TRAFFIC PERFORMANCE OF CELLULAR COMMUNICATION SYSTEMS WITH MIXED PLATFORM AND CALL TYPES

Stephen S. Rappaport

Department of Electrical Engineering
State University of New York
Stony Brook, New York 11790-2350

phone: *(516) 632-8394*
fax: *(516) 632-8494*
e-mail: *rappaport@sbee.sunysb.edu*

ABSTRACT

The proliferation of mobile, portable and personal communication systems will bring a variety of offered services. Practical systems must accommodate different types of calls including voice only, mixed voice and data, high speed data, low speed data, and image transmission as well as a mixture of platforms (such as persons, autos, and buses) having a range of mobility characteristics. In such environments, the bandwidth and/or resources needed for call sessions will not be identical. As a result calls will generally encounter different blocking and hand-off constraints. These effects are in addition to differences in blocking and forced (call) termination probabilities that are attributable to differing platform mobilities and (resource) channel quotas.

A cellular system with mixed platforms and call types is considered. We identify a suitable state characterization for the problem and demonstrate a framework for performance analysis. The model is used to generate example performance characteristics. These show carried traffic, blocking probability, and forced termination probability for each platform type and for each call type. The example results are discussed.

The author is grateful to Gustavo MONTE and Hua JIANG whose skillful computer programming, generated the performance characteristics presented herein.

The research reported in this paper was supported in part by the U.S. National Science Foundation under Grant No. NCR 9025131 and in part by IST/SDIO under Grant No. N0001491-J4063 administered by the U.S. Office of Naval Research.

INTRODUCTION

Practical cellular mobile communication systems currently envisioned must support a mixture of platform types having a range of mobility characteristics as well as a variety of services, including voice only, mixed voice and data, image transmission, and various data rates. Call sessions will require different types and amounts of network resources such as radio bandwidth, buffer allocations, and performance monitoring and call supervising processors. As a result they will encounter different blocking and hand-off constraints. These are in addition to differences that are attributable to differing mobilities and quotas that are associated with the various types of platforms which can be present in the system.

The development and evaluation of alternative network control strategies requires characterization of communications traffic performance and the development of tractable models that can accommodate a fair amount of physical complexity. The complexity includes heterogeneous call types and platform types in the same system. In recent work we have been developing a framework that is useful for tele-traffic performance analysis and modeling for a broad class of problems that arise in the context of cellular communication architectures. The approach, which is based on multi-dimensional birth-death processes is rich enough to permit modeling of many practical issues, and allows computation of theoretical performance characteristics. In particular we have devised tele-traffic performance models for cellular systems with the following features: single and multiple call platforms and hand-offs, priority and no priority for hand-off calls, mixtures of platforms having different mobility characteristics, lost call systems, delayed call systems, combined delay and loss systems, platform limits and quotas, channel limits and quotas, a broad class of platform mobility characteristics, and mixed macro and micro-cell configurations [2],[4],[6],[7],[8],[13],[19]. Related work includes [1],[3],[5],[10],[11],[12]. The reader should also note the nature of the hand-off issues which are considered in this paper whose focus is on *resource allocation and availability*. This is in contrast to another aspect of the hand-off problem which is concerned with *how the hand-off process is initiated*. This latter aspect of the problem has received some attention in the literature [14],[15],[16],[17],[20]. Reference [19] suggests an approach to combining these aspects of the problem.

In this paper, we mathematically formulate the problem outlined (in the first paragraph) above in a format that makes it amenable to solution using the approach that we are developing. In particular we consider a cellular system with mixed platform and call types. As a communicating platform moves out of range of the base station to which it is linked, a hand-off to an alternative base must be made. A hand-off attempt that fails results in a forced termination of the session. Since forced terminations are more obtrusive than (new) call blocking we also consider priority for hand-off attempts using a cut-off priority scheme. Other priority schemes are possible. We identify a suitable state characterization for the problem and demonstrate how the framework that we are developing can be applied. The model is used to generate theoretical performance characteristics. These show carried traffic, blocking probability, and forced termination probability for each platform type and for each call type. The example results are discussed.

DESCRIPTION OF THE ANALYTICAL APPROACH

The approach that we use requires a suitable state characterization which allows the problem to be considered in the framework set forth in [6],[7],[8],[13]. A cell state is described by a concatenation of integer state variables. Permissible states correspond to those concatenations whose elements satisfy certain constraints determined by the system resources. The underlying driving processes are identified. In the present context these include new call arrivals, call completions, hand-off call arrivals to a cell and hand-off departures from a cell. These processes are in general multi-dimensional. In the problem under discussion, for example, each of the driving processes is subdivided according to platform type and call type. Markovian assumptions for the driving processes are invoked. Because of hand-offs, the states of adjacent cells are coupled. A more complete characterization of the system would require consideration of the *system state* as a concatenation of all the cell states. The dimensionality of the problem is prohibitive even for modest system parameters [7]. As part of the overall methodology this difficulty is circumvented by use of a *conservation law* which relates the mean hand-off arrival rates (to a cell) to corresponding mean hand-off departure rates from a cell. This allows a decoupling of driving processes and a reduction in the number of states need to characterize performance. In particular it allows characterization for any cell without the mathematical encumbrance of accounting for the states of all cells simultaneously. The (cell) state transition flows are related to the underlying driving processes and the equilibrium state probability flow balance equations are formulated.

An algorithm is developed to numerically solve the resulting system of equations for the equilibrium state probabilities. Our algorithm used three levels of nested iterations based on Gauss-Seidel iteration, the bisection method, and successive substitutions. Details are given in [6] and [13]. Once the state probabilities are found for given system parameters, the various traffic performance measures are calculated.

PROBLEM DESCRIPTION

The development presented here proceeds along the lines of [6] and [13]. We present a brief description of the problem and then proceed with the mathematical formulation. We consider a large geographical region covered by cells, that are defined by proximity to designated network gateways. The region is traversed by large numbers of mobile platforms that are of several types. The platform types differ primarily in their mobility characteristics. Pedestrians with hand-held devices and autos with cellular phones are example platform types. Communication with a mobile platform is via a wireless base station. This is a gateway node (or several nodes) which define and are identified with each cell. These base stations (or gateways) allow communications between the system's radio segment and its fixed or wireline segment. We use the word *cell* in its generic sense to describe a spatial region serviced by a wireless gateway. The essential problem considered here is the same whether this is a macro-cell, micro-cell, pico-cell, zone, sector, or satellite beam. Thus the particular configuration is not very important in terms of demonstrating the applicability of the approach. References [6] and [7] contain various figures that suggest example configurations to which the analysis applies. In the development presented here, a platform can support at most one call. However, there are different call types each of which requires a (generally) different amount of communications resources when supported by the network. One of the resources is bandwidth (measured in appropriate units such as channels). Each supported call also needs access to a modem at the supporting gateway. Buffer space and call supervising processors are examples of other resources that may be required.

The wireless links can employ radio, optical, infra-red or acoustic signaling, and the multiple access scheme can be FDMA, TDMA, CDMA, or any hybrid. Channels can be organized using any mixture of frequency, time, space, and code division techniques - including hybrid schemes. However, channels that are simultaneously used in the same zone must be sufficiently separated by the time-space-frequency-code multiplex to allow acceptable communications on each. Circuit or virtual circuit switching is used so that the system operates by reserving some communications resources for any call (session) in progress.

We assume that there are G platform types, labeled g=1,2,...G, and that there are C channels assigned to each gateway. A cut-off priority scheme is used. That is, each gateway keeps C_h channels for use by hand-off calls. Specific channels are not reserved, just the number. In this way hand-off calls have access to more channels than new calls do, and increasing C_h provides increasing priority for hand-offs at the expense of blocking new call originations. Thus forced termination and blocking performance can be exchanged. For convenience we assume that the system is homogeneous. That is cells, gateways, and the respective driving processes that impinge upon each are statistically identical. Non-homogeneous systems can be considered essentially in the same way [6],[13],[19]. At each gateway there may be channel quotas, so that no more than J(g) channels can be occupied by platforms of type g at the same time in the same cell. Additional constraints on other network resources such as buffer space, directional radio beams, and call supervising processors may be present as well. In addition to quotas, which constrain use of network resources according to platform type, there are also limits on the amounts of resources available at a gateway. In general there can be I call types, labeled i=1,2,...I, and K kinds of resources labeled k=1,2,...K. We let r(g,i,k) denote the amount of resource k that is needed to support a call of type i on a platform of type g.

EXAMPLE PROBLEM STATEMENT - SINGLE CALL HAND-OFFS, CUT-OFF PRIORITY, MIXED PLATFORM TYPES, MIXED CALL TYPES.

There are G types of mobile platforms, indexed by g=1,2,...G.

No platform can support more than one call at any given time.

There are I types of calls, indexed by i=1,2,...I.

There are K types of resources to support calls. These resources are labeled, R_k, k=0,1,2,...K-1.

The new call origination rate *from a non-communicating g-type platform* is denoted $\Lambda(g,i)$.
We define $\alpha(g,i)= \Lambda(g,i) / \Lambda(1,i)$.

The number of *non-communicating* g-type platforms in any cell is denoted v(g,0). The total rate at which new calls of type i are generated from platforms of type g *in a cell* is denoted $\Lambda_n(g,i)$.
Thus, $\Lambda_n(g,i) = \Lambda(g,i) \cdot v(g,0)$.

Note: It is assumed that $v(g,0) >> C$, so that overall the population of non-communicating g-type platforms in a cell generates $\Lambda_n(g)$ calls per second. This *infinite* population model is consistent with a large population of non-communicating g-type platforms in each cell, only a small fraction of which are served at any time. This is, in fact, usually the case.

CHANNEL LIMIT: Each cell or gateway has C channels.
RESOURCE LIMITS: Each cell or gateway has R_k units of resource k.

> NOTE: The number of channels is in fact just another resource. So we can consider for example, $R_0=C$. But in the present discussion, for the sake of clarity, we consider the channel resource by name.

CHANNEL QUOTAS: At any gateway, the maximum number of channels that can be simultaneously used by g-type platforms is J(g).

RESOURCE QUOTAS: At any gateway, the maximum number of resources of type k, that can be simultaneously used by g-type platforms is R(g,k).

CUT-OFF PRIORITY: C_h channels in each cell are reserved for hand-off calls. Specific channels are not reserved, only the number C_h.
New calls will be blocked if the number of channels in use is $C-C_h$ or greater.
Hand-off attempts will fail if the number of channels in use is C.

We define the unencumbered call (session) duration as the time a call would remain in progress if it were not forced to terminate. For a call of type i this is taken as a negative exponentially distributed (ned) random variable, T(i), having a mean $\bar{T}(i) = 1/\mu(i)$.

The dwell time in a cell for a g-type platform is a ned random variable, $T_D(g)$ having a mean $\bar{T}_D(g)$.

The problem is to calculate relevant performance characteristics, including, for each platform type, and for each call type: Blocking Probability, Hand-off Failure Probability, Forced Termination Probability, Carried Traffic (or bandwidth utilization), and Hand-Off Activity.

Notes: We consider blocking probability to be the average fraction of ***new*** call originations that are denied access to a channel. Hand-off failure probability is the average fraction of hand-off "needs" that fail to gain access to a channel in the target zone. Forced termination probability is the probability that a call will suffer a hand-off *attempt* failure some time in the "lifetime" of the call. Hand-Off Activity is the average number of hand-off attempts for a call that receives service.

The mathematical analysis is similar to that used in [6],[7],[8],[13]. The major differences are in definition of the state variables, identification of the driving processes, and the equations that specify the state probability transition flows. In what follows we emphasize those aspects of the mathematical development that differ.

STATE CHARACTERIZATION

First consider a *single* cell. We define the state (of a cell) by a sequence of non-negative integers. This can be conveniently written as G n-tuples.

$$\begin{array}{lllll}
v_{11}, & v_{12}, & v_{13}, & \cdots & v_{1N(1)} \\
v_{21}, & v_{22}, & v_{23}, & \cdots\cdots\cdots\cdots & v_{2N(2)} \\
\vdots & \vdots & \vdots & & \vdots \\
v_{g1}, & v_{g2}, & v_{g3}, & \cdots\cdots\cdots\cdots\cdots\cdots & v_{gN(g)} \\
\vdots & \vdots & \vdots & & \vdots \\
v_{G1}, & v_{G2}, & v_{G3}, & \cdots\cdots\cdots\cdots\cdots\cdots\cdots & v_{GN(G)}
\end{array} \tag{1}$$

where v_{gi} { $g=1,2,...G;\ i=1,2,...I$ } is the number of platforms of type g that have a call of type i in progress. It is convenient to order the states using an index $s=0,1,2,...s_{max}$. Then the state variables v_{gi}, can be shown explicitly dependent on the state. That is, $v_{gi} = v(s,g,i)$.

When the cell (gateway) is in state, s, the following characteristics can be determined:

The number of channels being used by g-type platforms is

$$j(s,g) = \sum_{i=1}^{I} v(s,g,i) \quad . \tag{2}$$

The total number of channels in use is

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

The quantity of resources of type k being used by platforms of type g is

$$r(s,g,*,k) = \sum_{i=1}^{I} r(s,g,i,k) \tag{4}$$

in which $r(s,g,i,k)$ is the quantity of resource of type k that is being used by g-type platforms with i-type calls when the cell is in state s.

Note: It is reasonable (but not necessary) that $r(s,g,i,k)$ be proportional to $v(s,g,i)$. That is $r(s,g,i,k) = r(g,i,k) \cdot v(s,g,i)$. If this is the case, $r(s,g,*,k)$ is a linear combination of certain state variables.
For example,

$$r(s,g,*,k) = \sum_{i=1}^{I} r(g,i,k)\ v(s,g,i) \quad . \tag{5}$$

The quantity of resources of type k being used by calls of type i is

$$r(s,*,i,k) \quad = \quad \sum_{g=1}^{G} r(s,g,i,k) \qquad . \qquad (6)$$

The total quantity of resources of type k that is being used is

$$r(s,*,*,k) \quad = \quad \sum_{g=1}^{G} r(s,g,*,k) \qquad . \qquad (7)$$

Permissible states correspond to those sequences for which all constraints are met. Additional constraints can also be considered within this same framework. Here we have a *channel limit* which requires, $j(s) \leq C$, and *channel quotas* which require, $j(s,g) \leq J(g)$, for g=1,2,...G. In addition there are *resource limits* which require

$$r(s,*,*,k) \leq R_k \;, \quad k=1,2,...K. \qquad (8)$$

and *resource quotas* which require

$$r(s,*,g,k) \leq R_{gk} \;, \; g=1,2,...G; \; k=1,2,...K. \qquad (9)$$

in which R_{gk} denotes the maximum amount of resources of type k that the system will allow to be used by g-type platforms at any time. There can also be additional limit and quota constraints. For example, we note that the total number of platforms of type g that can be supported at a cell at any time can be constrained by V_g, g=1,2,....G. If we let w(s,g) denote the total number of g-type communicating platforms when the cell is in state s, we have

$$w(s,g) = \sum_{i=1}^{I} v(s,g,i) \qquad . \qquad (10)$$

Note: In this example w(s,g) is the same as j(s,g) given by (2) because no communicating platform can support more than one call at a time.

The set of permissible cell states corresponds all possible sequences of variables of the form (1) which satisfy all the limit and quota constraints.

A thorough formulation which accounts for direct coupling of ***cell state*** transitions of adjacent cells that are involved in a hand-off is circumvented by relating the average hand-off arrival and departure rates as in [6] and [7]. This avoids having to deal with an enormously (and usually intractably) larger number of ***system states*** represented by sequences of all simultaneously possible ***cell states***. Both *homogeneous* and *non-homogeneous* cellular systems can be treated in this way. The result is that we only have to consider a ***single cell*** and deal with the number of states needed to characterize its behavior.

The stated problem was solved using the general framework outlined in our previous work. The underlying driving processes for this problem are new call arrivals, (successful) call completions, and the dwell time(s) in a cell for each platform type. The mean hand-off arrival rates to a cell for calls of type i on platforms of type g is dependent on these other processes. The solution approach is to identify permissible cell states, relate the parameters of the underlying driving processes to the state probability transition rates, and solve the resulting set of equations for the state probabilities. The approach requires an iterative solution because of the coupling between the underlying driving processes and the hand-off arrival processes.

RESULTS

The approach was used to consider a problem with G=3 platform types, stationary platforms, low speed platforms, and high speed platforms. (Perhaps corresponding to stationary pedestrians, moving pedestrians and vehicles.) In addition we considered I=2 call types, low bandwidth calls, and high bandwidth calls. (Perhaps corresponding to low speed data, and voice plus high speed data.) High bandwidth calls were assumed to require 3 bandwidth units (channels) each while low bandwidth calls require only one bandwidth unit each. High speed platforms were assumed to move 5 times faster than low speed platforms. Of course the stationary platforms do not move. The only resource considered was bandwidth (measured in channels). No quota constraints were considered. The only resource limit for this example is the number of channels, C, which was taken as C=24. Channel access can be any FDMA or TDMA arrangement or hybrid. With proper interpretation the model also represents CDMA systems. The parameter C would then represent the amount of base station resources (modems, codecs) that are available. Different amounts of these resources would be engaged by high bandwidth and low bandwidth calls). In the example, new call arrival rates for all *platform types* were assumed to be the same, and new call arrival rates for all *call types* were assumed to be the same.

The parameter choices for this example were chosen only to demonstrate the applicability of the modeling approach and framework. The reader should not infer any loss in generality of the approach because of the example presented here.

Figure 1.a shows overall blocking and forced termination probabilities for different platform types plotted as a function of demand (from a single platform). The possible exchange (increase) of blocking probability for a decrease in forced termination probability can be seen as C_h increases from 0 to 4. Of course blocking and forced termination probabilities increase as demand increases. The overall forced termination probabilities were determined by calculating the forced termination probability for platform-call types (g,i) and using a weighted average according to the fraction of carried traffic that is of type (g,i).

Figure 1.b shows *forced termination probabilities* for high speed and low speed platforms on a similar plot. The improvement attainable by increasing C_h is shown.

Figure 1.c shows *blocking probabilities* for different call types. Since high bandwidth calls need more resources to obtain service, they experience more greater blocking probability when other parameters are the same. Also the increase in blocking (for each call type) as C_h is increased to favor hand-off calls can be seen.

Figure 1.d shows *forced termination probabilities* for different call types on a similar plot. The improvement obtainable by increasing C_h can be seen.

Figure 2 shows bandwidth utilization for various traffic components for $C_h=2$. Similar plots for other values of C_h can be obtained.

Figure 3 shows blocking and forced termination probabilities as a function of dwell time means. Because the dwell times are held to a fixed ratio, one can consider the abscissa, which is labeled as dwell time, to be proportional to cell size (radius). The figure shows the impact on performance measures as the abscissa (envisioned either as dwell time or as cell radius) increases.

Figure 4 shows a plot of bandwidth utilization for various traffic components. For small values of the abscissa (dwell times) (cell radius) bandwidth utilization by high speed platforms is decreased since such platforms are more likely to encounter forced terminations.

Figure 5 shows bandwidth utilization for various traffic components as a function of call demand.

REFERENCES

[1] D.C. Cox and D.O. Reudink, "Layout and control of high-capacity systems," Chap. 7 in *Microwave Mobile Communications*, ed. by W.C. Jakes, Jr., Wiley:New York, 1974, pp. 545-622.

[2] D. Hong and S.S. Rappaport, "Traffic model and performance analysis for cellular mobile radiotelephone systems with prioritized and non-prioritized hand-off procedures," IEEE Trans. Vehic. Technol., vol. VT-35, pp. 77-92, Aug. 1986.

[3] 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.

[4] D. Hong and S.S. Rappaport, "Priority oriented channel access for cellular systems serving vehicular and portable radio telephones," IEE (British) Proc., Part I, Commun., Speech and Vision, vol. 136, pt. I, no. 5, pp. 339-346, Oct. 1989.

[5] R. Guerin, "Queueing-blocking system with two arrival streams and guard channels," IEEE Trans. Commun., vol. COM-36, pp. 153-163, Feb. 1988.

[6] S.S. Rappaport, "The multiple-call hand-off problem in high-capacity cellular communications systems," IEEE Trans. Vehic. Technol., vol. 40, no. 3, pp. 546-557, Aug. 1991 .

[7] S. Rappaport, "Models for call hand-off schemes in cellular communication networks," in *Third Generation Wireless Information Networks*, ed. by S. Nanda and D.J. Goodman, Kluwer Academic Publishers: Boston, 1992, pp. 163-185.

[8] S.S. Rappaport, "Modeling the hand-off problem in personal communications networks," Proc. IEEE Vehicular Technology Conference, VTC '91, St. Louis, May 1991, pp. 517-523.

[9] R.B. Cooper, *Introduction to Queueing Theory, 2^{nd} ed.,* Elsevier North Holland:New York, 1981.

[10] R. Steele, "Deploying personal communication networks," IEEE

Communications Magazine, September 1990, pp. 12 ff.

[11] D.C. Cox, "Personal communications - a viewpoint," IEEE Communications Magazine, November 1990, pp. 8 ff.

[12] A.J. Viterbi, "Wireless digital communication: a view based on three lessons learned," IEEE Communications Magazine, September 1991, pp. 33 ff.

[13] S.S. Rappaport, "Blocking, hand-off and traffic performance for cellular communication systems with mixed platforms and imperfect hand-off initiation," CEAS Technical Report no. 660, College of Engineering and Applied Sciences, State University of New York, Stony Brook, New York 11794-2350, November 26, 1991, Appendix.

[14] A. Murase, I.C. Symington, and E. Green, "Handover criterion for macro and microcellular systems," Proc. IEEE Vehicular Technology Conference, VTC '91, St. Louis, May 1991, pp. 524-530.

[15] 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.

[16] M. Gudmundson, "Analysis of handover algorithm," Proc. IEEE Vehicular Technology Conference, VTC '91, St. Louis, May 1991, pp. 537-542.

[17] O. Grimlund and B. Gudmundson, "Handoff strategies in microcellular systems," Proc. IEEE Vehicular Technology Conference, VTC '91, St. Louis, May 1991, pp. 505-510.

[18] F.P. Kelly, *Reversibility and Stochastic Networks,* Wiley:New York, 1979.

[19] S.S. Rappaport, "Communications traffic performance for cellular systems with mixed platform types," Workshop Record, Winlab Workshop on Third Generation Wireless Information Networks," East Brunswick, NJ, April 28-29, 1992, pp. 131-158. Available from Rutgers University, WINLAB, Box 909, Piscataway, NJ 08855-0909, USA.

[20] 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.

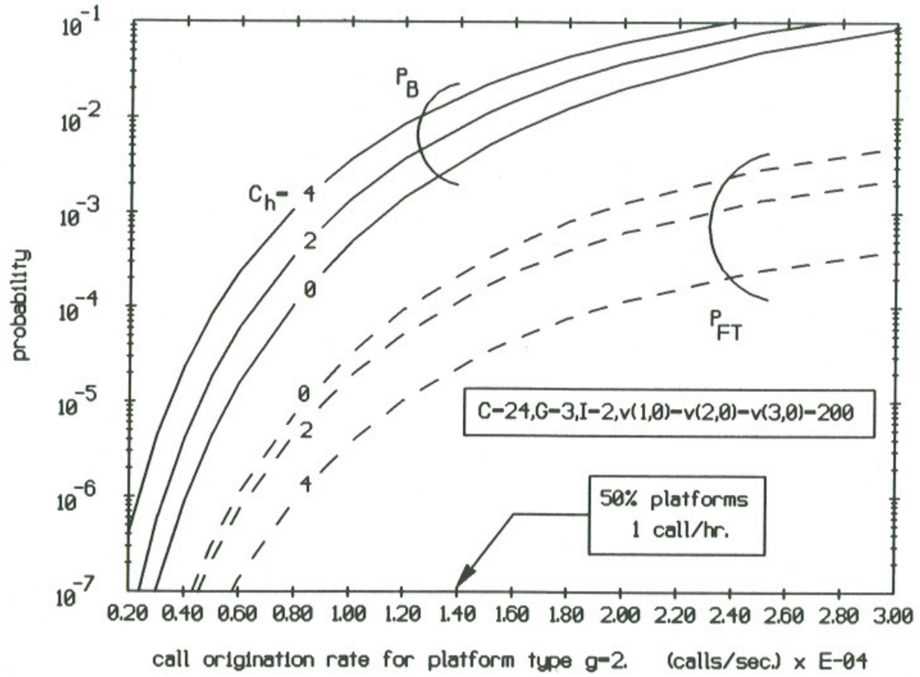

Figure _.1.a Overall Blocking and Forced Termination Probabilities

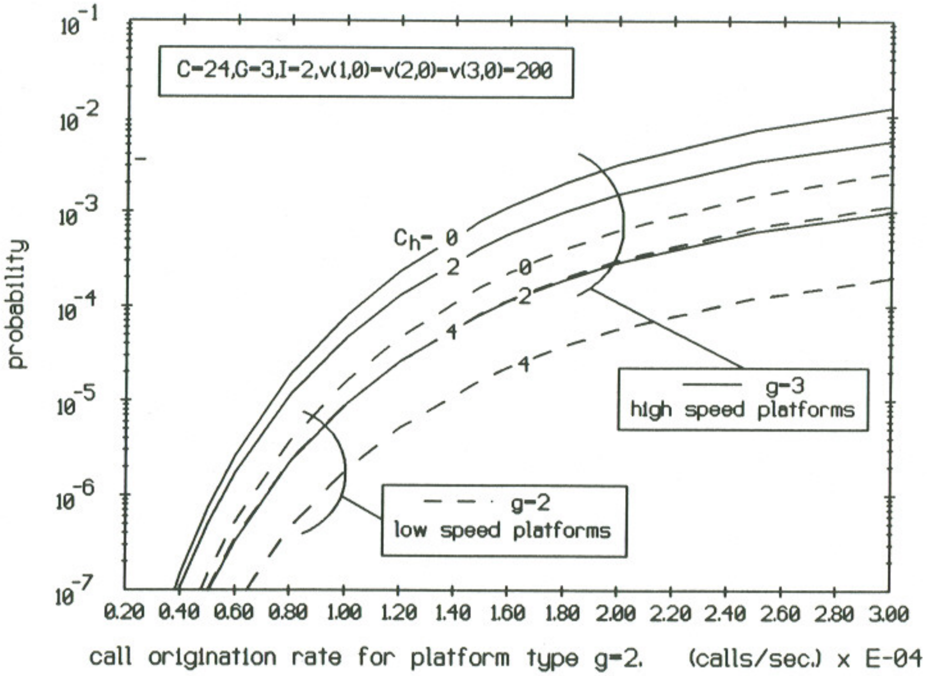

Figure _.1.b Forced Termination Probabilities
for Different Platform Types

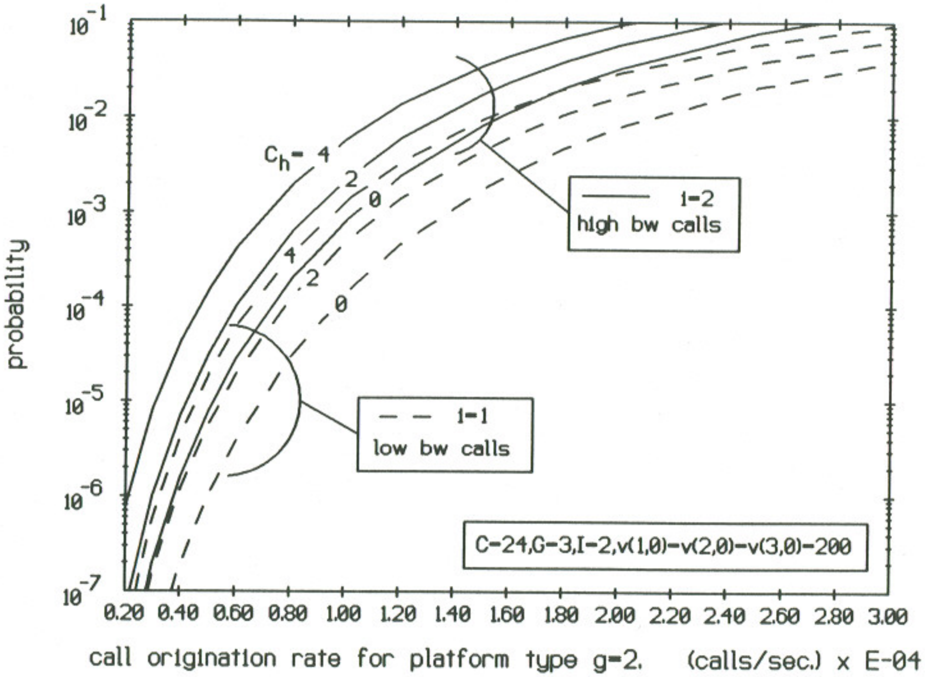

Figure _.1.c Blocking Probabilities for Different Call Types

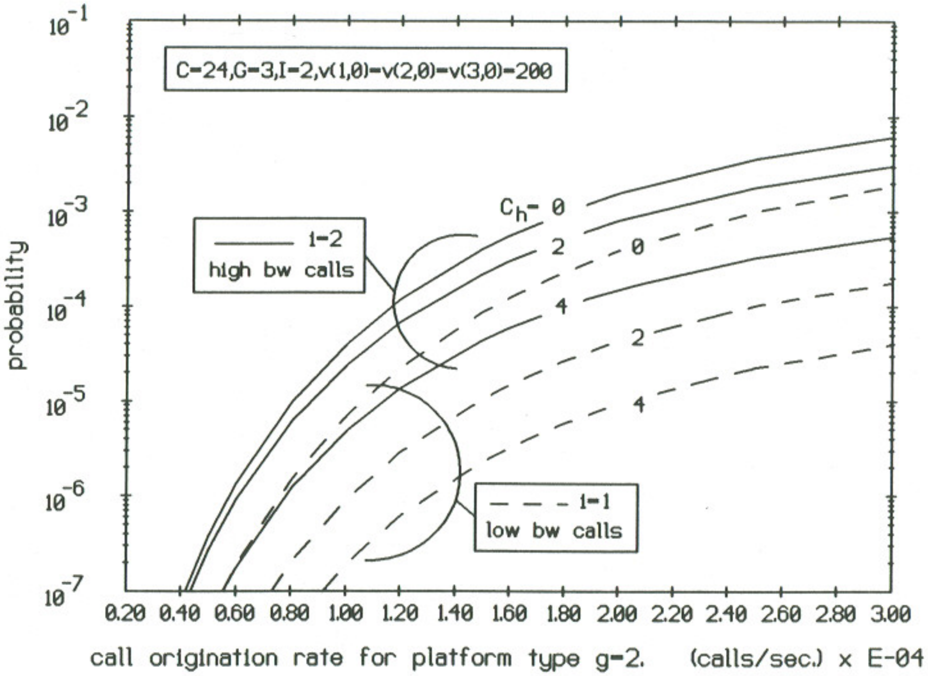

Figure _.1.d Forced Termination Probabilities
for Different Call Types

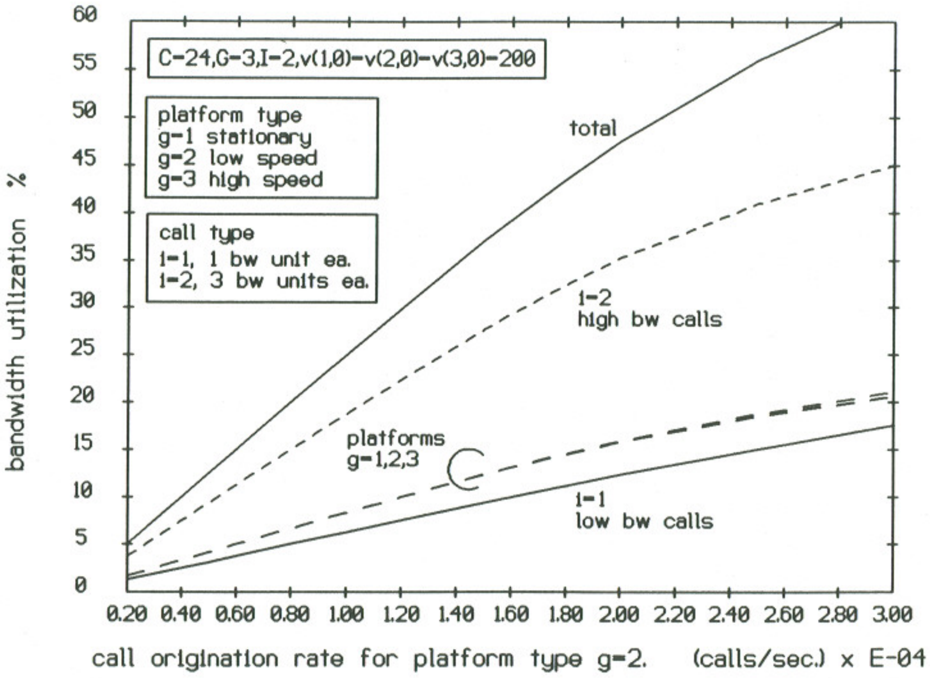

Figure _.2 Bandwidth Utilization for Various Traffic Components. ($C_h=2$)

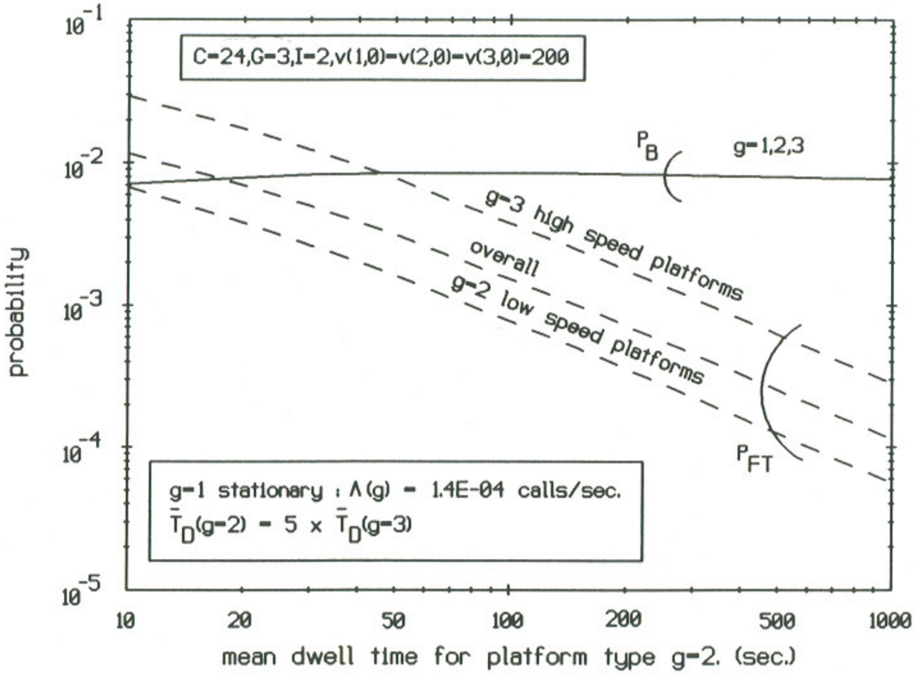

Figure _.3 Blocking and Forced Termination Probabilities Depend on Dwell Time Means (C_h=2)

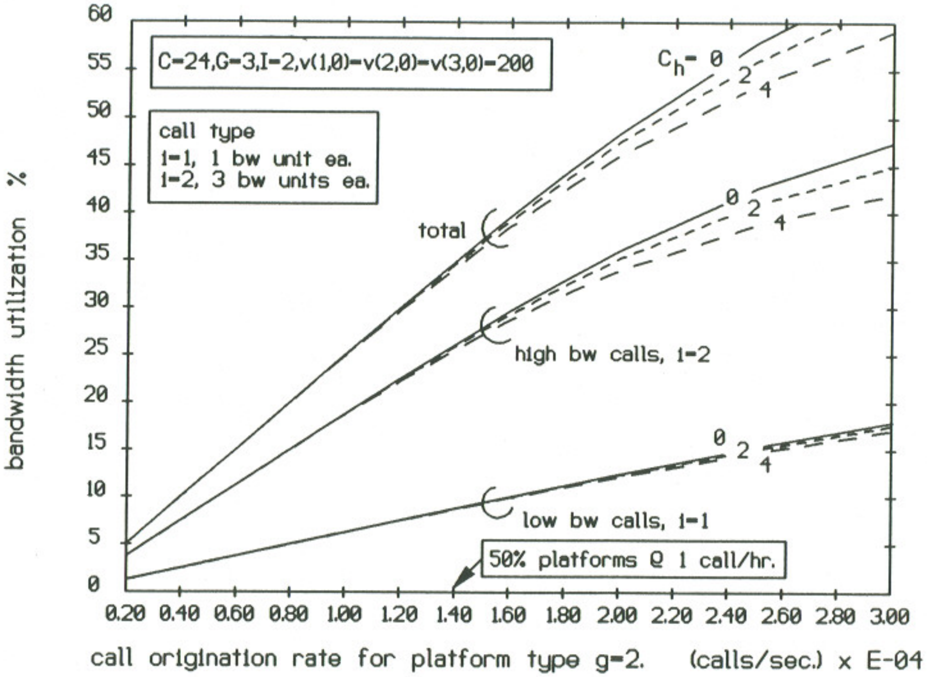

Figure _.5 Bandwidth Utilization for
Various Traffic Components (C_h=0,2,4)