

Traffic Performance Analysis for Cellular Communication Systems with Mixed Platform Types and Queued Hand-Offs

by

Cezary Purzynski and Stephen S. Rappaport
Department of Electrical Engineering
State University of New York
Stony Brook, NY 11794-2350

Keywords: cellular communication systems, mobile radiotelephone, wireless communication networks.

ABSTRACT

Cellular communication systems that support a mixture of platform types and queueing of hand-off calls are considered. In pure loss systems, if there are no channels available at the target gateway, the hand-off attempt fails and the call is forced to terminate. In delay systems however, hand-off calls can be held in queue while the supporting mobile is within a transition region where acceptable performance can be provided by at least two different gateways (base stations). When only a finite number of hand-off calls can be queued, mixed delay/loss configurations arise. An analytically tractable model is developed to predict communications traffic performance in such systems.

A suitable state characterization is identified which allows the problem to be formulated within the framework of multi-dimensional birth-death processes - an approach which has been used in some recent work that considers loss (lost call) systems with cut-off priority for hand-off attempts. The present paper extends the approach and demonstrates how delay and combined delay/loss systems can be considered within the analytical framework that is being developed. The formulation allows consideration (within the same system) of a mixture of platform types distinguished by mobility characteristics and multiplicity of calls that can be supported.

Theoretical performance characteristics obtained show trade-offs among blocking and forced termination probabilities and carried traffic.

The Multiple Call Hand-Off Problem with Queued Hand-Offs and Mixed Platform Types

by

Cezary Purzynski and Stephen S. Rappaport
Department of Electrical Engineering
State University of New York
Stony Brook, NY 11794-2350

Keywords: cellular communication systems, mobile radiotelephone, wireless communication networks.

ABSTRACT

Cellular communication systems that support a mixture of platform types and queueing of hand-off calls are considered. In pure loss systems, if there are no channels available at the target gateway, the hand-off attempt fails and the call is forced to terminate. In delay systems however, hand-off calls can be held in queue while the supporting mobile is within a transition region where acceptable performance can be provided by at least two different gateways (base stations). When only a finite number of hand-off calls can be queued, mixed delay/loss configurations arise. An analytically tractable model is developed to predict communications traffic performance in such systems.

A suitable state characterization is identified which allows the problem to be formulated within the framework of multi-dimensional birth-death processes - an approach which has been used in some recent work that considers loss (lost call) systems with cut-off priority for hand-off attempts. The present paper extends the approach and demonstrates how delay and combined delay/loss systems can be considered within the analytical framework that is being developed. The formulation allows consideration (within the same system) of a mixture of platform types distinguished by mobility characteristics and multiplicity of calls that can be supported.

Theoretical performance characteristics obtained show trade-offs among blocking and forced termination probabilities and carried traffic.

INTRODUCTION

A general framework for analytically tractable models of traffic performance of cellular systems with hand-offs is presented in [1], [2], [3] and [4]. These focus primarily on systems using loss type service disciplines. That is calls that cannot obtain resources are immediately cleared from the system. Mixed delay/loss disciplines for systems with single call hand-offs are considered in [5], [6] and [7]. Related work appears in [8]. Extension of the general framework (which allows a multiplicity of platform types) to *combined delay/loss systems* is suggested in [1] and [3] but not developed there. The extension primarily requires a suitable state characterization which makes the class of problems amenable to solution within the same framework. This paper identifies a suitable state characterization for this purpose. The extended framework is then used to solve several example problems. Computed performance characteristics are obtained and discussed. It should be noted that this paper's focus is on one aspect of the hand-off problem - the resource availability issue. Another aspect of the problem - the hand-off initiation issue is investigated in [9],[10],[11],[12] and [13].

MODEL DESCRIPTION

We consider a large geographical region, tessellated by cells that are defined by proximity to designated network gateways (base stations). We consider a cell to be an area in which a communication link of acceptable quality can be established between a mobile platform and the cell gateway. Generally, the cell coverage areas overlap. A cell can be conceptually partitioned into two zones. In one, *ONLY* the cell's own gateway is able to provide service. We call this an *inner zone*. In the other, called the *transition zone*, at least two gateways can provide a platform with a link of acceptable quality.

The service region is traversed by large numbers of mobile platforms. The platforms are of several types, differing in mobility and numbers of communication ports on a platform. There are G platform types, labeled $g=1,2,3 \dots G$. A platform of type g can support up to $N(g)$ calls. Mobility of a g -type platform can be described by defining a *dwelt time in an inner zone* and a *dwelt time in a transition zone*. The *dwelt time in an inner zone* is the time that a g -type platform remains in communication range of only a single gateway. This is a random variable denoted $T_D(g)$. The *dwelt time in a transition zone* is the time that a g -type platform remains in the transition zone. It is also a random variable denoted $T_T(g)$. A platform that has at least one call in progress is called a *communicating platform*. A communicating platform with a link in a *source* cell can move to a *target* cell. If i calls are in progress on a platform that needs resources in the target cell, a demand for i channels at the target gateway is placed. The event is called a *hand-off attempt of order i* . In general a hand-off can be serviced according to a *loss type* service discipline, for which calls that cannot be assigned resources in the target cell immediately are cleared from the system. An alternative is a *delay type* service discipline in which a platform with hand-off needs is placed in queue awaiting the release of a sufficient number of channels to satisfy its demand. The source gateway is able to provide service to such a platform for as long as the platform remains in the transition zone. If the platform leaves the transition zone before sufficient resources are available at

the target gateway, at least some of the calls in progress on the platform will be terminated. When a call in progress is terminated, a *hand-off failure* is said to occur. It should be emphasized that calls can be successfully completed while the supporting platform is in the transition zone. Such an event is not a hand-off failure. Many variations of the two basic strategies described above are possible. In what follows we will present an analysis of a hybrid delay/loss system, i.e. a system with a hand-off queue and limited waiting space of size Q . A more detailed description of the discipline is presented in subsequent sections. The *unencumbered* session duration is the amount of time that a call would remain in progress if it were not terminated by hand-off failure. It is a random variable $T(g)$.

At any given instant, a single cell in the system can be characterized by its state. The state of the cell is specified by $G+2$ n-tuples of integers. The first G n-tuples describe the conditions prevailing at the target gateway. The g -th n-tuple consists of $N(g)$ integers: $v_{g1} v_{g2} \dots v_{gN(g)}$. The integer v_{gn} denotes the number of communicating platforms of type g with exactly n calls in progress (at the gateway). The remaining two n-tuples specify the conditions in the hand-off queue i.e. the status of platforms awaiting resources at the target gateway. Both consist of Q integers. The n-tuple $z_1 z_2 \dots z_Q$ specifies the number of calls in progress on queued platforms. In particular, z_q is the number of calls in progress on the q -th platform in queue. The n-tuple, $x_1 x_2 \dots x_Q$, specifies the types of the queued platforms, that is, x_q is the platform type of the q -th platform in queue. Succinctly, the state of a cell can be written as

$$\left\{ \begin{array}{cccc} v_{11} & v_{12} & \dots & v_{1N(1)} \\ v_{21} & v_{22} & \dots & v_{2N(2)} \\ \dots & \dots & \dots & \dots \\ v_{G1} & v_{G2} & \dots & v_{GN(G)} \\ z_1 & z_2 & \dots & z_Q \\ x_1 & x_2 & \dots & x_Q \end{array} \right\} \quad (1)$$

Permissible states correspond to all possible collections of n-tuples of this form, for which the channel limit and quota constraints as well as the platform limit and quota constraints are met [4]. The state of a cell changes as time progresses. The state transitions are driven by underlying random processes, which include the following: (1) Generation of new calls; (2) Call completions; (3) Arrival of communicating platforms at the cell (*hand-off arrivals*); (4) Departure of communicating platforms from the cell (*hand-off departures*); (5) Departure of communicating platforms from the transition zone (*defections from the hand-off queue*).

The entire cellular system, of course consists of many cells. Since certain hand-off events will result in a change of state of two cells simultaneously, a complete model should characterize *system state*. System state is a concatenation of the corresponding cell states. However, this approach leads to a system description burdened by overwhelming dimensionality. An alternative approach is to consider only one cell and balance the average rate of hand-off attempt arrivals and departures [1], [2].

SYSTEM STATES AND GATEWAY CHARACTERISTICS.

The cell state is described by $G+2$ N-tuples. The permissible states of the cell can be enumerated using an index $s=0,1,\dots,s_{\max}$. Then the state variables v_{gn} , z_q , x_q are explicitly dependent on the current state s of the system. That is $v_{gn}=v(s,g,n)$, $z_q=z(s,q)$ and $x_q=x(s,q)$. The states were enumerated first according to increasing values of the number of occupied channels; next according to increasing values of the total number of communicating platforms; next according to increasing values of $v_{11}(s)$; next according to increasing values of $v_{21}(s)$; and so forth thru $v_{GN(G)}$; next according to number of calls in progress on first platform waiting in queue; next according to number of calls on 2-nd platform in queue; and so forth thru $w_{QL}(s)$; and lastly according to increasing platform type.

A number of interesting gateway characteristics can be determined from the state variables. The total number of channels being used by g -type platforms in state s , is

$$j(s,g) = \sum_{i=1}^{N(g)} i \cdot v(s,g,i) \quad (2)$$

The total number of channels in use at a gateway in state s is

$$j(s) = \sum_{g=1}^G j(s,g) \quad (3)$$

The number of communicating platforms of type g is

$$w(s,g) = \sum_{i=1}^{N(g)} v(s,g,i) \quad (4)$$

and the total number of communicating platforms is

$$w(s) = \sum_{g=1}^g w(s,g) \quad (5)$$

DRIVING PROCESSES AND SYSTEM DYNAMICS

The number of g -type noncommunicating platforms per cell is assumed to be known and independent of the cell state. It is denoted by v_{g0} . Since a g -type platform is equipped with $N(g)$ communication ports, the number of idle ports on g -type platforms is $N(g) \cdot v_{g0}$. It is further assumed that the number of idle ports per cell on each platform

type is much larger than the number of channels in a cell, C . The new call arrival rate per port on a noncommunicating platform is also given and denoted by Λ_0 . Therefore the new call arrival rate on g -type noncommunicating platforms is

$$\text{new call arrival rate on } g\text{-type noncommunicating platforms} = N(g) \cdot \Lambda_0 \cdot v_{g0} \quad (6)$$

In a similar manner a new call origination rate on g -type communicating platforms can be found. The new call arrival rate per port on a g -type platform with i calls in progress is denoted as Λ_{gi} , which in general can be different from Λ_0 . Defining $\alpha_{gi} = \Lambda_{gi} / \Lambda_0$, then new call arrival rate on a g -type platform with i calls in progress is $\alpha_{gi} \cdot \Lambda_0 \cdot (N(g) - i)$. Since in state s there are $v(s, g, i)$ g -type platforms with i calls in progress, the new call arrival rate on that platform type is

$$\text{new call arrival rate on } g\text{-type communicating platforms} = \alpha_{gi} \cdot \Lambda_0 \cdot (N(g) - i) \cdot v(s, g, i) \quad (7)$$

The total new call arrival rate in state s can be expressed as

$$\Lambda_n(s) = \Lambda_0 \sum_{g=1}^G v_{g0} \cdot N(g) \left[1 + \sum_{i=1}^{N(g)} \alpha_{gi} (1 - i/N(g)) \cdot v(s, g, i) / v_{g0} \right] \quad (8)$$

The following assumptions, along with those previously discussed, render the problem amenable to solution using the notion of multi-dimensional birth-death processes.

- (1) The new call arrival process is a Poisson point process with state dependent means.
- (2) The unencumbered session duration $T(g)$ is an exponentially distributed random variable with mean $\bar{T}(g) = \mu^{-1}(g)$.
- (3) The hand-off arrival process is a Poisson point process.
- (4) The dwell time of a g -type platform in the inner zone is an exponentially distributed random variable with mean $\bar{T}_D(g) = \mu_D^{-1}(g)$.
- (5) The dwell time of a g -type platform in the transition zone is an exponentially distributed random variable with mean $\bar{T}_T(g) = \mu_T^{-1}(g)$.

The state transition flows, resulting from the driving processes are developed in subsequent sections.

The operation of the system under consideration here can be succinctly stated. The target gateway uses a first come first served discipline for hand-off attempts. Some hand-off attempts awaiting resources at the target gateway leave the queue because the supporting mobile leaves the transition zone. NO new call will be served at a gateway where hand-off attempts are waiting. A sophisticated reader may notice some similarity between the queueing discipline described here and classical schemes. Overall this is a head of the line priority queueing system with priority for hand-off attempts, no waiting room for new call originations and a finite waiting room and *impatience* for hand-off attempts. An important distinction however is that impatience applies to platforms with

multiple hand-off call attempts as well as to hand-off call attempts corresponding to calls that are successfully completed while the supporting platform is in the transition zone. As in [1] -[4], the analytical approach is to write the probability flow balance equations and solve for the cell state probabilities. From these, system performance measures are calculated. The set of $s_{\max}+1$ (non-linear) flow balance equations is of the following form

$$\sum_{s=0}^{s_{\max}} q(\xi,s) \cdot p(s) = 0 \quad \xi = 0, 1 \dots, s_{\max}-1 \quad (9)$$

$$\sum_{s=0}^{s_{\max}} p(s) = 1 \quad (10)$$

In the above equations $q(\xi,s)$, for $\xi \neq s$, represents the net probability flow from state s to state ξ and $q(\xi,\xi)$ is the total flow out of state ξ . Flow into a state is taken as a positive quantity. The total state-to-state flows, $q(\xi,s)$ can be determined by considering components of flow from state s to state ξ , due to underlying driving processes.

PROBABILITY FLOW EQUATIONS

In this section we consider a cellular system operating as described above. No channel quota, platform quota nor platform limit restrictions were imposed on the system. Although the inclusion of such restrictions is straightforward, for the sake of clarity it is not pursued here.

In the following, we assume that the cell is in state s and we find the rates of flow out state s due to events generated by the underlying processes. We also show the form of the successor state, ξ as a function of the current state, s .

Flow due to new call originations

A new call arising when the cell is in state s can be served only if there is a channel available for immediate assignment to that call. Since cut-off priority [1] can be combined with hand-off request queueing, a new call will be served only if the number of channels in use in state s , $j(s)$, is less than $C-C_h$ (where C_h is the number of channels reserved exclusively for hand-offs) and no hand-off requests are queued awaiting channel release. Otherwise a new call is blocked and cleared from the system. A call that was not blocked, originating on a noncommunicating g -type platform "transforms" that platform into a g -type platform with 1 call in progress. As a result we find the system in state t with $v(\xi,g,1) = v(s,g,1) + 1$. The new call origination rate on g -type noncommunicating platforms is independent of the cell state and given by (6). Summarizing, we find that the component of flow out of state s due to new call arrivals on g -type noncommunicating platforms is

$$r_n(s,g,0) = \begin{cases} 0 & \text{if } j(s) \geq C - C_h \text{ or queue not empty} \\ \Lambda_0 \cdot N(g) \cdot v_{g0} & \text{otherwise} \end{cases} \quad (11)$$

and the corresponding successor state is

$$\left\{ \begin{array}{c} \cdot \\ \cdot \\ v_{g,1+1} \\ \cdot \\ \cdot \end{array} \right\} \quad (12)$$

Note: Dots represent state variables that are the same in both the current and successor states.

In a similar manner, the flow out of state s due to new call arrivals on g -type communicating platforms with i calls in progress can be found. A new call which is not blocked "transforms" a g -type platform with i calls in progress into a g -type platform with $(i+1)$ calls in progress. That is reflected by a transition to state ξ for which the state variable $v(\xi,g,i) = v(s,g,i) - 1$ and the state variable $v(\xi,g,i+1) = v(s,g,i+1) + 1$. In state s , the new call arrival rate on g -type platforms with i -calls in progress is given by (7). The flow out of state s due to new call arrivals on g -type platforms with i calls in progress is found to be

$$r_n(s,g,i) = \begin{cases} 0 & \text{if } j(s) \geq C - C_h \text{ or queue not empty} \\ \alpha_{gi} \cdot \Lambda_0 \cdot (N(g)-i) \cdot v(s,g,i) & \text{otherwise} \end{cases} \quad (15)$$

The corresponding successor state is

$$\left\{ \begin{array}{c} \cdot \\ \cdot \\ v_{g,i-1} \quad v_{g,i+1+1} \\ \cdot \\ \cdot \end{array} \right\} \quad (16)$$

Flow due to call completions.

After a call completion on g -type platform (located in the inner zone) with i -calls in progress, we find the system in state ξ for which the state variable $v(\xi,g,i) = v(s,g,i) - 1$ and the variable $v(\xi,g,i-1) = v(s,t,i-1) + 1$. The state transition reflects the fact that after the call completion we have one less g -type platform with i calls in progress and one more with $i-1$ calls. The transition rate due to call terminations on g -type platforms (located in the inner zone) with i calls in progress in state s is found to be

$$r_c(s,g,i) = v(s,g,i) \cdot i \cdot \mu(g) \quad (17)$$

and the corresponding successor state is

$$\left\{ \begin{array}{cccc} \cdot & \cdot & \cdot & \cdot \\ \cdot & v_{g,i-1}+1 & v_{g,i-1} & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{array} \right\} \quad (18)$$

After a call completion on a g-type platform with 1 call in progress, the platform will become a noncommunicating platform. Since the number of noncommunicating platforms is not included in the state description, the successor state is

$$\left\{ \begin{array}{cccc} \cdot & \cdot & \cdot & \cdot \\ \cdot & v_{g,i-1} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{array} \right\} \quad (18)$$

Transitions due to call completions on the q-th platform in the hand-off attempt queue occur at the rate

$$r_{ct}(s,q) = \mu(x(s,q)) \cdot z(s,q) \quad (19)$$

Call completion on the q-th platform in queue will result in a transition to state ξ for which $z(\xi,q) = z(s,q) - 1$. Recall that z_q is the number of calls in progress on the q-th platform in queue. The successor state is

$$\left\{ \begin{array}{cccc} \cdot & \cdot & \cdot & \cdot \\ z_1 & \cdot & z_{q-1} & \cdot & z_Q \\ x_1 & \cdot & x_q & \cdot & x_Q \end{array} \right\} \quad (20)$$

If $z(s,q) = 1$, then after the call termination on the q-th platform in queue that platform becomes the noncommunicating platform. As a result the q-th platform leaves the queue and all platforms in positions $q+1, q+2, \dots, Q$ will move forward to positions $q, q+1, \dots, Q-1$. The resulting successor state is

$$\left\{ \begin{array}{cccccc} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ z_1 & \cdot & z_{q-1} & z_{q+1} & \cdot & z_Q & 0 \\ x_1 & \cdot & x_{q-1} & x_{q+1} & \cdot & x_Q & 0 \end{array} \right\} \quad (21)$$

Slightly more involved is the case when in state s , all but one of the first queued platform's calls have channels reserved at the target gateway (i.e. $j(s)+z(s,1)=C+1$) and a call supported by that platform terminates. Since after the call termination all of the platform's calls can be served by the target gateway, the platform must not be queued any more. The event leads to the transition to state ξ , for which $v(\xi, x(s,1), z(s,1)-1) = v(s, x(s,1), z(s,1)-1) + 1$ and platforms remaining in queue will advance by one position. The successor state is shown below

$$\left\{ \begin{array}{ccccccc} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & v_{x_1, z_1-1+1} & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ z_2 & \cdot & z_q & \cdot & z_Q & 0 & \cdot \\ x_2 & \cdot & x_q & \cdot & x_Q & 0 & \cdot \end{array} \right\} \quad (22)$$

Flow due to hand-off arrivals.

Hand-off arrivals of order i on g -type platforms occur at the rate

$$\Lambda(g,i) = \Lambda_h \cdot F(g,i) \quad , \quad (23)$$

where Λ_h is the total hand-off arrival rate and $F(g,i)$ is a fraction of hand-offs that are of order i and occur on g -type platforms. For now, these parameters are assumed to be known. In a subsequent section it will be shown how these are determined from the underlying processes and the system dynamics.

A hand-off attempt of order i will result in a state transition only when the hand-off attempt queue is not full. If the queue is empty and there are at least i channels available, all calls that were handed-off are immediately served by the destination gateway. The successor state is

$$\left\{ \begin{array}{ccc} \cdot & \cdot & \cdot \\ \cdot & v_{g,i+1} & \cdot \\ \cdot & \cdot & \cdot \end{array} \right\} \quad (24)$$

If however, the queue is not empty or there are less than i channels available, the arriving hand-offs will be placed in queue, awaiting the release of a sufficient number of channels. If q platforms were queued before the hand-off arrival, the arriving platform's type and number of calls in progress will be recorded in variables x_{q+1} and z_{q+1} respectively.

Flow due to hand-off departures.

Hand-off departures of order i on g -type platforms when the gateway is in state s occur at the rate

$$r_d(s,g,i) = \mu_D(g) \cdot v(s,g,i) \quad . \quad (25)$$

The expression for the successor state, ξ , cannot be presented in a compact form. The state variable that is always decreased by one is the number of g -type platforms with i calls in progress, that is $v(\xi,g,i)=v(s,g,i)-1$. If however the queue is not empty the channels released by the departing platform will be assigned to a possibly large number of platforms waiting in the queue. More specifically, the number of channels reserved for use by the hand-off attempts on the first platform in queue in state s is $C-j(s)$. Therefore after

the hand-off departure of order i , the total pool of channels available to queued hand-offs is $C-j(s)+i$. Those channels are assigned to the queued hand-offs on the head of the line priority basis. As a result, a recursive procedure has to be used to obtain the expression for the successor state. To illustrate such a transition we will give one example in which the number of channels available at the gateway after hand-off departure of a g -type platform with i calls in progress ($C - j(s) + i$) is sufficient to serve all calls on the first and second platforms in the queue ($z(s,1) + z(s,2)$). We further assume that the number of channels available at the gateway after the hand-off departure is smaller than the number of channels required to serve all calls in progress on the first three platforms in queue ($z(s,1) + z(s,2) + z(s,3)$). The conditions are expressed mathematically below

$$[z(s,1) + z(s,2)] \leq [C - j(s) + i] < [z(s,1) + z(s,2) + z(s,3)]$$

In effect, the first two platforms leave the queue. The remainder of the channels released by the hand-off departure is held to serve a fraction of calls on the third platform in queue. The successor state is shown below

$$\left\{ \begin{array}{cccccc} v_{11} & v_{12} & \cdot & \cdot & \cdot & \cdot & v_{1N(1)} \\ v_{21} & v_{22} & \cdot & \cdot & \cdot & \cdot & v_{2N(2)} \\ \cdot & \cdot & \cdot & v_{x_1 z_1 + 1} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & v_{g_i - 1} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & v_{x_2 z_2 + 1} & \cdot & \cdot & \cdot \\ v_{G1} & v_{G2} & \cdot & \cdot & \cdot & \cdot & v_{GN(G)} \\ z_3 & \cdot & z_i & \cdot & z_Q & 0 & 0 \\ x_3 & \cdot & x_i & \cdot & x_Q & 0 & 0 \end{array} \right\} \quad (26)$$

Flow due to defections from the hand-off queue.

The transition rate due to the q -th queued platform defecting from the queue is

$$r_d(s,q) = \mu_T(x_q(s)) \quad (27)$$

The defection of the i -th platform in the queue will result in the removal of that platform from the queue and the successor state is

$$\left\{ \begin{array}{cccccc} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ z_1 & \cdot & z_{q-1} & z_{q+1} & \cdot & z_Q & 0 \\ x_1 & \cdot & x_{q-1} & x_{q+1} & \cdot & x_Q & 0 \end{array} \right\} \quad (28)$$

A special case occurs when the first platform in queue defects. Since $(C - j(s))$ of its calls can be served by the target gateway, only the remainder will be terminated. The

gateway effectively "sees" an arriving hand-off of order $(C-j(s))$. The event results in a transition to the following state, ξ

$$\left\{ \begin{array}{cccccc} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & v_{x_1, (C-j(s))} + 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ z_2 & \cdot & z_q & \cdot & z_Q & 0 \\ x_2 & \cdot & x_q & \cdot & x_Q & 0 \end{array} \right\} \quad (29)$$

DETERMINATION OF HAND-OFF ARRIVAL RATES.

In the analysis it was assumed that the average hand-off arrival rates of each type and order were given. Actually these parameters have to be determined from the dynamics of the process. The observation that a hand-off departure from one cell corresponds to a hand-off arrival of the same type and order in another cell leads to the conclusion that in a homogeneous system the average hand-off arrival rate of g-type platforms with i calls in progress has to be equal to the average hand-off departure rate of the same type and order. Let $\Lambda_h(g,i)$ denote the i-th order hand-off arrival rate on g-type platforms. Similarly, let $\Delta_h(g,i)$ denote the i-th order hand-off departure rate on g-type platforms. The probability of a hand-off departure of order i on g-type platform when the system is in state s is just a ratio of flow out of state s due to these events to the total flow out of state s. Let $d_{gij|s}$ denote the probability of a hand-off departure of order i on g-type platform when the cell is in state s. Then, it follows that

$$d_{gij|s} = \mu_D(g) \cdot v(s,g,i) / r(s) \quad , \quad (30)$$

where $r(s) = -q(s,s)$.

The overall probability of a hand-off departure of order i on g-type platform is

$$d_{gi} = \sum_{s=0}^{s_{\max}} [\mu_D(g) \cdot v(s,g,i) / r(s)] \cdot \tilde{p}(s) \quad , \quad (31)$$

where $\tilde{p}(s)$ is the probability that the birth-death process visits state s. The relationship between $\tilde{p}(s)$ and $p(s)$ can be expressed as

$$\tilde{p}(s) = r(s) \cdot p(s) / \sum_s r(s) \cdot p(s) \quad . \quad (32)$$

Hence

$$d_{gi} = \frac{\sum_{s=0}^{s_{\max}} \mu_D(g) \cdot v(s,g,i) \cdot p(s)}{\sum_s r(s) \cdot p(s)} \quad (33)$$

The denominator of (33) can be recognized as an average rate at which transition events occur. The numerator is the average rate at which hand-off departures of order i on g -type platforms occur. That is, hand-off departures of order i for g -type platform occur at an average rate given by

$$\Delta_h(g,i) = \sum_s p(s) \cdot \mu_D(g) \cdot v(s,g,i) \quad (34)$$

PERFORMANCE MEASURES

Blocking probability

Calls originating when the channel demand (defined as current channel use plus the number of channel requests registered at the gateway by queued hand-offs) exceeds $C-C_h$ are not admitted into service and are cleared from the system. Define the following set of states B_b

$$B_b = \{ s : (j(s) + z(s,1)) \geq C - C_h \} \quad .$$

If the cell is in state $s \in B_b$ and a new call arises, blocking occurs. Therefore the blocking probability is given by

$$P_b = \sum_{s \in B_b} p(s) \quad (35)$$

Hand-off failure probability

A communicating platform arriving into the cell encounters one of the following conditions prevailing at the destination gateway.

- (a) The number of available channels is equal or greater than the number of calls currently in progress on the arriving platform. All calls can be handed off successfully without delay.
- (b) The number of available channels is smaller than the number of calls currently in progress. The platform is placed in queue awaiting the release of a sufficient number of channels. If the platform leaves the transition zone before some (all) of the queued calls were able to reserve a channel, all calls that did not obtain a channel at the target gateway will be terminated.

- (c) There are no channels available at the destination gateway. If there is room in the queue, the platform is placed there. If it leaves the transition zone before getting to the first place in the queue, all of its calls are terminated. Otherwise their fate is determined as in case (b).
- (d) If the queue is full all calls will be terminated immediately. The average rate at which g-type platforms defect from the i-th place in queue can be expressed as

$$\sum_{s \in H_{gq}} \mu_T(x_i(s)) \cdot p(s) \quad , \quad (36)$$

where H_{gq} is the set of states in which the q-th position in the hand-off queue is occupied by a platform of type g i.e. $H_{gq} = \{s: x_q(s) = g\}$.

As described in (b) above, only a fraction of calls on a platform defecting from the first place in queue is terminated, namely $z_1(s) - (C - j(s))$, the average rate at which calls are terminated due to g-type platforms defecting from the first place in queue is found to be

$$D(g,1) = \sum_{s \in H_{g1}} \mu_T(x_1(s)) \cdot p(s) \cdot (z_1(s) - (C - j(s))) \quad . \quad (37)$$

A platform defecting from the q-th position ($q \neq 1$) in queue will have all of its calls terminated and therefore the average rate of forced call terminations due to g-type platforms defecting from the q-th position in the queue is

$$D(g,q) = \sum_{s \in H_{gq}} \mu_T(x_q(s)) \cdot p(s) \cdot z_q(s) \quad (q \neq 1) \quad . \quad (38)$$

The g-type platforms with i calls in progress arrive at the gateway at an average rate (platforms/sec) $\Lambda_h(g,i) = F(g,i) \cdot \Lambda_h$. Hence, the average hand-off *call* arrival rate on g-type platforms is

$$L(g) = \Lambda_h \cdot \sum_{i=1}^{N(g)} i \cdot F(g,i) \quad . \quad (39)$$

The call hand-off failure probability due to defections of g-type platforms can be expressed as a ratio of the number of calls terminated due to hand-off failure on g-type platforms to the total number of hand-off calls on g-type platforms impinging on the gateway. This probability can be expressed as

$$P_{H(g)}^A = \sum_{q=1}^Q D(g,q) / L(g) \quad (40)$$

The second component of hand-off failure probability was stated in (d). With a Poisson hand-off arrival process, an arriving hand-off on a g-type platform will find the target cell queue full with probability $P_{H(g)}^B$ given by

$$P_{H(g)}^B = \sum_{s \in H_F} p(s) \quad (41)$$

In (41) H_F is the set of permissible states for which the queue is full. That is, $H_F = \{s: z(s,Q) \neq 0\}$. It is worth noting that this component of P_H is independent of platform type considered. The hand-off failure probability for hand-off attempts on g-type platforms is

$$P_{H(g)} = P_{H(g)}^A + P_{H(g)}^B \quad (42)$$

Similarly, the overall hand-off failure probability can be found as

$$P_H = \frac{\sum_{g=1}^G \sum_{q=1}^Q D(g,q)}{\sum_{g=1}^G L(g)} + P_H^B \quad (43)$$

Forced termination probability.

Probability of hand-off failure gives the average fraction of hand-off attempts that fail. From the individual user's point of view a more interesting measure of performance is the probability that a call, which was not initially blocked, will be allowed to continue until satisfactory completion. *Forced termination probability* P_{FT} can be defined as a probability that a call, which was not blocked, will be interrupted in its lifetime due to hand-off failure. If we let $a(g)$ denote the probability that a call on g-type platform will make a hand-off attempt and will fail on that attempt. Similarly, $b(g)$ denotes the probability that a call on g-type platform will make a hand-off attempt and succeed. Using the Markovian properties of the model we get

$$a(g) = \mu_D(g) \cdot P_{Hg} / (\mu(g) + \mu_D(g)) \quad (44)$$

and

$$b(g) = \mu_D(g) \cdot (1 - P_{Hg}) / (\mu(g) + \mu_D(g)) \quad (45)$$

Assuming hand-offs independence we get

$$P_{FT}(g) = \sum_{i=0}^{\infty} a(g) \cdot b(g)^i \quad (46)$$

Summing (46) and using (44), (45) one finds

$$P_{FT}(g) = \mu_D(g) \cdot P_{Hg} / (\mu(g) + \mu_D(g) \cdot P_{Hg}) \quad (47)$$

Carried traffic.

The carried traffic on the g-type platforms per cell, $A_C(g)$, is the average number of channels occupied by calls supported by g-type platforms and is given by

$$A_C(g) = \sum_{s=0}^{s_{\max}} j(s,g) \cdot p(s) \quad (48)$$

The total carried traffic can be found as

$$A_C(\text{TOTAL}) = \sum_{g=1}^G A_C(g) = \sum_{s=0}^{s_{\max}} j(s) \cdot p(s) \quad (49)$$

DISCUSSION OF RESULTS

The approach described above was used to generate numerical performance results for a cellular system whose parameters are presented below. The number of channels per cell, C , is 20. There are three platforms types in the system ($G=3$). Type 1 is a low mobility platform, with one communication port per platform. Type 2 is a high mobility platform equipped with one communication port. Type 3 is a medium mobility platform, able to support up to two simultaneous calls. The platforms' parameters are summarized in the table below. In all example results the new call origination rate per *platform* was kept constant for any platform with an unoccupied port.

Type	N	\bar{T}	\bar{T}_T	\bar{T}_D	v_{g0}	α_{g1}
1	1	100s	100s	1000s	240	N/A
2	1	100s	20s	200s	120	N/A
3	2	100s	40s	400s	60	2.0

Table 1.

Fig. 1 and Fig. 2 show the behavior of the blocking and forced termination probabilities as the new call origination rate per port on a non-communicating platform is varied from $5.0E-05$ to $5.0E-04$ calls/s. Two values for the number of channels reserved for exclusive use by hand-offs were used, $C_h=0$ and $C_h=2$. The length of the queue's waiting space was stepped from $Q=0$ to $Q=2$ (platforms). As can be seen from Fig. 1 and Fig. 2 adding a hand-off waiting space of size 1 can improve the forced call termination probability by a factor of 3 or more for both $C_h=0$ and $C_h=2$. The improvement depends on the platform's mobility, the greatest improvement is for slow moving platforms. An additional improvement in forced termination probability can be obtained by reserving two channels for exclusive hand-off use (Fig. 3). Employing a hybrid channel reservation/hand-off queueing scheme decreases the hand-off failure probability by two orders of magnitude. However, while the improvement in forced termination probability due to hand-off queueing came at no noticeable cost in other performance measures, the improvement due to channel reservation for exclusive hand-off use caused an increase in blocking probability by an order of magnitude at low offered load and a drop in carried traffic at high offered load. Fig. 1 and Fig. 2 also show an additional improvement obtainable by adding one more waiting space (for a total of $Q=2$). The gain is most pronounced for low mobility platforms, with high mobility platforms benefiting very little. It can be concluded from Fig. 6 that carried traffic is highly insensitive to the size of the hand-off queue's waiting space.

Fig. 7, Fig. 8, Fig. 9 and Fig. 10 present the behavior of blocking probability, forced termination probability and carried traffic as a function of type-1 platform's increasing dwell time in the inner zone. The other dwell times were scaled along with $T_D(1)$ to preserve dwell time ratios calculated from table 1. Such scaling of parameters corresponds to changing of the cell size. The call origination rate Λ_0 was fixed at $2.75E-4$ and the unencumbered session duration remained unchanged at $T=100s$. Again two values for C_h were used, $C_h=0$ and $C_h=2$ as well as for Q , $Q=1$ (Fig. 7 and Fig. 8) and $Q=2$ (Fig. 9 and Fig. 10). A steady improvement in forced termination probability is observed as the dwell times increase (Fig. 7). The observed behavior is due to the declining probability of requesting a hand-off during a session. Blocking probability is insensitive to changes in dwell times, although there is a slight decrease in blocking as the dwell times become shorter. This behavior can be explained by noticing a rather steep rise in forced termination probability as dwell times are shortened. A large number of hand-off failures, with the new call origination rate practically constant, results in a lower total traffic impinging on the gateway and results in a lower blocking probability. A large number of hand-off failures is also a reason behind lower carried traffic at small dwell times as depicted in Fig. 8. For larger dwell times, the carried traffic remains constant. The

comparison of Fig.7 and Fig.9 reveals that increasing the size of the waiting space to $Q=2$ improves the P_{FT} consistently in a wide range of dwell time values.

The results of the study presented in the report support the conclusion that hand-off queueing allows significant improvement in forced termination probability. While improvements in forced termination probability attainable with hand-off queueing are lower than performance gains obtained by using a combined hand-off queueing/channel reservation scheme, hand-off queueing does not have the adverse effects on blocking probability and carried traffic associated with channel reservation for hand-offs.

REFERENCES

- [1] S.S. Rappaport, "The multiple call hand-off problem in high capacity cellular communication systems" IEEE Trans. Veh. Technol., vol.VT-40 ,pp.546-557, Aug.1991
- [2] S.S. Rappaport, "Traffic performance for cellular communication systems with mixed platforms and imperfect hand-off initiation", Tech. Report 660, College of Engineering and Applied Sciences, State Univ. of New York, Stony Brook, NY, 11794
- [3] S.S.Rappaport, "Modeling the hand-off problem in personal communication networks", IEEE Vehic. Tech. Conf. , VTC '91, St.Louis, pp.517-523
- [4] S.S.Rappaport, "Models for call hand-off schemes in cellular communication networks", in "Third generation wireless information networks", ed. Sanjiv Nanda and David J. Goodman pp.163-186, Kluwer Academic Publishers 1992.
- [5] D.Hong and S.S.Rappaport, "Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized hand-off procedures", IEEE Trans. Veh. Technol., vol. VT-35, pp.77-92, Aug. 1986
- [6] D.Hong and S.S.Rappaport, "Priority oriented channel access for cellular systems serving vehicular and portable radio telephones", Inst.Elec.Eng.Proc., pt.I, Communications, Speech and Vision, vol.136, pt.I, no.5, pp.339-346,Oct. 1989
- [7] R.Guerin, "Queueing-blocking system with two arrival streams and guard channels", IEEE Trans. Commun., vol.36, pp. 153-163, Feb. 1988.
- [8] S.A. El-Dolil, W-C. Wong and R.Steele, "Teletraffic performance of highway microcells with overlay macrocell" , IEEE J. Select. Areas Commun., vol. 7, no. 1, pp. 71-78, Jan. 1989.
- [9] A. Murase, I.C. Symington, E. Green, "Handover criterion for macro and microcellular systems", Proc. IEEE Vehicular Technology Conference, VTC '91, St. Louis, May 1991, pp.524-530.
- [10] S.T.S. Chia, "The control of handover initiation in microcells", Proc. IEEE Vehicular Technology Conference, VTC '91, St. Louis, May 1991, pp.531-536.
- [11] M. Gudmundson, "Analysis of handover algorithm", Proc. IEEE Vehicular Technology Conference, VTC '91, St. Louis, May 1991, pp.537-542.
- [12] O. Grimlund and B. Gudmundson, "Handoff strategies in microcellular systems", Proc. IEEE Vehicular Technology Conference, VTC '91, St. Louis, May 1991, pp.505-510.
- [13] J.M. Holtzman and R. Vijayan, "Analysis of hand-off algorithms using non-stationary signal strength measurements", Proc. IEEE Globecom '92, Orlando, Dec. 1992, pp. 1405-1409.

BLOCKING AND FORCED TERMINATION PROBABILITIES DEPEND ON CALL ORIGINATION RATE

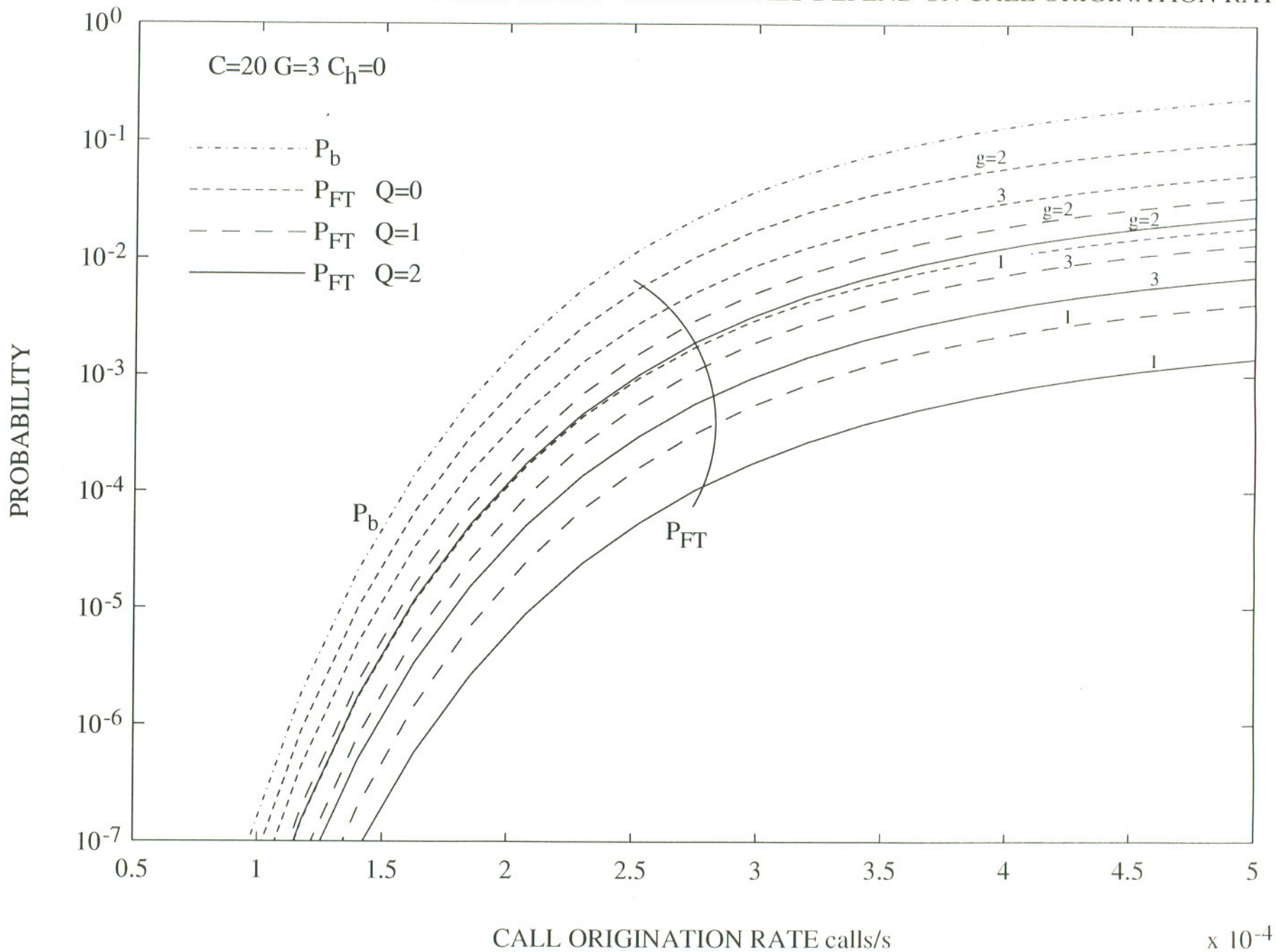
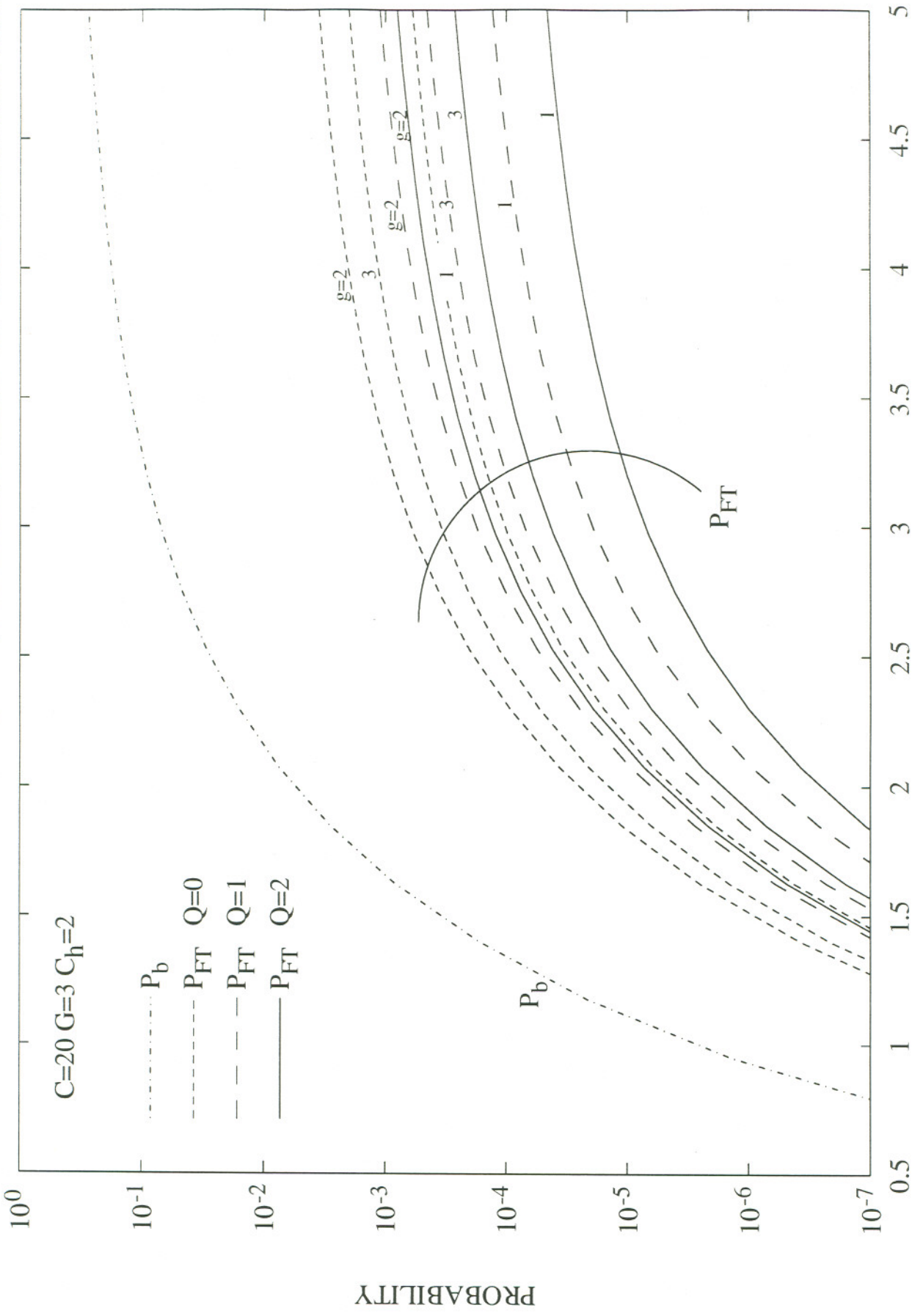


FIG.1

BLOCKING AND FORCED TERMINATION PROBABILITIES DEPEND ON CALL ORIGINATION RATE



CALL ORIGINATION RATE calls/s $\times 10^{-4}$

FIG.2

BLOCKING AND FORCED TERMINATION PROBABILITIES DEPEND ON CALL ORIGINATION RATE

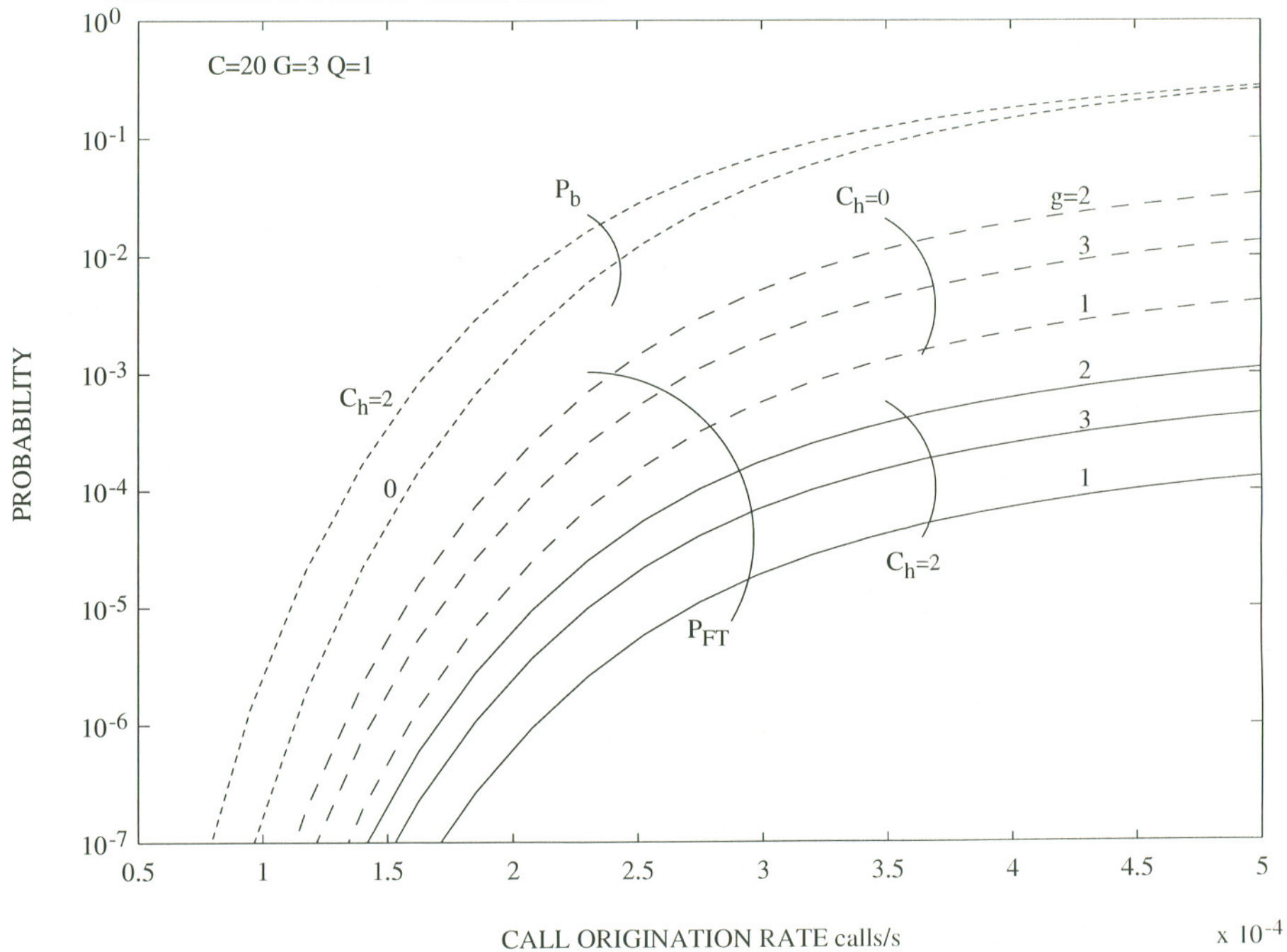


FIG.3

CARRIED TRAFFIC DEPENDS ON CALL ORIGINATION RATE

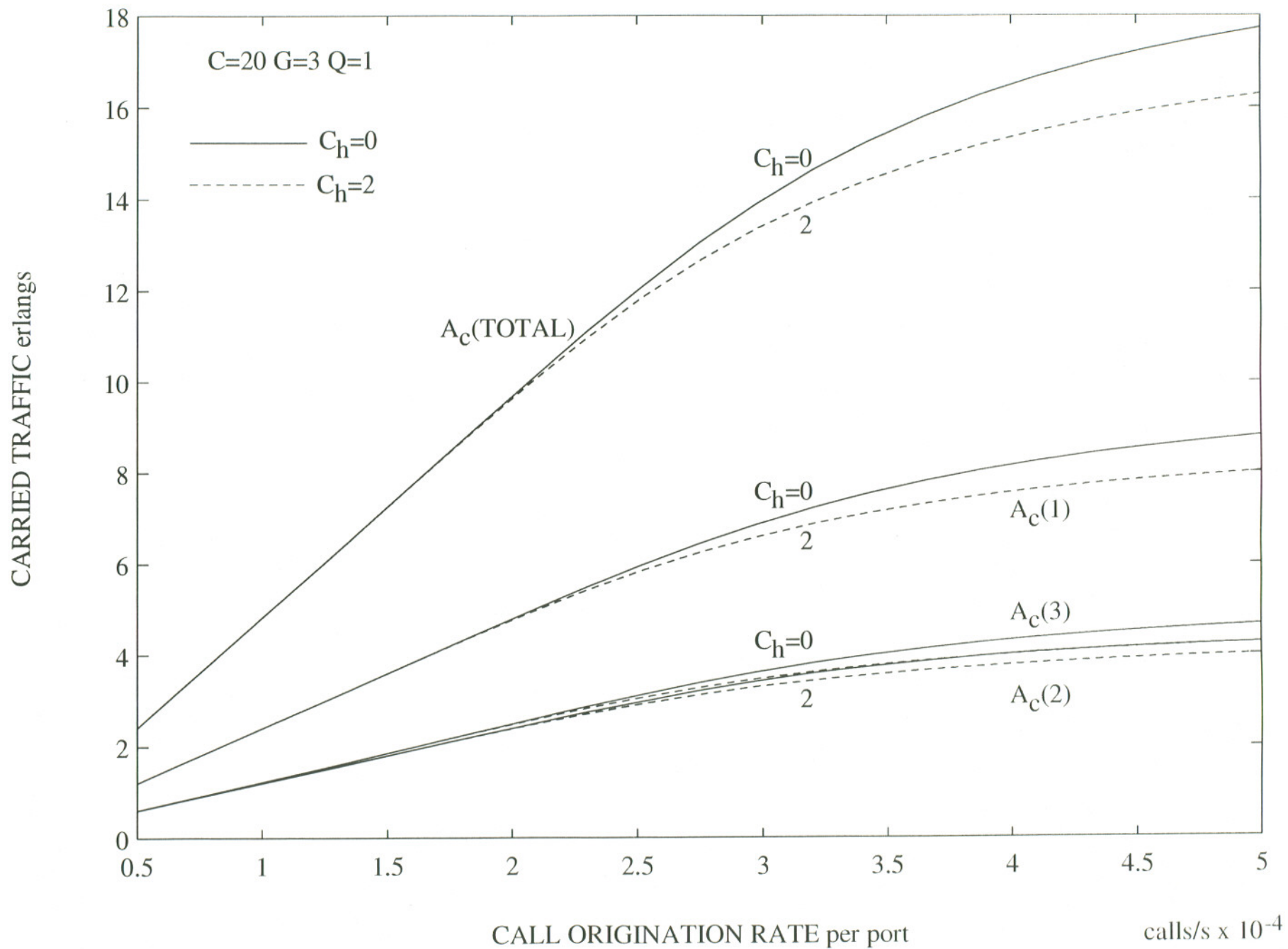


FIG.4

BLOCKING AND FORCED TERMINATION PROBABILITIES DEPEND ON CALL ORIGINATION RATE

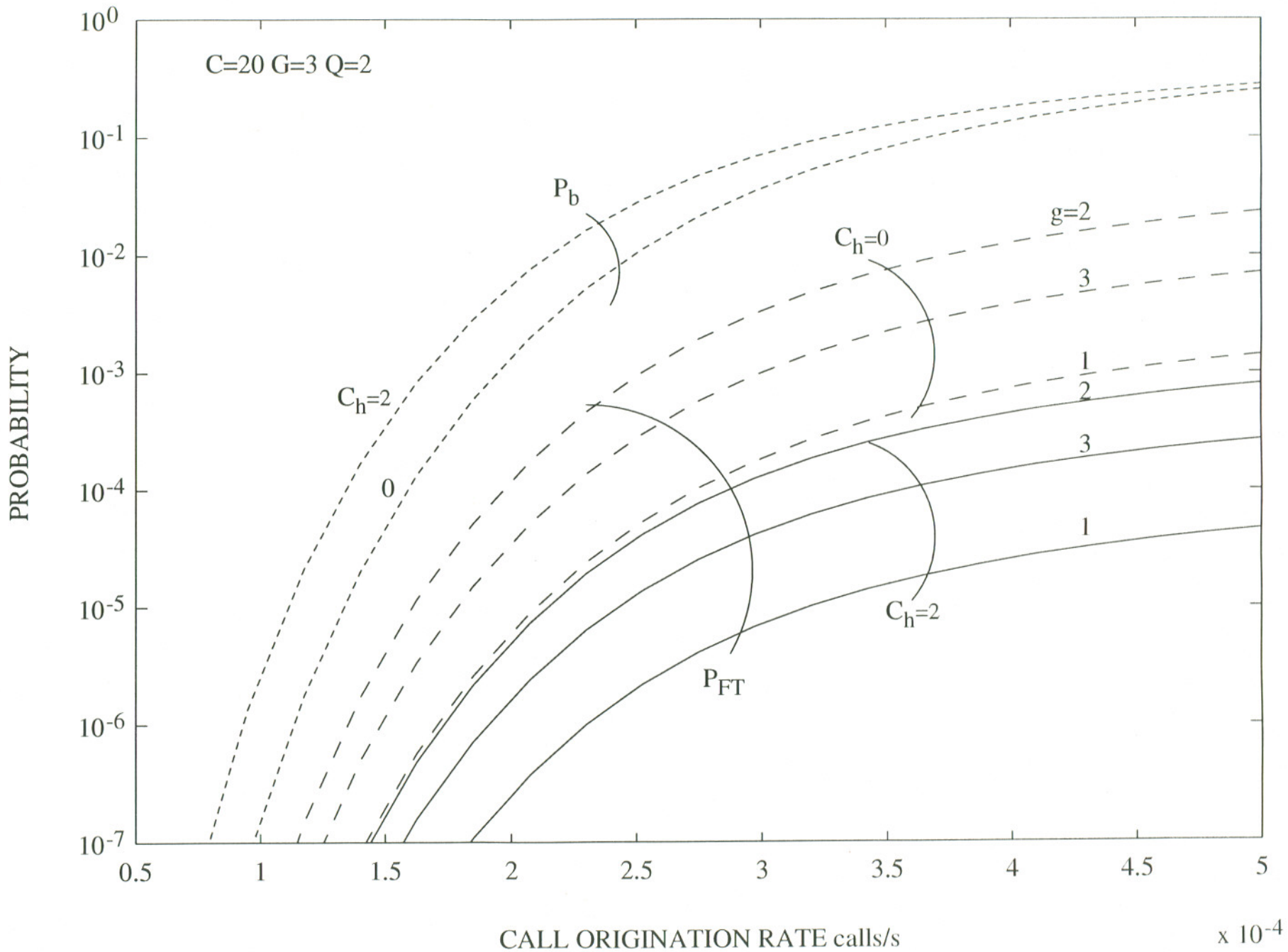


FIG.5

CARRIED TRAFFIC DEPENDS ON CALL ORIGINATION RATE

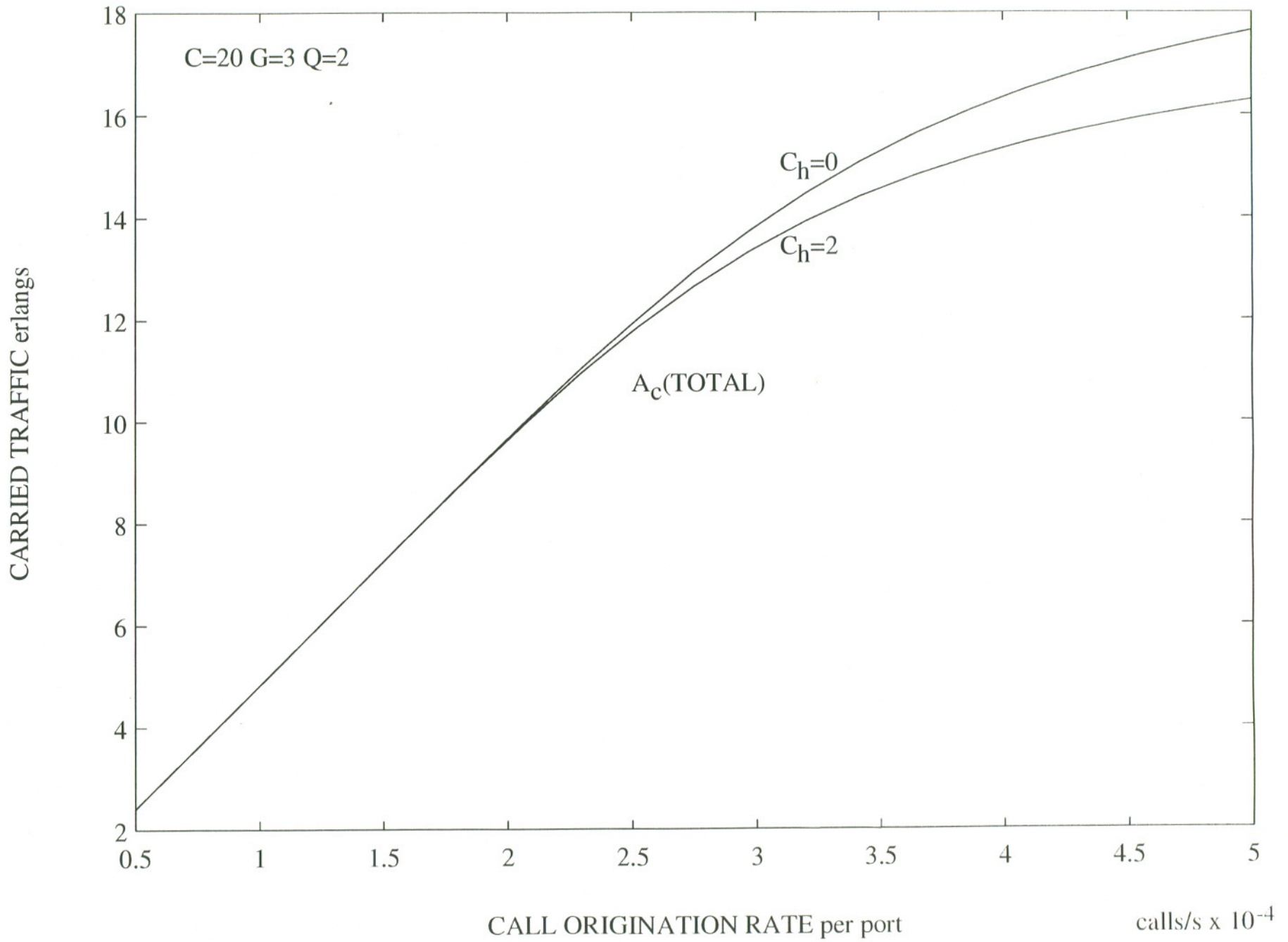


FIG.6

BLOCKING AND FORCED TERMINATION PROBABILITIES DEPEND ON DWELL TIMES

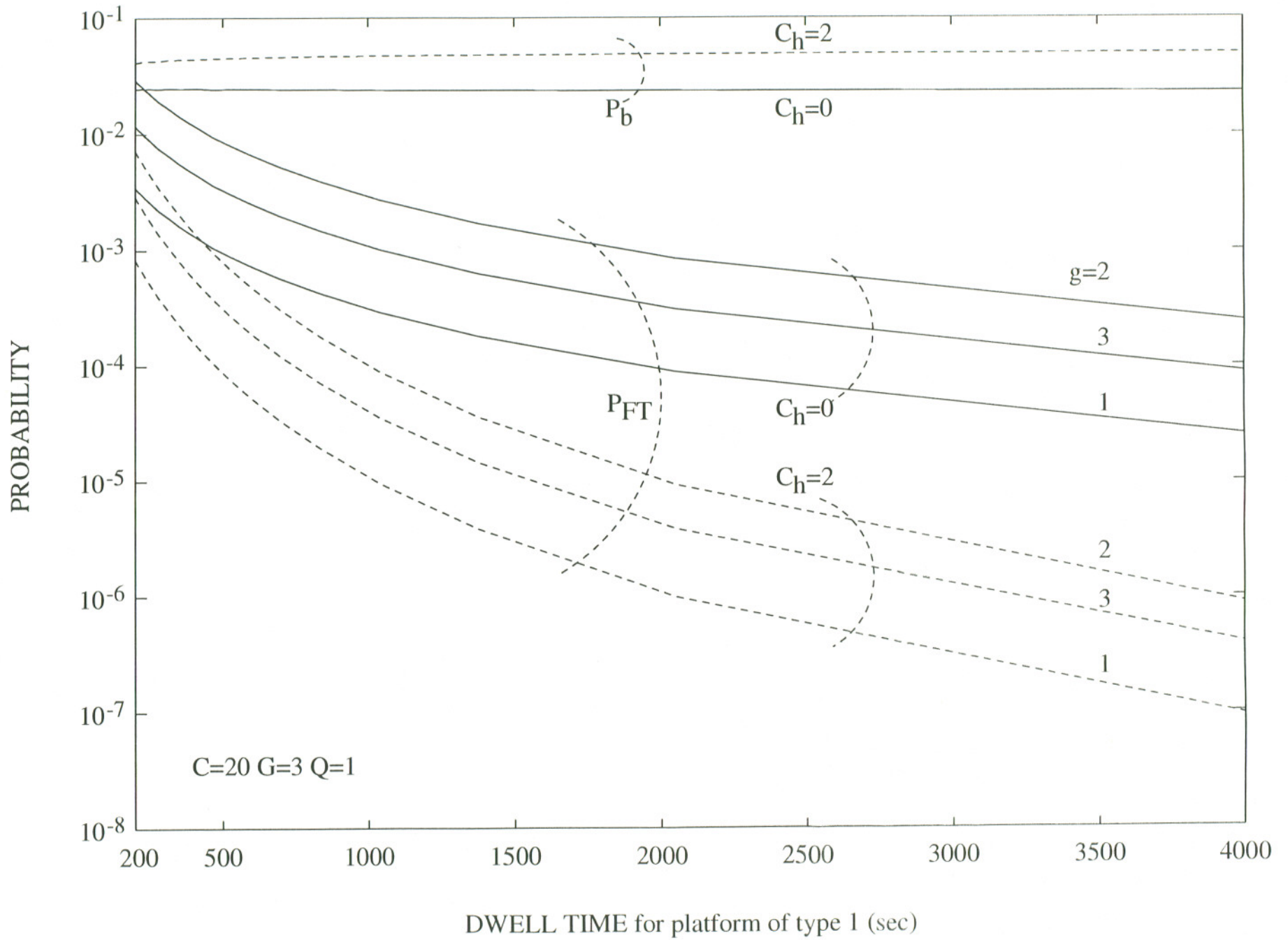


FIG.7

CARRIED TRAFFIC DEPENDS ON DWELL TIME

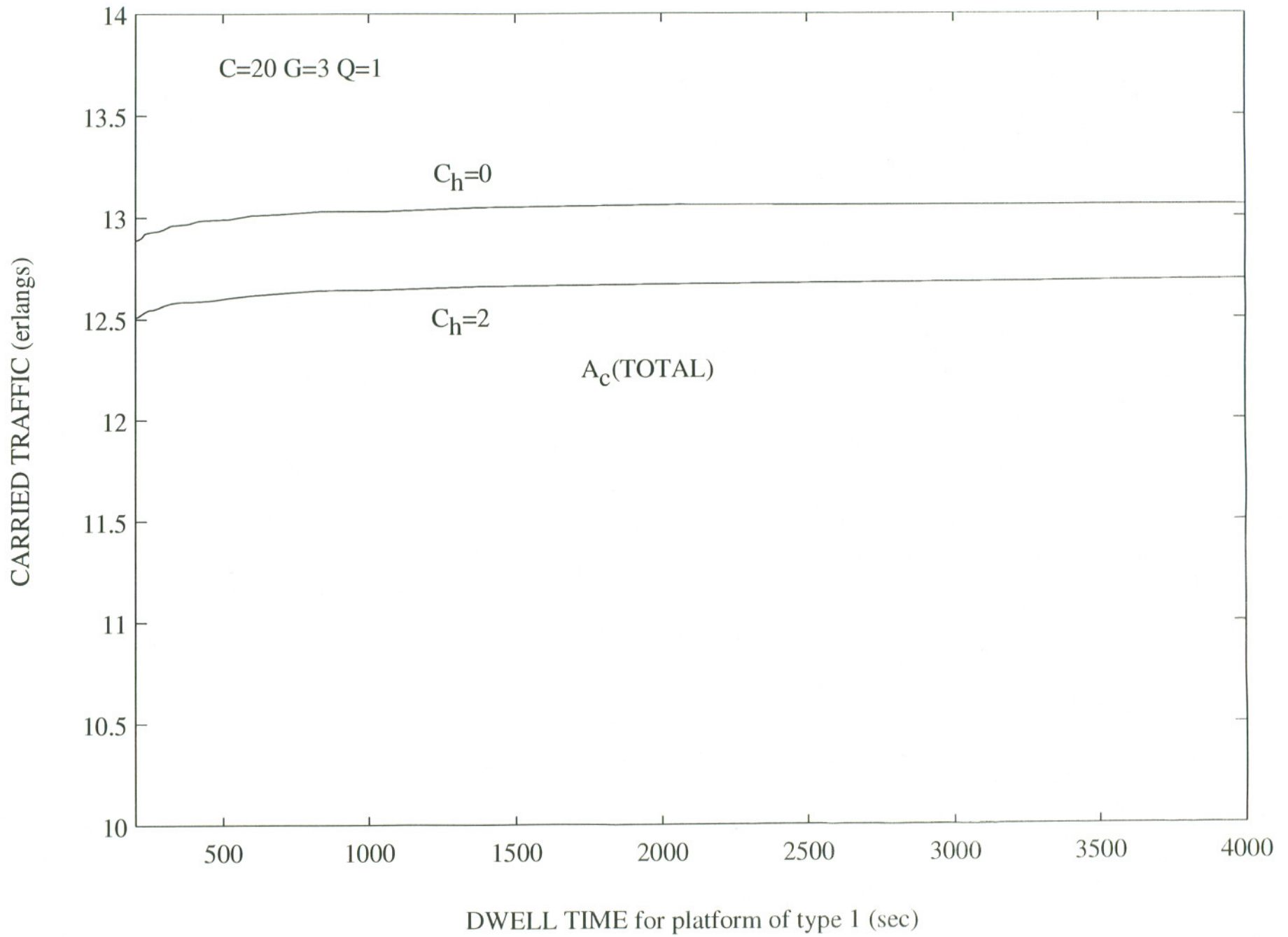


FIG.8

BLOCKING AND FORCED TERMINATION PROBABILITIES DEPEND ON DWELL TIMES

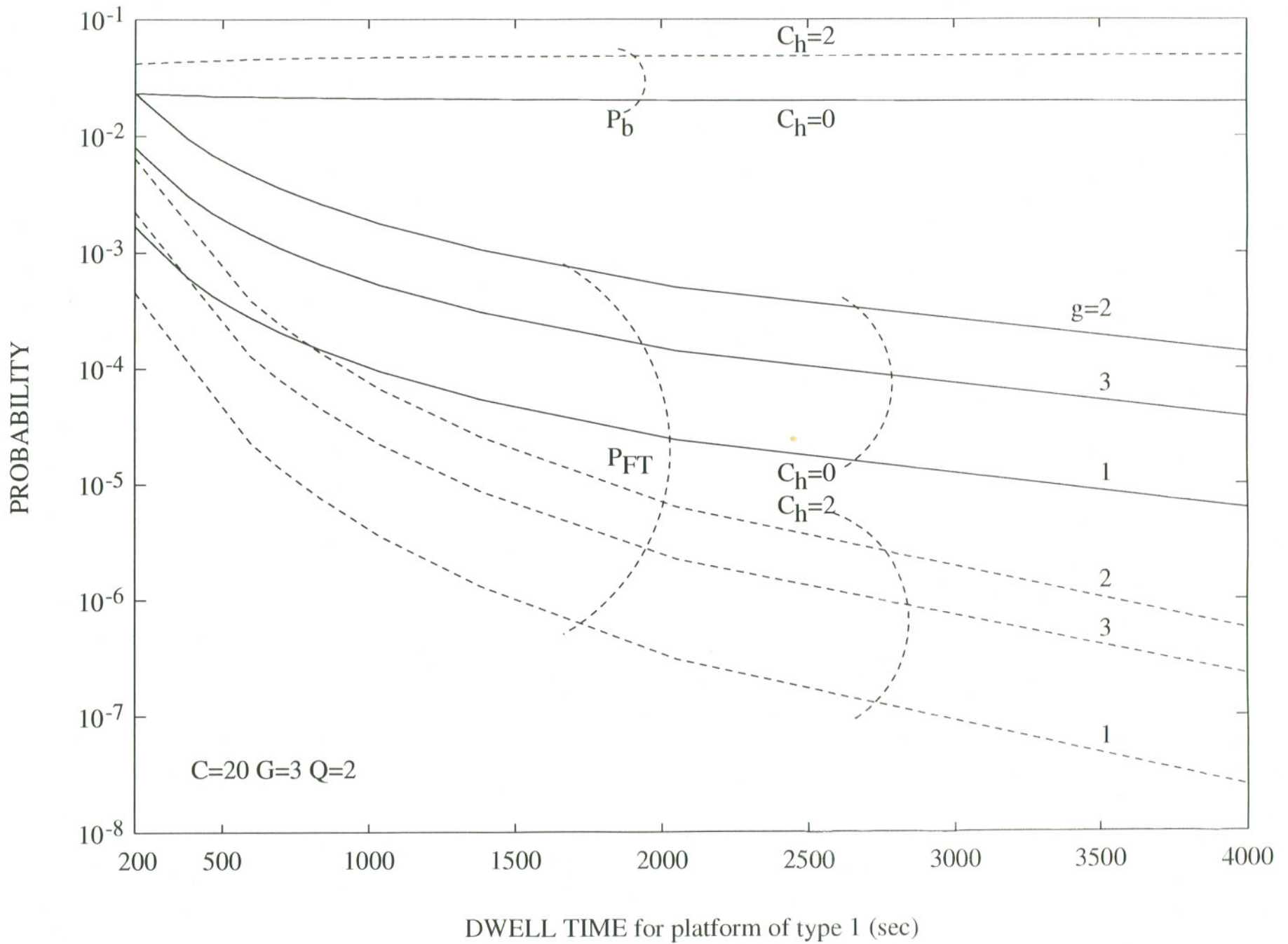


FIG.9

CARRIED TRAFFIC DEPENDS ON DWELL TIME

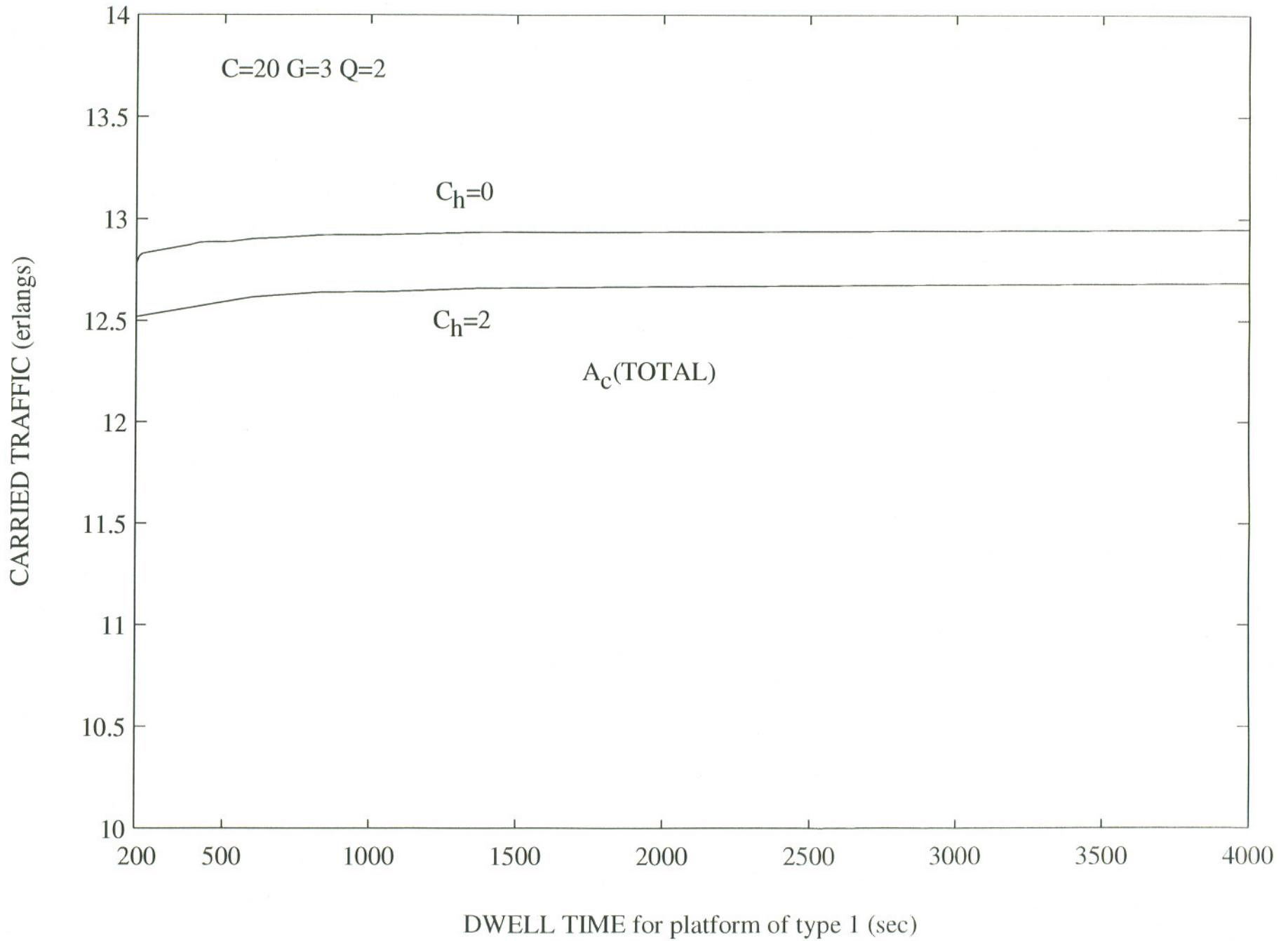


FIG.10