\overline{T}_v(g) = 1/\mu_v(g)$. Then, the transition flow into state s from x_{cv} due to a voice call completion is given by

$$\gamma_{cv}(s, x_{cv}, g) = \begin{cases} \mu_v(g) \times v(x_{cv}, g) & s \in R_1 \cup R_4 \cup B_1 \cup B_6 \cup B_7 \\ P(s, g) \times \mu_v(g) \times v(x_{cv}, g) & s \in R_2 \cup R_3 \cup B_2 \cup B_3 \cup B_4 \cup B_5. \end{cases} \tag{60}$$

7.4 data call completions

In the region R_1, R_4, B_1, or B_6, a transition into state s, due to a data call completion on a g-type platform when the cell is in state x_{cw}, will simply cause the state variable $z(x_{cw}, g)$ to be decreased by 1. Thus a permissible state x_{cw} is a predecessor state of s for data call completion on g-type platforms, if the state variables are related by

$$v(x_{cw}, k) = v(s, k)$$

$$z(x_{cw}, g) = z(s, g) + 1$$

$$z(x_{cw}, k) = z(s, k) \quad , \quad k \neq g$$

$$w(x_{cw}, k) = w(s, k) \tag{61}$$

and $s \in R_1 \cup R_4 \cup B_1 \cup B_6$.

If a data call completes on a g-type platform when the system is in R_2, R_3, B_2, B_3, B_4, or B_5. the system may choose a data call held for reactivating. Suppose a data call on l-type platform is reactivated due to the completion of a data call on g-type of mobile platform, then state variables

28

are related by

$(i) g = l$

$$
\begin{aligned}
v(x_{cw}, k) &= v(s, k) \\
z(x_{cw}, k) &= z(s, k) \\
w(x_{cw}, l) &= w(s, l) + 1 \\
w(x_{cw}, k) &= w(s, k) \quad , \quad k \neq l
\end{aligned}
$$

$(ii) g \neq l$

$$
\begin{aligned}
v(x_{cw}, k) &= v(s, k) \\
z(x_{cw}, g) &= z(s, g) + 1 \\
z(x_{cw}, l) &= z(s, l) - 1 \\
z(v_{cw}, k) &= z(s, k) \quad , \quad k \neq g, l \\
w(x_{cw}, l) &= w(s, l) + 1 \\
w(x_{cw}, k) &= w(s, k) \quad , \quad k \neq l,
\end{aligned}
\tag{62}
$$

and $s \in R_2 \cup R_3 \cup B_2 \cup B_3 \cup B_4 \cup B_5$.

The unencumbered data transmission duration on a g-type platform is a ned random variable, $T_w(g)$, having a mean $\overline{T}_w(g) = 1/\mu_w(g)$, Then the transition flow into state s from x_{cw} due to a data call completion is given by

$$
\gamma_{cw}(s, x_{cw}, g) = \begin{cases} \mu_w(g) \times z(x_{cw}, g) & s \in R_1 \cup R_4 \cup B_1 \cup B_6 \\ P(s, g) \times \mu_w(g) \times z(x_{cw}, g) & s \in R_2 \cup R_3 \cup B_2 \cup B_3 \cup B_4 \cup B_5. \end{cases}
\tag{63}
$$

7.5 voice hand-off call arrivals

In the region R_1 or R_4, a transition into state s, due to a voice hand-off call arrival on a g-type platform when the cell is in state x_{hv}, will simply cause the state $v(x_{hv}, g)$ to be incremented by 1.

29

Thus a permissible state x_{hv} is a predecessor state of s for a voice hand-off call arrival on g-type platform, if the state variables are related by

$$
\begin{aligned}
v(x_{hv}, g) &= v(s, g) - 1 \\
v(x_{hv}, k) &= v(s, k) \quad, \quad k \neq g \\
z(x_{hv}, k) &= z(s, k) \\
w(x_{hv}, k) &= w(s, k)
\end{aligned}
\tag{64}
$$

and $s \in R_1 \cup R_4$.

If a voice hand-off call arrives on a g-type platform when the cell is in a state in R_2, R_3, B_1. or B_6. one data call will be suspended in order to serve the higher priority voice hand-off call. Let us suppose that a data call which is on l-type platform is suspended due to an arrival of a voice hand-off call on g-type of mobile platform, then the state variables are related by

$$
\begin{aligned}
v(x_{hv}, g) &= v(s, g) - 1 \\
v(x_{hv}, k) &= v(s, k) \quad, \quad k \neq g \\
z(x_{hv}, l) &= z(s, l) + 1 \\
z(x_{hv}, k) &= z(s, k) \quad, \quad k \neq l \\
w(x_{hv}, l) &= w(s, l) - 1 \\
w(x_{hv}, k) &= w(s, k) \quad, \quad k \neq l
\end{aligned}
\tag{65}
$$

and $s \in R_2 \cup R_3 \cup B_1 \cup B_6$.

When the system is in the regions B_2 or B_4, a voice hand-off c call arrival causes an active data call suspension and a held data call termination at the same time. Suppose that a data call which is on an l-type platform is suspended and a data call on an m-type platform is terminated due to an arrival of voice hand-off call on g-type of mobile platform, then the state variables are related by

$(i) m \neq l$

$$
v(x_{hv}, g) = v(s, g) - 1
$$

30

$$v(x_{hv}, k) = v(s, k) \quad , \quad k \neq g$$

$$z(x_{hv}, l) = z(s, l) + 1$$

$$z(x_{hv}, k) = z(s, k) \quad , \quad k \neq l$$

$$w(x_{hv}, l) = w(s, l) - 1$$

$$w(x_{hv}, m) = w(s, m) + 1$$

$$w(x_{hv}, k) = w(s, k) \quad . \quad k \neq l, m$$

$(ii) m = l$

$$v(x_{hv}, g) = v(s, g) - 1$$

$$v(x_{hv}, k) = v(s, k) \quad , \quad k \neq g$$

$$z(x_{hv}, l) = z(s, l) + 1$$

$$z(x_{hv}, k) = z(s, k) \quad , \quad k \neq l$$

$$w(x_{hv}, k) = w(s, k) \tag{66}$$

and $s \in B_2 \cup B_4$.

Let Λ_{hv} be the average rate at which voice hand-off call arrivals impinge on the cell, and F_{vg} denote the fraction of voice call hand-off arrivals that are from g-type platform. Λ_{hv} and F_g can be determined from the dynamics of the process itself using an iterative method. Then the transition flow into state s from x_{hv} due to a voice hhand-off call arrival is given by

$$\gamma_{hv}(s, x_{hv}, g) = \begin{cases} \Lambda_{hv} \times F_{vg} & s \in R_1 \cup R_4 \\ Q_s(s, g) \times \Lambda_{hv} \times F_{vg} & s \in R_2 \cup R_3 \cup B_1 \cup B_6 \\ Q_s(s, g) \times Q_t(s, g) \times \Lambda_{hv} \times F_{vg} & s \in B_2 \cup B_4. \end{cases} \tag{67}$$

7.6 data hand-off call arrivals

In the region R_1, or R_4, a transition into state s, due to a data hand-off call arrival on a g-type platform when the cell is state x_{hw}, will cause the state variable $z(x_{hw}, g)$ to incremented by 1. Thus a permissible state x_{hw} is a predecessor state of s for a data hand-off call arrival on g-type platform.

31

if the state variables are related by

$$
\begin{aligned}
v(x_{hw}, k) &= v(s, k) \\
z(x_{hw}, g) &= z(s, g) - 1 \\
z(x_{hw}, k) &= z(s, k) \quad , \quad k \neq g \\
w(x_{hw}, k) &= w(s, k)
\end{aligned}
\tag{68}
$$

and $s \in R_1 \cup R_4$.

Even though all the channels in the target system are busy, an arriving hand-off data call can wait for its transmission opportunity if a waiting space in the system is available. That event happens when the system is in R_2, R_3, B_1, B_3, B_6, or B_7. Therefore, a hand-off data call which arrives on a g-type platform when the system is in one of those regions makes the state variables be related by

$$
\begin{aligned}
v(x_{hw}, k) &= v(s, k) \\
z(x_{hw}, k) &= z(s, k) \\
w(x_{hw}, g) &= w(s, g) - 1 \\
w(x_{hw}, k) &= w(s, k) \quad , \quad k \neq g
\end{aligned}
\tag{69}
$$

and $s \in R_2 \cup R_3 \cup B_1 \cup B_3 \cup B_6 \cup B_7$.

When the system is in B_2, B_4, or B_5, an arriving hand-off data call will be blocked due to the lack of waiting spaces. Let Λ_{hw} be the average rate at which data hand-off call arrivals impinge on the cell, and F_{wg} denote the fraction of data call hand-off arrivals that are from g-type platform.. Then the transition flow into state s from x_{hw} due to a hand-off data call arrival is given by

$$
\gamma_{hw}(s, x_{hw}, g) = \Lambda_{hw} \times F_{wg} \qquad\qquad s \in R_1 \cup R_2 \cup R_3 \cup R_4 \cup B_1 \cup B_3 \cup B_6 \cup B_7.
\tag{70}
$$

7.7 voice hand-off call departures

In the region R_1, R_4, B_1, B_6, or B_7, a transition into state s due to a voice hand-off call departure on a g-type platform when the cell is in state x_{dv}, will cause the state variables $v(x_{dv}, g)$

32

to be decreased by 1. Thus a permissible state x_{dv} is a predecessor state of s for voice hand-off call departure on a g-type platform, if the state variables are related by

$$
\begin{aligned}
v(x_{dv}, g) &= v(s, g) + 1 \\
v(x_{dv}, k) &= v(s, k) \quad , \quad k \neq g \\
z(x_{dv}, k) &= z(s, k) \\
w(x_{dw}, k) &= w(s, k)
\end{aligned}
\tag{71}
$$

and $s \in R_1 \cup R_4 \cup B_1 \cup B_6 \cup B_7$.

If a channel is available due to a voice call hand-off departure when the system is in R_2, R_3, B_2, B_3, B_4, or B_5, one data call will be chosen for reactivation. The system will operate by choosing with probabilities $P(s, g)$ for service. Therefore, there exist G permissible predecessor states, x_{dv}, based upon the type of mobile platform of reactivating data call. Suppose that a data call which is on l-type platform is reactivated due a departure of a voice hand-off call on g-type platform. Then, the state variables are related by

$$
\begin{aligned}
v(x_{dv}, g) &= v(s, g) + 1 \\
v(x_{dv}, k) &= v(s, k) \quad , \quad k \neq g \\
z(x_{dv}, l) &= z(s, l) - 1 \\
z(x_{dv}, k) &= z(s, k) \quad , \quad k \neq l \\
w(x_{dv}, l) &= w(s, l) + 1 \\
w(x_{dv}, k) &= w(s, k) \quad , \quad k \neq l
\end{aligned}
\tag{72}
$$

and $s \in R_2 \cup R_3 \cup B_2 \cup B_3 \cup B_4 \cup B_5$.

Therefore the corresponding transition flow is given by

$$
\gamma_{dv}(s, x_{dv}, g) = \begin{cases} \mu_D(g) \times v(x_{dv}, g) & s \in R_1 \cup R_4 \cup B_1 \cup B_6 \cup B_7 \\ P(s, g) \times \mu_D(g) \times v(x_{dv}, g) & s \in R_2 \cup R_3 \cup B_2 \cup B_3 \cup B_4 \cup B_5 \ . \end{cases}
\tag{73}
$$

33

7.8 data hand-off call departures

7.8.1 active data call hand-off departures

In the regions R_1, R_4, B_1, B_6, or B_7, a transition into state s, due to an active data call hand-off departure on a g-type platform when the cell is in state x_{dw}, will cause the state variable $z(x_{dw}, g)$ to be decreased by 1. Therefore, a permissible state x_{dw} is a predecessor state of s for an active data call hand-off departure on a g-type platform, if the state variables are related by

$$
\begin{aligned}
v(x_{dw}, k) &= v(s, k) \\
z(x_{dw}, g) &= z(s, g) + 1 \\
z(x_{dw}, k) &= z(s, k) \quad , \quad k \neq g \\
w(x_{dw}, k) &= w(s, k)
\end{aligned}
\tag{74}
$$

and $s \in R_1 \cup R_4 \cup B_1 \cup B_6 \cup B_7$.

If an active data hand-off call departs when the system is in R_2, R_3, B_2, B_3, B_4, or B_5, one data call will be chosen to be reactivated by system. Suppose that a data call which is on an l-type of mobile platform is reactivated due to an active data call hand-off on a g-type platform. Then, the state variables are related by

$(i)g = l$

$$
\begin{aligned}
v(x_{dw}, k) &= v(s, k) \\
z(x_{dw}, k) &= z(s, k) \\
w(x_{dw}, l) &= w(s, l) + 1 \\
w(x_{dw}, k) &= w(s, k) \quad , \quad k \neq l
\end{aligned}
$$

$(ii)g \neq l$

$$
\begin{aligned}
v(x_{dw}, k) &= v(s, k) \\
z(x_{dw}, g) &= z(s, g) + 1 \\
z(x_{dw}, l) &= z(s, l) - 1
\end{aligned}
$$

34

$$z(x_{dw}, k) = z(s, k) \quad , \quad k \neq g, l$$

$$w(x_{dw}, l) = w(s, l) + 1$$

$$w(x_{dw}, k) = w(s, k) \quad , \quad k \neq l \tag{75}$$

and $s \in R_2 \cup R_3 \cup B_2 \cup B_3 \cup B_4 \cup B_5$.

The corresponding transition flow is given by

$$\gamma_{dwa}(s, x_{dw}, g) = \begin{cases} \mu_D(g) \times z(x_{dw}, g) & s \in R_1 \cup R4 \cup B_1 \cup B_6 \cup B_7 \\ P(s, g) \times \mu_D(g) \times z(x_{dw}, g) & s \in R_2 \cup R_3 \cup B_2 \cup B_3 \cup B_4 \cup B_5. \end{cases} \tag{76}$$

7.8.2 held data call hand-off departures

Since a mobile may move from a cell to another cell, a data call held will hand-off to a target cell when a mobile move to target cell. This type of flow is generated when the system is in region R_2, R_3, B_2, B_3, B_4, or B_5. In this region, a transition into state s, due to a held data call hand-off departure on a g-type platform when the cell is in state x_{dw}, will cause the state variable $z(x_{dw}, g)$ to be decreased by 1. Thus a permissible state x_{dw} is a predecessor state of s for held data call hand-off departures on a g-type platform, if the state variables are related by

$$v(x_{dw}, k) = v(s, k)$$

$$z(x_{dw}, k) = z(s, k)$$

$$w(x_{dw}, k) = w(s, k) \quad , \quad k \neq g$$

$$w(x_{dw}, g) = w(s, g) + 1 \tag{77}$$

and $s \in R_2 \cup R_3 \cup B_2 \cup B_3 \cup B_4 \cup B_5$.

The corresponding transition flow is given by

$$\gamma_{dws}(s, x_{dw}, g) = \mu_D(g) \times z(x_{dw}, g) \quad s \in R_2 \cup R_3 \cup B_2 \cup B_3 \cup B_4 \cup B_5 \tag{78}$$

Therefore, the total corresponding transition flow due to the data call hand-off departure is given by

$$\gamma_{dw}(s, x_{dw}, g) = \gamma_{dwa}(s, x_{dw}, g) + \gamma_{dws}(s, x_{dw}, g). \tag{79}$$

35

C \ W	10	15	20	25	30
10	19591	39611	66781	101101	142571
12	31395	63245	106470	161070	227045
15	56916	114036	191556	289476	407796
17	80085	159885	268185	404985	570285
20	125741	249711	417956	630476	887271

Table 1: Number of permissible cell states.
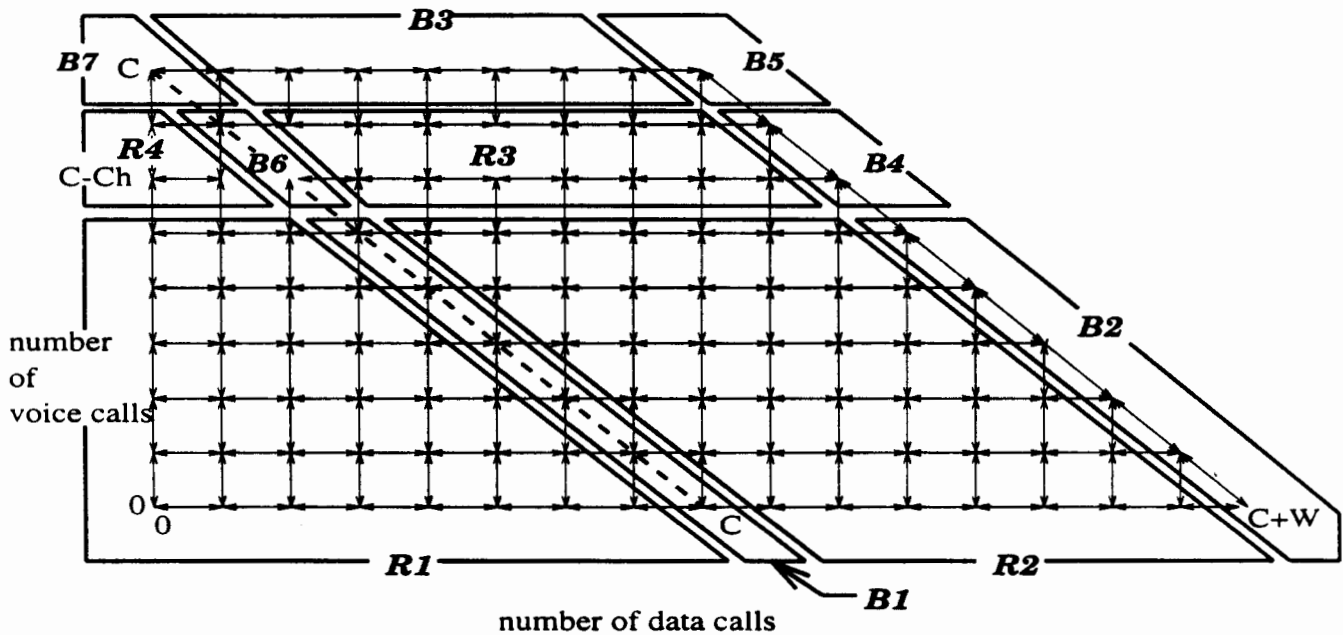
Figure 1: State space and transition diagram when $G = 1$.

	Call type	Call generation				Call completion			
		voice		data		voice		data	
Sub-spaces		new	hand-off	new	hand-off	end	hand-off	end	hand-off
Region I	**(R1)**	a	a	a	a	u	u	u	u
Region II	**(R2)**	a&s	a&s	h	h	r	r	r	r | u
Region III	**(R3)**	b	a&s	h	h	r	r	r	r | u
Region IV	**(R4)**	b	a	a	a	u	u	u	u
Boundary I	**(B1)**	a&s	a&s	h	h	u	u	u	u
Boundary II	**(B2)**	a&t	a&t	b	b	r	r	r	r | u
Boundary III	**(B3)**	b	b	h	h	r	r	r	r | u
Boundary IV	**(B4)**	b	a&t	b	b	r	r	r	r | u
Boundary V	**(B5)**	b	b	b	b	r	r	r	r | u
Boundary VI	**(B6)**	b	a&s	h	h	u	u	u	u
Boundary VII	**(B7)**	b	b	h	h	u	u	u	u

a : arriving call gains access to a channel

b : new call is blocked

h : arriving call is held

u : unchanged

r : a held data call is reactivated

s : an active data call is suspended

t : a data call is terminated

| ; or

& ; and

Table 2: Regions and boundaries in state space.
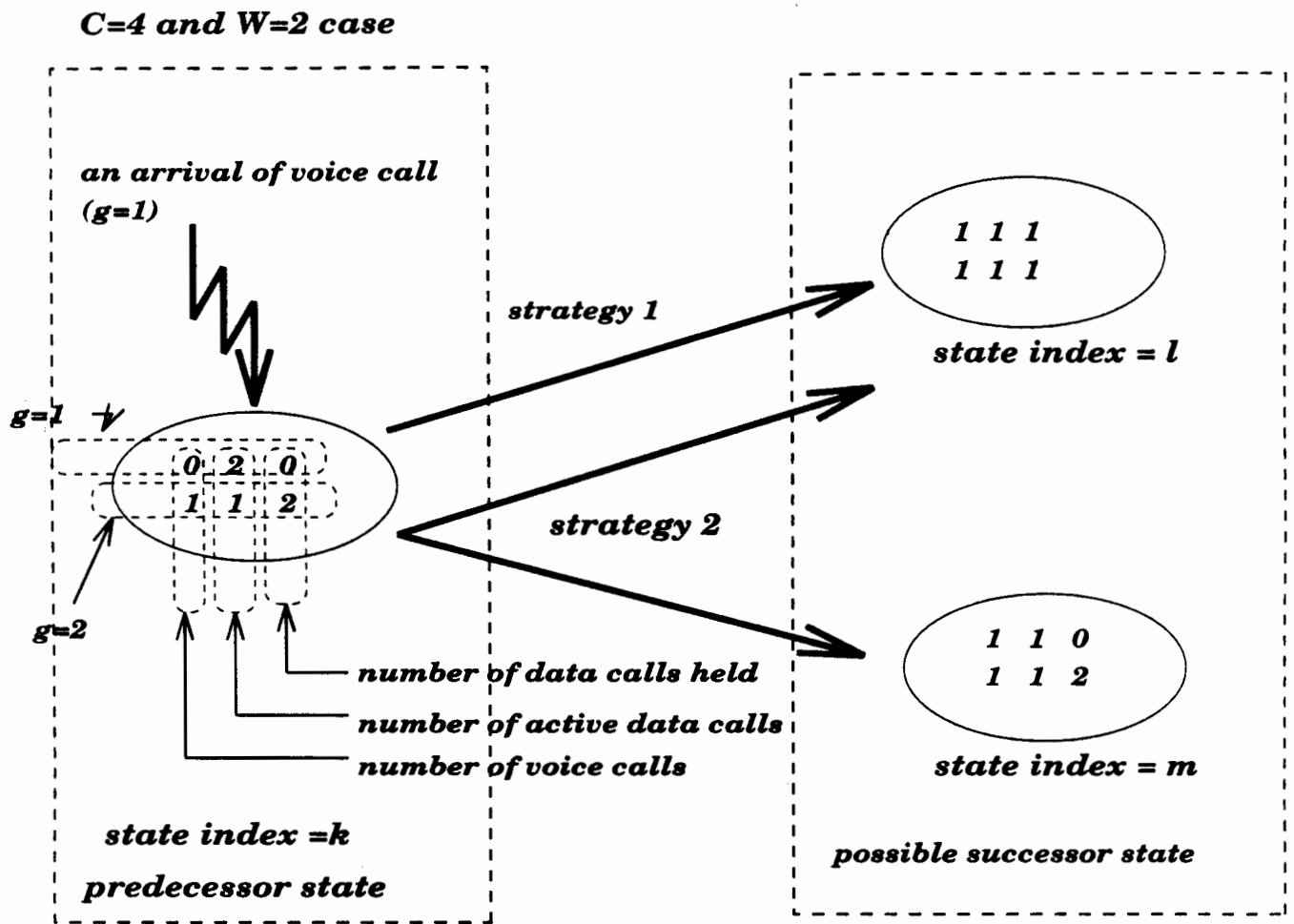
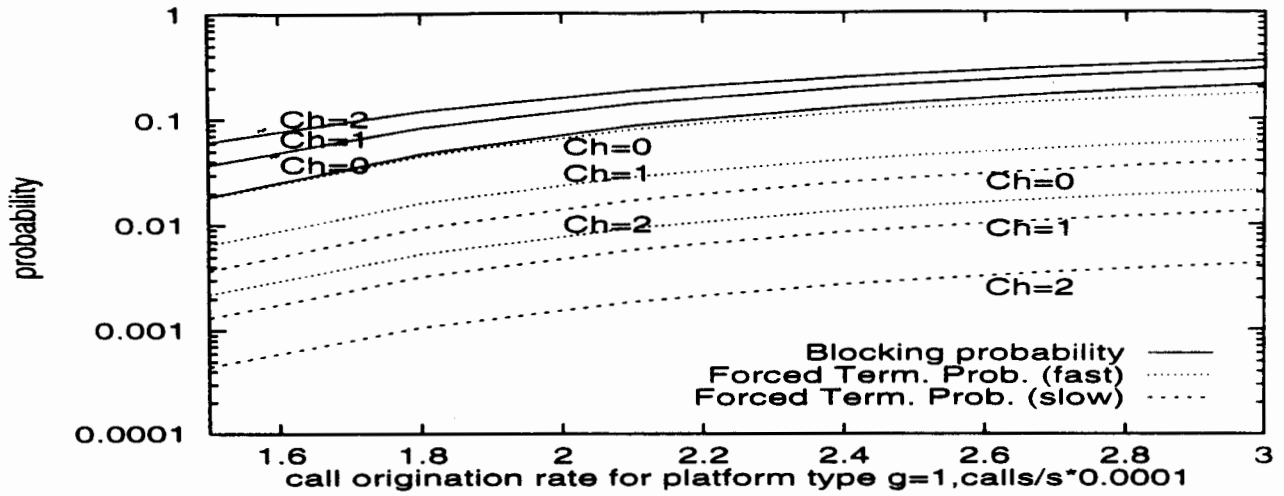Figure 2: An example of state transition with different strategies.

Figure 3: Performance evaluation of voice call traffic: C=15, G=2, $v(1,0) = v(2,0) = 300$, $\alpha(2)=\Lambda_{nv}(2)/\Lambda_{nv}(1)$=1.0, $\beta(1)=\Lambda_{nw}(1)/\Lambda_{nv}(1)$=1.0, $\beta(2)=\Lambda_{nw}(2)/\Lambda_{nv}(2)$=1.0, $\overline{T}_v(1)=\overline{T}_v(2)$=100s, $\overline{T}_w(1)=\overline{T}_w(2)$=20s, $\overline{T}_D(1)$=500s, and $\overline{T}_D(2)$=100s were assumed.
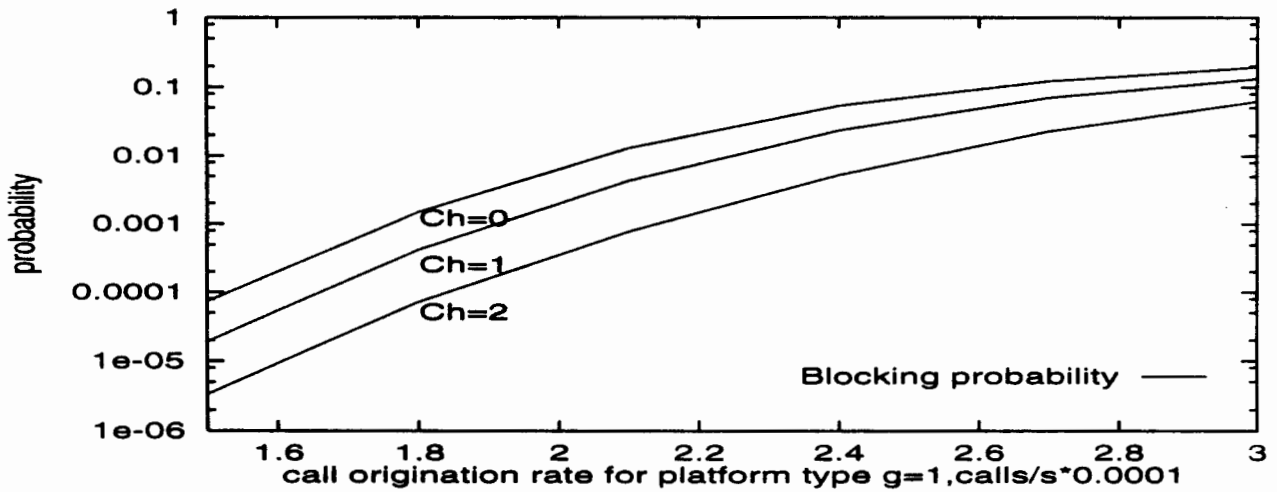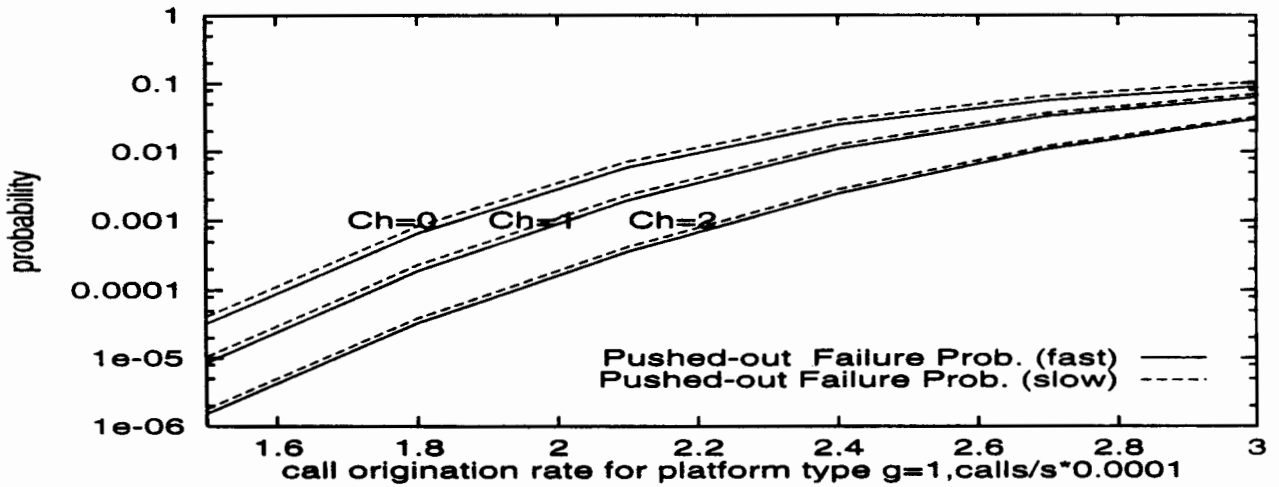

Figure 4: Blocking probability of data call: C=15, W=13, G=2, $v(1,0) = v(2,0) = 300$, $\alpha(2)=\Lambda_{nv}(2)/\Lambda_{nv}(1)$=1.0, $\beta(1)=\Lambda_{nw}(1)/\Lambda_{nv}(1)$=1.0, $\beta(2)=\Lambda_{nw}(2)/\Lambda_{nv}(2)$=1.0, $\overline{T}_v(1)=\overline{T}_v(2)$=100s, $\overline{T}_w(1)=\overline{T}_w(2)$=20s, $\overline{T}_D(1)$=500s, and $\overline{T}_D(2)$=100s were assumed.
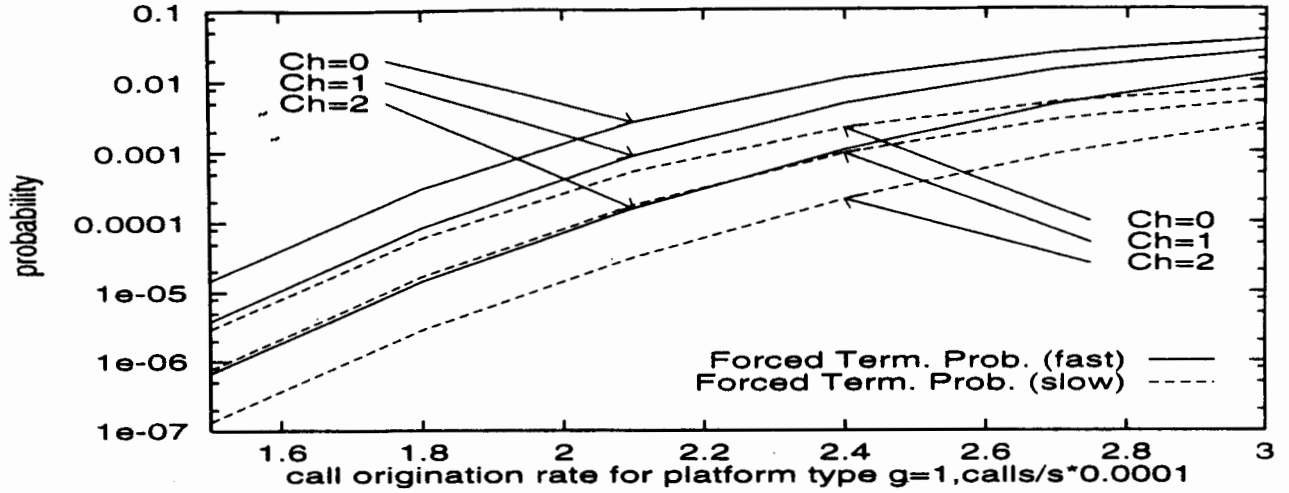
Figure 5: Forced termination probability of data call: $C=15$, $W=13$, $G=2$, $v(1,0) = v(2,0) = 300$, $\alpha(2)=\Lambda_{nv}(2)/\Lambda_{nv}(1)=1.0$, $\beta(1)=\Lambda_{nw}(1)/\Lambda_{nv}(1)=1.0$, $\beta(2)=\Lambda_{nw}(2)/\Lambda_{nv}(2)=1.0$, $\overline{T}_v(1)=\overline{T}_v(2)=100s$, $\overline{T}_w(1)=\overline{T}_w(2)=20s$, $\overline{T}_D(1)=500s$, and $\overline{T}_D(2)=100s$ were assumed.

Figure 6: Pushed-out failure probability of data call: $C=15$, $W=13$, $G=2$, $v(1,0) = v(2,0) = 300$, $\alpha(2)=\Lambda_{nv}(2)/\Lambda_{nv}(1)=1.0$, $\beta(1)=\Lambda_{nw}(1)/\Lambda_{nv}(1)=1.0$, $\beta(2)=\Lambda_{nw}(2)/\Lambda_{nv}(2)=1.0$, $\overline{T}_v(1)=\overline{T}_v(2)=100s$, $\overline{T}_w(1)=\overline{T}_w(2)=20s$, $\overline{T}_D(1)=500s$, and $\overline{T}_D(2)=100s$ were assumed.
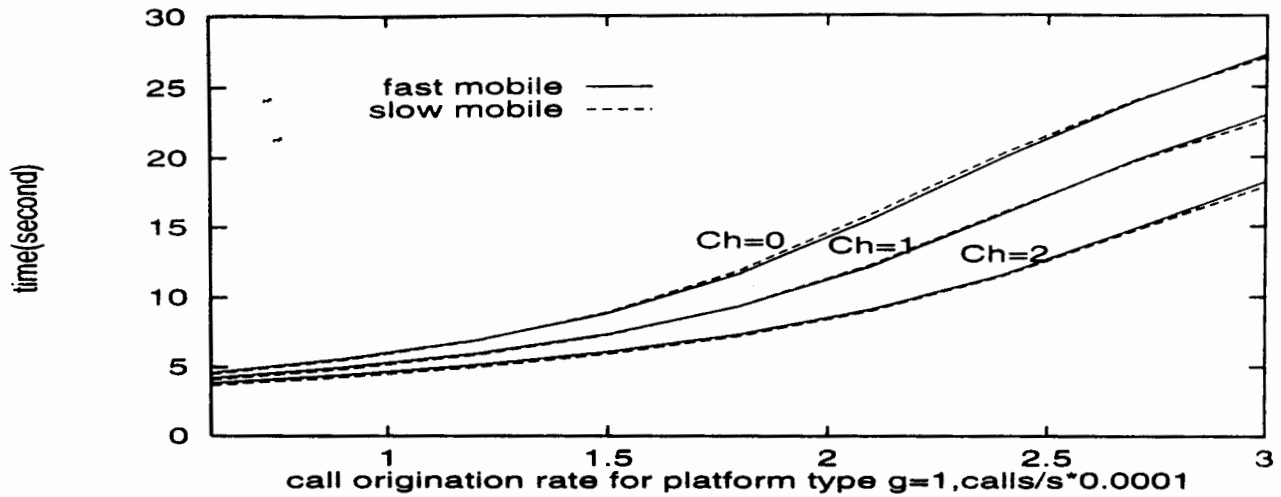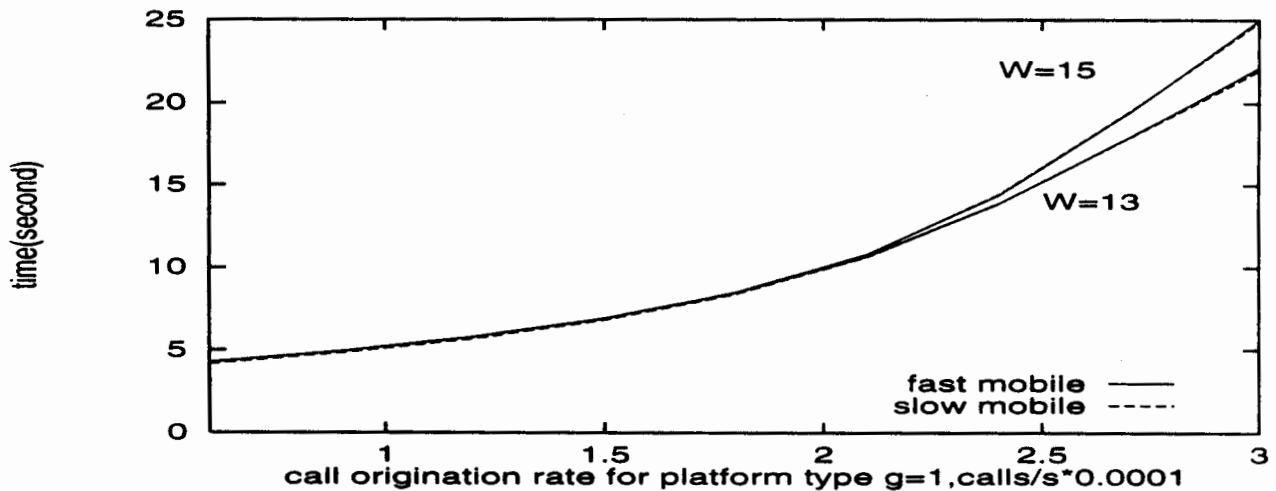
40

Figure 7: Expected waiting time of data call: $C=15$, $W=13$, $G=2$, $v(1,0) = v(2,0) = 300$, $\alpha(2)=\Lambda_{nv}(2)/\Lambda_{nv}(1)=1.0$, $\beta(1)=\Lambda_{nw}(1)/\Lambda_{nv}(1)=1.0$, $\beta(2)=\Lambda_{nw}(2)/\Lambda_{nv}(2)=1.0$, $\overline{T}_v(1)=\overline{T}_v(2)=100$s, $\overline{T}_w(1)=\overline{T}_w(2)=20$s, $\overline{T}_D(1)=500$s, and $\overline{T}_D(2)=100$s were assumed.

Figure 8: Expected waiting time of data call: $C=15$, $C_h=0$, $G=2$, $v(1,0) = v(2,0) = 300$, $\alpha(2)=\Lambda_{nv}(2)/\Lambda_{nv}(1)=1.0$, $\beta(1)=\Lambda_{nw}(1)/\Lambda_{nv}(1)=1.0$, $\beta(2)=\Lambda_{nw}(2)/\Lambda_{nv}(2)=1.0$, $\overline{T}_v(1)=\overline{T}_v(2)=100$s, $\overline{T}_w(1)=\overline{T}_w(2)=20$s, $\overline{T}_D(1)=500$s, and $\overline{T}_D(2)=100$s were assumed.

41
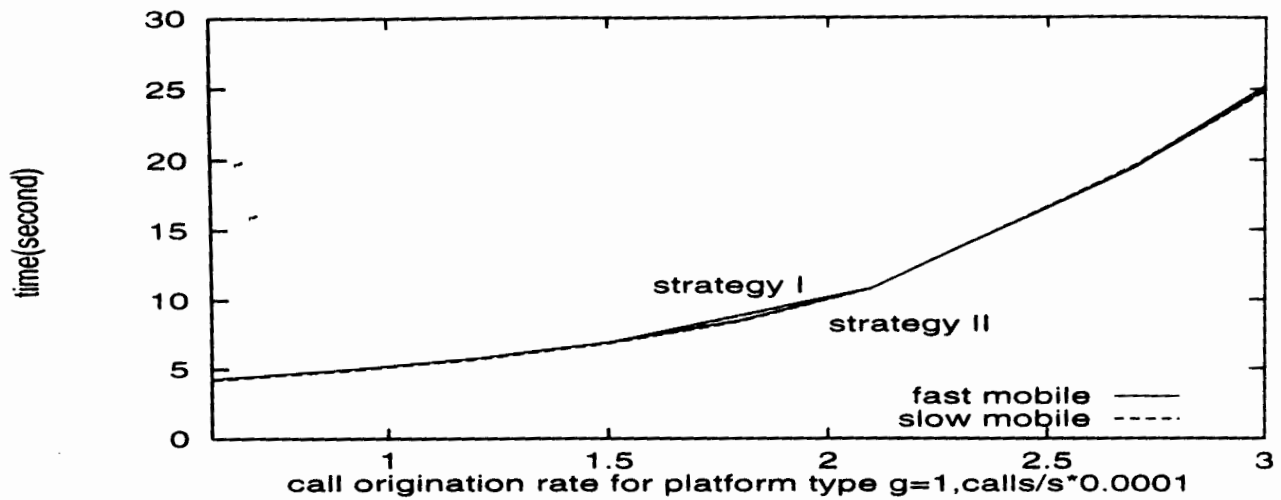
Figure 9: Comparison of the expected waiting time using two different data call management strategies: $C=15$, $C_h=0$, $G=2$, $v(1,0)=v(2,0)=300$, $\alpha(2)=\Lambda_{nv}(2)/\Lambda_{nv}(1)=1.0$, $\beta(1)=\Lambda_{nw}(1)/\Lambda_{nv}(1)=1.0$, $\beta(2)=\Lambda_{nw}(2)/\Lambda_{nv}(2)=1.0$, $\overline{T}_v(1)=\overline{T}_v(2)=100$s, $\overline{T}_w(1)=\overline{T}_w(2)=20$s, $\overline{T}_D(1)=500$s, and $\overline{T}_D(2)=100$s were assumed.
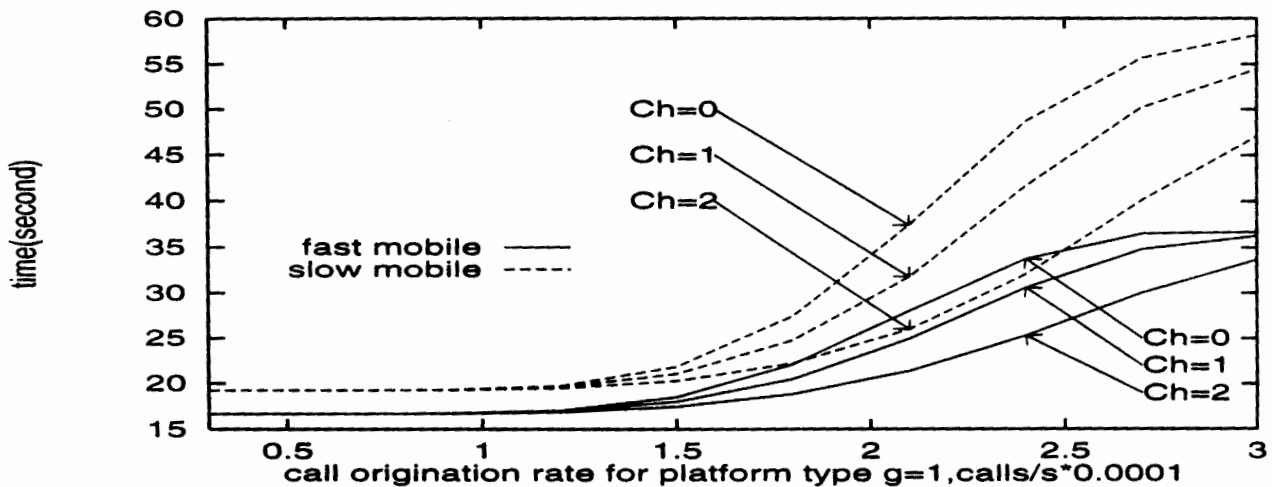

Figure 10: Average time a data call spends in the system: $C=15$, $W=13$, $G=2$, $v(1,0)=v(2,0)=300$, $\alpha(2)=\Lambda_{nv}(2)/\Lambda_{nv}(1)=1.0$, $\beta(1)=\Lambda_{nw}(1)/\Lambda_{nv}(1)=1.0$, $\beta(2)=\Lambda_{nw}(2)/\Lambda_{nv}(2)=1.0$, $\overline{T}_v(1)=\overline{T}_v(2)=100$s, $\overline{T}_w(1)=\overline{T}_w(2)=20$s, $\overline{T}_D(1)=500$s, and $\overline{T}_D(2)=100$s were assumed.