Performance Analysis of Cellular Networks

By Myuran RAJARATNAM

University of Natal

2000

Submitted in fulfilment of the academic requirements for the degree of Doctor of Philosophy in the School of Electrical and Electronic Engineering, University of Natal, Durban, South-Africa.

Abstract

Performance analysis in cellular networks is the determination of customer orientated grade-of-service parameters, such as call blocking and dropping probabilities, using the methods of stochastic theory. This stochastic theory analysis is built on certain assumptions regarding the arrival and service processes of user-offered calls in a network. In the past, cellular networks were analysed using the classical assumptions, Poisson call arrivals and negative exponential channel holding times, borrowed from earlier fixed network analysis. However, cellular networks are markedly different from fixed networks, in that, they afford the user a unique opportunity: the ability to communicate while on the move. User mobility and various other cellular network characteristics, such as customer-billing, cell-layout and hand-off mechanisms, generally invalidate the use of Poisson arrivals and negative exponential holding times. Recent measurements on live networks substantiate this view. Consequently, over the past few years, there has been a noticeable shift towards using more generalised arrival and service distributions in the performance analysis of cellular networks. However, two shortcomings with the resulting models are that they suffer from state space explosion and / or they represent hand-off traffic as a state dependent *mean* arrival rate (thus ignoring the higher moments of the hand-off arrival process).

This thesis's contribution to cellular network analysis is a *moment-based* approach that avoids full state space description but ensures that the hand-off arrival process is modelled beyond the first moment. The thesis considers a performance analysis model that is based on Poisson new call arrivals, generalised hand-off call arrivals and a variety of channel holding times. The thesis shows that the performance analysis of a cellular network may be loosely decomposed into three parts, a generic *cell traffic characterising model*, a generic *cell traffic blocking model* and a *quality of service evaluation* model. The cell traffic characterising model is employed to determine the mean and variance of hand-off traffic offered by a cell to its neighbour. The cell traffic-blocking model is used to determine the blocking experienced by the various traffic streams offered to each cell. The quality of service evaluation part is essentially a *fixed-point iteration* of the cell traffic characterising and cell traffic blocking parts to determine customer orientated grade-of-service parameters such as blocking and dropping probabilities. The thesis also presents detailed mathematical models for user mobility modelling. Finally, the thesis provides extensive results to validate the proposed analysis and to illustrate the accuracy of the proposed analysis when compared to existing methods.

Preface

The research work discussed in this thesis was performed by Mr. Myuran Rajaratnam, under the supervision of Professor Fambirai Takawira, at the School of Electronic Engineering, University of Natal. The work was supported by Alcatel Altech Telecom and Telkom South Africa through the Centre of Excellence programme at the Centre for Radio Access Technologies at the University Of Natal

The work presented in this thesis has been published by the author in the IEEE Transactions in Vehicular Technology and has been presented by the author at six different IEEE / ITC conferences.

The whole thesis, unless specifically indicated to the contrary in the text, is the author's work, and has not been submitted in part, or in whole to any other University.

Acknowledgements

First and foremost thanks goes to Professor Fambirai Takawira for his help, encouragement and enthusiasm during the past three years. Secondly a great big thanks is owed to my wife Pranisha for her encouragement and support, especially during the trying times when the thesis did not seem worth the effort.

Thanks are also owed to Alcatel Altech Telecom and Telkom SA Ltd for funding this PhD study. Thanks is also owed to Professor Takawira, Alcatel Altech Telecom, Telkom SA Ltd, and Mr. Rajaratnam's current employer, BT Cellnet, for providing him with the funds to travel abroad and present papers at various international conferences.

Finally thanks is owed to my mathematics guru who lent me his expertise: Mr. Balakanapathy Rajaratnam, my brother, (University Of Witwatersrand).

TABLE OF CONTENTS

List of figures

Chapter 1

Chapter 2

No figures

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Chapter 8

No figures

Appendix

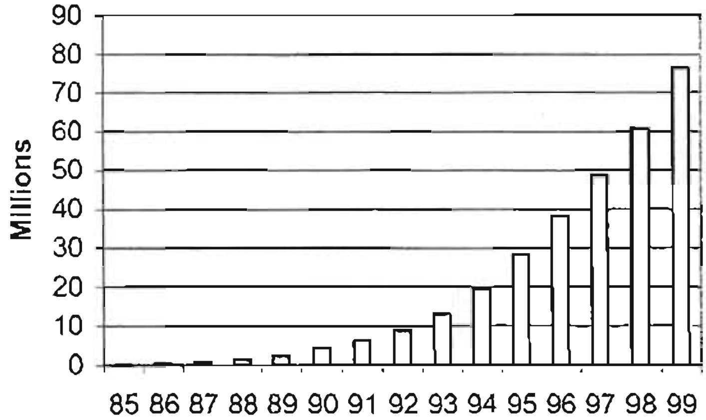
CHAPTER 1

1.1 Introduction

The mobile phone has become an integral part of every day human life. No longer a luxury of the rich, the mobile phone has become a readily available and reasonably cheap tool for every man, woman and child to conduct business and to pursue pleasure. The growth in mobile subscriber population has been nothing short of astounding. In the Scandinavian countries, e.g. Sweden and Finland, the penetration (the percentage of the population that has a mobile phone) is somewhere in the range of 60 to 70 percent. The country with the largest number of mobile users is the United States of America. Figure 1-1 illustrates the yearly growth in the US in the number of mobile subscribers.

The reason for the phenomenal success of the mobile communications industry and the accompanying huge popularity of the mobile handset can be attributed to man's desire for mobility in communications and the companion desire to be free from tethers, that is, from physical connections to communications networks. These desires have fuelled the rapid growth of mobile communications even when the well-established wired telephony service

Figure 1-1: Mobile Subscriber population in the US



is available at a fraction of the cost. Bullish industry pundits predict that by the year 2010 cellular subscribers will outnumber fixed network subscribers.

To accommodate such phenomenal growth, network operators, as a matter of survival, have had to *dimension* and *analyse the performance* of their networks accurately. Dimensioning is essentially a non-linear programming problem where given certain requirements for adequate *grade-of-service*, operators provide appropriate capacity where necessary. *Performance analysis* goes hand in hand with dimensioning and is the process whereby customer orientated grade-of-service parameters are evaluated for a network. The topic of this PhD thesis, in a nutshell, is the performance analysis of cellular mobile networks serving voice subscribers. We examine the ideal customer orientated grade-of-service parameters that may be used to measure the performance of a wireless network. We then consider how these parameters may be determined for a given wireless network configuration. For this purpose, we employ the methods of probability theory, stochastic process theory and tele-traffic analysis. Before we delve into this, we first examine the architecture and the associated technologies of popular wireless mobile network standards.

1.2 Cellular mobile networks

Modern wireless mobile networks are designed based on a *cellular architecture*, where the allocated radio spectrum is reused in areas that are sufficiently far away so as not to cause interference. The cellular architecture may be decomposed into a backbone network and a radio network. The backbone network connects the fixed base stations that serve each cell. The radio network connects the mobile subscribers to each base station. The area over which a particular base station and mobile subscribers in the vicinity can maintain a clear two way communication link is called a *cell*. The geographical shape of cells is determined by effects such as radio propagation losses, reflection, fading, antenna gain, antenna beam pattern, terrain etc. A set of channels is allocated to each base station. These channels may take the form of time-slots in Time Division Multiple Access (TDMA) systems, frequencies in Frequency Division Multiple Access (FDMA) systems and codes in Code Division Multiple Access (CDMA) systems. Subscribers communicate with each other as well as those users served by other fixed or wireless networks using the channel allocated to them by the base station. The use of a radio link between the mobile subscriber and the base station permits the user to be free of a physical connection to the network. Cell designers ensure that cellular coverage over a geographical area is contiguous so that communication is not interrupted when subscribers move out of the coverage area of a particular cell. Cellular networks have the facility to *hand-over* (or *hand-off*) the control of the call when the subscriber moves out of the coverage area of one cell and into the coverage area of another cell.

1.2.1 Competing Cellular Network Standards

As it is with most aspects of human life, political and geographical consideration has effected and affected the implementation of a wide variety of mobile communications standards. Different geographical regions in the world have proposed and implemented different wireless networks that operate in different manners. In North America, an analogue system called AMPS (Advanced mobile phone service) based on the 800-MHz band and later a digital version called D-AMPS have been quite popular. In Japan, the PDC (personal digital cellular system) was proposed and implemented. The pan-European effort was a digital standard called GSM (Global system for mobile communications) based on the 900-MHz band. The GSM standard is a hybrid FDMA-TDMA system that has become very popular around the world (except in North America & Japan). Another standard that has been proposed and accepted in North America is the IS-95: CDMA standard, where, code division multiple access rather than FDMA or TDMA is used for mobile communications. A further standard is the PCS1800 standard, which operates in the 1800 MHz region and is essentially an adaptation of the GSM900 standard. All the above standards are generally considered to be second generation wireless network standards and were proposed with one-to-one voice communications as their primary objective. More recently, a third generation wireless network standard for UMTS (Universal mobile telecommunications services) has been proposed based on a W-CDMA (Wideband-CDMA) access methodology. This third generation standard has been proposed with voice, data and multimedia communication services in mind. It is envisaged that data bit rates of 384 kbit/s for high mobility subscribers and 2 Mbit/s for limited mobility subscribers are possible in UMTS networks. It was hoped that the third generation standard would be a single global standard that allows the same handset to be used world-wide. However, due to spectrum allocation difficulties in certain countries and technology preferences by certain equipment vendors, there are, in fact, a few competing third generation standards. The joint-proposal by ETSI (European standards body) and ARIB (Japanese standards body) is currently the most popular. The architecture of most 2nd generation systems and to a certain extent 3rd generation systems is similar. Considering that GSM is the most popular 2nd generation standard, we examine its system architecture and discuss its functional elements.

1.3 GSM system Architecture

The GSM system architecture is a hierarchical one [1] as shown in figure 1-2. The various functional elements, in increasing order of hierarchy, are mobile stations (MS), Base Transceiver Stations (BTS), Base Station Controller (BSC) and Mobile Switching Center (MSC). Other relevant functional elements include the Operation and Maintenance Center (OMC), the Home Location Register (HLR), the Visitor Location Register (VLR), Authentication Center (AC) and Equipment Identity Register (EIR).

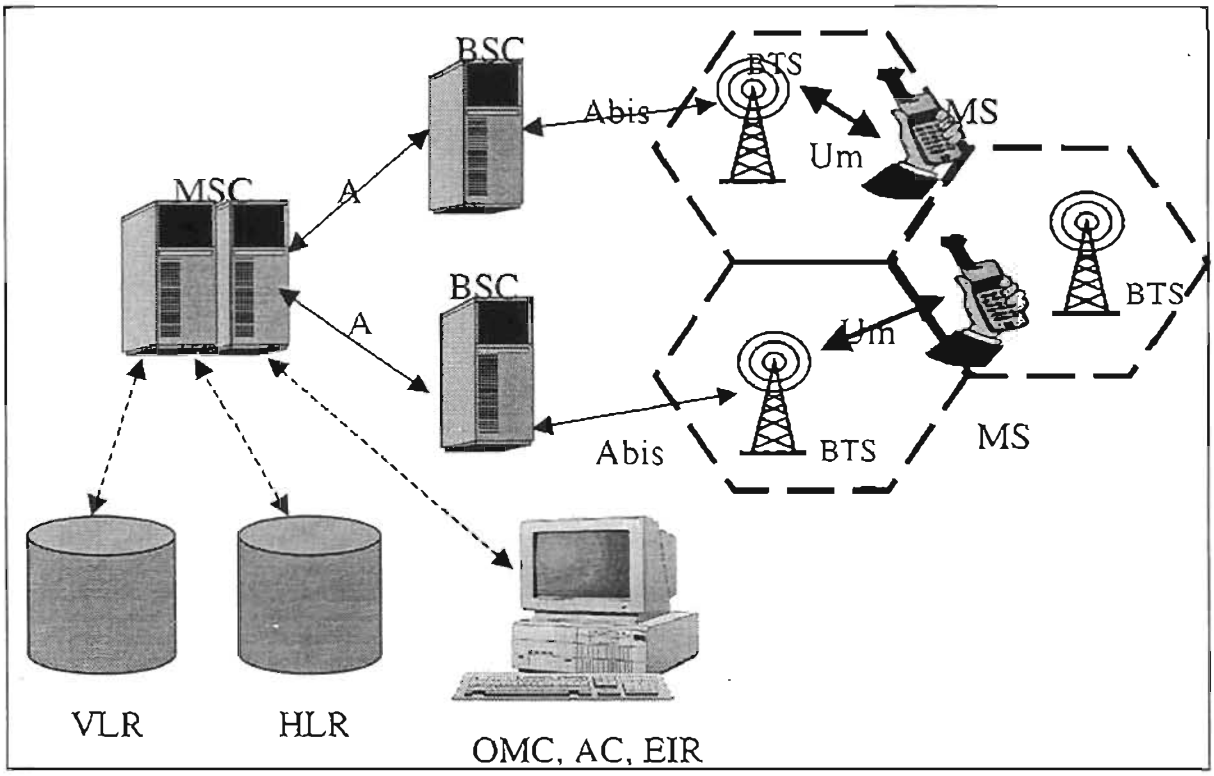


Figure 1-2: The GSM Architecture

Mobile Stations

The mobile stations range from large and bulky vehicle mounted units to small convenient hand-held units. The vehicle-mounted units have a maximum RF output of about 20W and the hand-held units up to 2W. With consistent improvements in battery technology, the mobile handsets are becoming smaller and smaller. This convenience in size has in fact given the digital mobile communications market a great boost in popularity.

The SIM or Subscriber Identity Module provides mobile equipment with a unique identity. The mobile handset does not operate without a SIM card (except for emergency calls). The SIM is a smart card with a processor and memory chip permanently installed in a plastic card the size of a credit card. The SIM stores various subscriber parameters, personal phone numbers as well as a PIN (personal identification number) for security reasons.

Base Transceiver Station

The BTS, which is normally located at the centre of a cell, forms the mobile subscriber's interface to the network. The antenna subsystem, normally mounted on a mast, may be omni-directional, sectored or various combinations depending on the traffic density and propagation requirements of the area. A base station has between one and sixteen transceivers, each of which represents a different RF channel. In GSM terminology, the radio interface between the mobile handset and the BTS is termed the Um-interface.

Base Station Controller

The BSC performs frequency administration, BTS monitoring and control functions, and exchange functions. The BSC also performs a traffic concentrator function by compressing speech in the direction of the BTS and decompressing it in the direction of the MSC. In GSM terminology, the interface between the BTS and the BSC is termed the Abis-interface.

Mobile Switching Centre

The MSC monitors and controls several BSCs, and forms the interconnection between the cellular network and other networks such as the PSTN. The MSC is a complete exchange with routing and switching functionality. In GSM terminology, the interface between the BSC and the MSC is termed the A-interface.

Operation and Maintenance Centre

The OMC plays an integral part in operation and maintenance of the GSM network. The OMC has connectivity to both the MSCs and the BSCs. It handles error messages coming from the network, and controls the traffic load of the BSC and the BTS. Even though the OMC does not have direct access to the BTS, an OMC operator may configure the BTS via the BSC.

Home Location Register

The HLR is a semi-permanent data store for subscriber information such as International Mobile Subscriber Number (IMSI), the users cell-phone number, authentication key, subscriber's category data (the permitted services for each subscriber). The HLR supports network call delivery by maintaining an up-to-date record of the current geographical location of subscribers in the network.

Visitor Location Register

The VLR is associated to a specific MSC and it contains a mirror image of the relevant HLR data for all mobiles currently located in the area served by the MSC. The VLR supports the MSC during the call establishment procedure and the authentication procedure by furnishing data specific to the subscriber.

Authentication center & Equipment Identity Register

The AC is related to the HLR. The AC contains certain ciphering algorithms that provide the HLR with different sets of parameters to complete the authentication of a mobile subscriber. The EIR is an optional security system that uses the unique International Mobile Equipment Identity (IMEI) of each handset to determine whether it should be allowed or barred access to the network depending on whether it is defective or perhaps stolen.

1.4 GSM Technology and Enhancements

Government bodies and technical standard committees dictate the frequency spectrum over which a mobile communications network is to operate. With many competing users such as the military, radio & television stations, and various other industries, the radio spectrum is allocated sparingly. The radio spectrum is considered to be the most precious commodity and perhaps the largest bottleneck in the mobile communications industry. In addition to this, the radio spectrum is very unreliable due to propagation factors such as path loss, fading, reflection, diffraction etc. To overcome this unreliability and to best use the limited radio spectrum, the GSM standard employs various techniques [1]:

RF Channel

The multiple access scheme in GSM is a hybrid FDMA/TDMA system. It is FDMA in the sense that different operators are generally allocated different portions of the radio spectrum. Furthermore, the spectrum is employed in a FDD (frequency division duplex) fashion whereby different bands are allocated for the uplink (handset to the BTS) and the downlink (BTS to the handset). The GSM uplink is at 935-960 MHz and the downlink is at 890-915 MHz. The frequency bands are divided into 125 channels with widths of 200kHz each. Time Division Multiple Access (TDMA) is used within each channel to subdivide it into 8 time slots each of which may be assigned to an individual user.

Speech Coding, Channel coding and Modulation.

In fixed networks, 64kb/s is the standard bit rate for speech. In GSM networks, low bit rate speech coding is used to increase the number of users served per MHz. The GSM standard employs a codec named RPE-LTP, which stands for regular pulse excitation and long-term prediction. Using RPE-LTP, speech is compressed to about 13 kb/s. GSM employs ½ rate convolutional channel coding and the digital modulation technique, Gaussian Minimum-Shift Keying. More information on these techniques can be found in [1].

Network Level Enhancements

The explosive growth in subscriber population has fuelled various innovative developments in second generation wireless networks. These developments include Dynamic Channel Allocation methods that optimally re-allocate spare capacity in the network [2,3,4,5,6], micro-cellular networks that accommodate larger subscriber populations [7,8], hierarchical cellular networks that accommodate subscribers with varied mobilities [7,9,10,11], and GPRS (General Packet Radio Service) for packet based data transfers (email and web-browsing) [12].

1.5 Customer-orientated Grade-Of-Service parameters

A mobile subscriber, termed the A-party, enters the B-party number (the person with whom he/she wants to communicate) on his/her handset and then presses the "send" button to initiate communication. Transparent to the A-party subscriber, the mobile handset and the base station perform electronic "hand-shaking" on the signalling channel. Provided a radio channel is available between the A-party and the base station, provided that a complete path from the base station to the B-party subscriber may be set up, and provided that the B-party is available (in person or in the form of some sort of a answering service) to answer the call, communication between the two parties may take place. A call may not be initiated if any one of the above provisions is not satisfied. A call may also fail for various other reasons such as signalling link failure, HLR / VLR authorisation failure etc. However, it is common in the literature to concentrate the tele-traffic study of wireless networks to the availability of resources in the radio interface layer. Due to the unreliability and the limited spectrum allocated in the radio interface layer, the possibility of call failure in the radio interface layer is generally an order of magnitude greater than the possibility of call failure in the remaining elements of a cellular network. Three customer orientated grade-of-service parameters are of paramount importance in the performance analysis of cellular networks. They are *new call blocking*, *hand-off call blocking* and *forced termination probability*. New call blocking is the congestion experienced by the A-party when he/she presses the "send" button to initiate communication and finds that no free channels are available for his/her use. If, however, a channel is available for his/her use, and if the remaining portion of the call set-up process goes ahead without any problems, communication between the two parties will commence. The channel that was allocated to the A-party may be released for one of two reasons: (i) the call is completed after the normal course of conversation; (ii) the A-party moves to another cell before the call is completed and in which case a hand-off is required. Hand-off blocking is the congestion experienced by the A-party when he/she moves into the coverage area of a target cell and no free channel is available for use by the A-party in the target cell. Forced termination probability aggregates the effect of various hand-off scenarios and is the probability that the A-party, after being provided with the first channel in the cell where he/she initiated the call, will *ever* be dropped by the network due to a hand-off failure when crossing various cell boundaries. Therefore, one minus the forced termination probability is the probability with which the A-party, after being admitted into the network, will be able to communicate with the B-party without interruption due to the lack of resources in the wireless network serving the A-party. Since the arrival of calls and the service of such calls are of a stochastic nature, a study on the grade-of-service parameters involves the methods of probability theory. Certain assumptions are made regarding the arrival and service processes, and the methods of tele-traffic theory are then employed to determine the grade-of-service parameters.

1.5.1 Classical Analysis

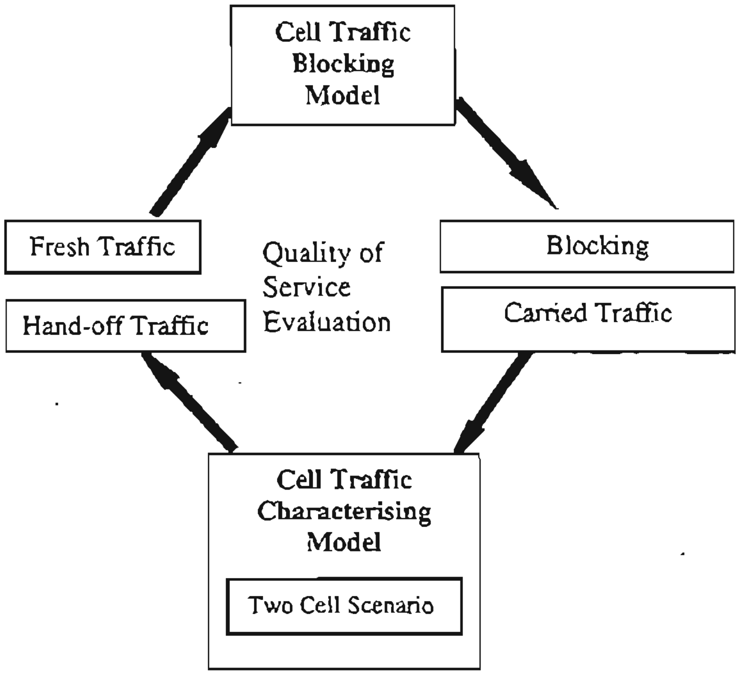
In the seventies, fixed or landline telephone networks were analysed using M/M/C loss systems [13]. Efficient numerical algorithms for determining network Quality of Service (QOS) parameters were derived [14] under the simplistic assumptions of Poisson call arrivals and negative exponential service times. These algorithms were easy to implement and were computationally efficient in relation to the capabilities of the computers of that era. Cellular network analysis, in the eighties and early nineties, has been performed by simply borrowing the above-mentioned assumptions from the fixed network era [15,16]. However, there is conflicting evidence in the literature on the validity of Poisson arrivals and negative exponential channel holding times [17-30]. Recent measurements on live networks [31-33] have generally invalidated these classical assumptions. Consequently, over the past few years, there has been a noticeable shift towards using more generalised arrival and service time distributions in the performance analysis of cellular networks. Two shortcomings with most recent approaches is that they suffer from state space explosion (possibly leading to intractable solutions) [21,26-30] and / or they represent hand-off traffic as a state dependent *mean* arrival rate (thus ignoring the higher moments of the hand-off arrival process) [15,16,21,26-30].

1.5.2 Proposed Analysis

Our contribution in the performance analysis of cellular networks is a moment-based solution that avoids full state space description but allows the hand-off arrival process to be modelled beyond the first moment. Unlike existing methods, we make no presumptions regarding the hand-off traffic process. The hand-off traffic process is implicitly and necessarily defined by the new call arrival process, the service process, user mobility and the network configuration. We show that a performance analysis model that is based on Poisson new call arrivals, generalised hand-off call arrivals and negative exponential service times is reasonably valid for general cellular network layouts. However, certain networks, such as highway cellular networks, require further extensions. For networks such as the highway network we present a moment-based approach that employs Poisson new call arrivals, generalised hand-off call arrivals and more appropriate service time distributions.

Our proposed approach is quite modular. We show that the performance analysis of a cellular network may be loosely decomposed into three parts, a generic *cell traffic characterising model*, a generic *cell traffic blocking model* and a *quality of service evaluation* model. See figure 1-3.

Figure 1-3: Performance Analysis Algorithm



In the cell traffic characterising model we determine the mean and variance of fresh traffic offered to each cell, and the mean and variance of hand-off traffic offered by a cell to its neighbour. In the cell traffic blocking model we determine the blocking experienced by the various traffic streams offered to each cell and the carried traffic in each cell. The quality of service evaluation part is used to determine customer orientated grade-of-service parameters and is essentially a fixed-point iteration of the cell traffic characterising and cell traffic blocking parts. Fixed-point iteration, also known as the relaxation method or repeated substitution, is a simple technique for solving the non-linear equations that describe the system. As shown in figure 1-3, the solution process is an iterative cycle whereby, the outputs of the cell traffic characterising model (offered traffic values) are used as the inputs of the cell traffic blocking model. In return, the outputs of cell traffic blocking model (blocking and carried traffic values) are fed back in as the inputs of the cell traffic characterising model. This is repeated until convergence. As can be seen in figure 1-3, at the heart of our cell traffic characterising model, there lies the stochastic theory analysis of a simple *two-cell scenario*. The two-cell scenario is fundamental to our performance analysis algorithm and is used as "building blocks" to analyse large-scale networks. Typically, we determine the mean and variance of hand-off traffic offered by a cell to its neighbour in our two-cell scenario, and then extend the results to apply for multi-cellular networks.

1.6 Thesis Layout

In chapter 2 of this document, we define the various tele-traffic parameters that need to be considered in the performance analysis of wireless networks. These parameters are new call arrivals, hand-off call arrivals, call holding times, cell dwell times and channel holding times. We discuss the various models in the literature for these parameters, based on analytical modelling, simulation studies and live network measurements. We discuss the usefulness as well as shortcomings of these approaches. In chapter 3, we consider the various performance analysis algorithms in the literature that employ the different tele-traffic models that we presented in chapter 2. We consider earlier single moment methods and the more recent state dependent methods for the performance analysis of cellular networks. We show how customer orientated grade of service parameters may be determined using these performance analysis algorithms. In chapter 4 we examine our two-cell model which we use as building blocks for analysing large-scale multi-cell networks. In chapter 5 we consider our cell traffic blocking models for determining the blocking experienced by the various offered traffic in a cell. In chapter 6, we consider the quality of service algorithm, where we show how the cell traffic characterising and cell traffic blocking models may be applied in a iterative manner to determine customer orientated grade-of-service parameters such as new call blocking, hand-off call blocking and forced termination probability. In chapters 4 to 6 we consider the accuracy of the various proposed methods by comparison with existing models and simulation. Characterising user mobility is important, since, only then it is possible to determine the time spent by the user employing the resources of the network. In chapter 7, we present some work on mobility models. In chapter 8 we present our conclusions.

1.7 Published work

Original work that we present in this thesis has been published at various IEEE journals and IEEE conferences. These publications are:

1.7.1 IEEE Journal Papers:

[1] Rajaratnam, M., and Takawira, F., "Non-classical traffic modelling and performance analysis of cellular mobile networks with and without channel reservation," *IEEE Transactions on Vehicular Technology*, Vol. 49, Number 03, pp 817-834, May 2000.

[2] Rajaratnam, M., and Takawira, F., " Hand-off traffic characterisation in cellular networks under non-classical arrivals and service time distributions," *submitted for IEEE Transactions on Vehicular Technology.*

1.7.2 IEEE / ITC Conferences

[1] Rajaratnam, M., and Takawira, F., "Hand-off traffic modelling in cellular networks", *Proceedings of IEEE Global Communications Conference (GLOBECOM'97)*, Phoenix, Arizona, pp. 131-137, Nov 1997.

[2] Rajaratnam, M., and Takawira, F., "The two moment performance analysis of cellular mobile networks with and without channel reservation," *Proc IEEE Conference on Universal Personal Communications (ICUPC'98)*, pp. 1157-1161, Florence, Oct 98.

[3] Rajaratnam, M., and Takawira, F., "Performance Analysis of Highway Cellular Networks using generalised arrival and generalised service time distributions", *Proceeding of International Teletraffic Congess (ITC-16)*, Vol. 3a: Teletraffic Science and Engineering, pp 11-22, June 1999, Edinburgh, Scotland.

[4] Rajaratnam, M., and Takawira, F., "Hand-off traffic characterisation in cellular networks under Engset new call arrivals and generalised service times" *Proceedings of*

IEEE Wireless & Networking Conference (WCNC'99), pp. 423-427, New Orleans, September 1999.

[5] Rajaratnam, M., and Takawira, F., "Asymptotic approximation for hand-off traffic characterisation in cellular networks under non-classical arrival and service time distributions", *Proceedings of the ITC mini-seminar*, pp 95-106, Lillehammer, Norway, Mar 2000.

[6] Rajaratnam, M., and Takawira, F., "Hand-Off Traffic Characterisation In Cellular Networks Under Non-Classical Arrivals And Gamma Service Time Distributions", *Proceedings of the IEEE Personal, Indoor, Mobile Communications conference PIMRC 2000 conference*, pp. 1635-153, Volume 2, September 2000, London.

1.7.3 South-African Telecommunications Conferences:

[1] Rajaratnam, M. & Takawira, F., "Practical Approach to mobility modelling in cellular mobile networks", SATNAC'98, pp456-465, Capetown, September 1998.

[2] Rajaratnam, M. & Takawira, F., "Two moment analysis of channel reservation in cellular networks", COMSIG'99, pp345-349, Capetown, September 1999.

1.8 Original work in this thesis

The original work presented in this thesis is as follows:

In Chapter 4 we analyse a simple two-cell scenario. We derive expressions for the hand-off traffic offered by a cell to its neighbour under various channel holding time distributions. They include:

(i) Negative exponential channel holding times;

(ii) Deterministic channel holding times;

(iii) Det-neg channel holding times (a combination of negative exponential and deterministic distributions (defined later));

(iv) Gamma channel holding times;

Chapter 5 discusses the cell traffic blocking models. We consider the following two cell traffic-blocking models from fixed network analysis:

(i) Delbrouck's two moment model

(ii) Sanders, Haemers and Wilcke's two moment model

We then extend the above to apply for cellular networks and show how to determine congestion probabilities and carried traffic values for the case where

(i) There is no channel reservation

(ii) Channel reservation of level r is implemented.

Channel reservation is used in mobile networks to give higher priority to hand-off calls at the expense of new calls by reserving a certain number of channels, r, for the exclusive use of hand-off calls.

In chapter 6, we show how the cell traffic characterising model and the cell traffic blocking models may be applied in a fixed-point iteration to determine customer orientated grade-of-service parameters. In chapter 7 we consider our proposed mobility models. Assuming certain distributions for the velocity and the distance travelled by a vehicle on a highway, and assuming a certain geometry for the highway cell, we analytically derive the distributions of the time spent by mobile users (travelling in vehicles) within the coverage area of a highway cell. We also show that, in the literature, to simplify analysis, very simple assumptions are made regarding the mobility of users, thus resulting in mobility models that are not necessarily an accurate reflection of reality. To circumvent this, we show how one may performance analyse an *established* cellular network without actually having to derive the relevant dwell time distributions. Our proposed approach is useful in established networks where it is often easier to measure mobility-related traffic parameters from the network than explicitly derive the cell dwell time distributions.

CHAPTER 2

2.1 Introduction

The performance analysis of cellular mobile networks is carried out using the techniques of queuing theory. In queuing theory analysis, it is necessary to characterise accurately the arrival and service time distributions such that the resulting stochastic model reflects reality. A poor knowledge of these distributions contributes to an inefficient design of the network because the engineer has to be conservative to cope with a possible error margin. In this chapter, we consider the various arrival and service processes relevant to cellular network performance analysis. In the subsequent chapter, we consider the performance analysis algorithms where these arrival and service processes are employed.

There are in essence four different stochastic variables that need to be modelled accurately for the performance analysis of cellular networks. They are new call arrivals, hand-off call arrivals, cell dwell times and unencumbered call holding times. The cell dwell times and the unencumbered call holding times together determine the nature of another important variable in cellular networks, the channel holding time. Various models have been proposed in the literature for these parameters based on mathematical analysis, computer simulation or live network measurements. In subsequent sections, a literature survey is performed on these models, highlighting their advantages, their disadvantages and the validity of the assumptions made in each of them.

2.2 New call arrivals

2.2.1 Analytic model for new call arrivals

A new or fresh call is the initial request for service that occurs when a cellular subscriber (the A-party) enters a B-party number on his handset and presses the "send" button. In almost all of the literature, the new call arrival process is assumed to be Poisson [9,15,16,21,26-30]. The Poisson assumption implies that the distribution of the inter-arrival times between successive new call arrivals is negative-exponential in nature. Given that the mean new call arrival rate is λ, the probability $p_k(t)$ that k call arrivals occur within the time interval $[0,t)$ is:

$$p_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \tag{2.1}$$

The Poisson assumption has been borrowed from the days of fixed (wire-line) telephony [13,14]. In the loss system scenario of fixed telephone service, there were many advantages to the Poisson call arrival assumption. Firstly there exists a *memory-less* attribute associated with the negative exponential inter-arrival times that reasonably mimics the real-life effect of subscribers offering calls in a random and non-premeditated manner. Secondly, the analysis of a finite server loss system under Poisson call arrivals is very simple: the famous *Erlang-B formula* explicitly solves for the blocking or congestion experienced in a finite server loss system. The blocking E experienced by a Poisson arrival process with mean arrival rate λ and mean service rate μ in a finite server system of size C is given by:

$$E = \frac{\dfrac{\left(\frac{\lambda}{\mu}\right)^C}{C!}}{\displaystyle\sum_{j=0}^{C} \frac{\left(\frac{\lambda}{\mu}\right)^j}{j!}} \tag{2.2}$$

However, one disadvantage of the Poisson arrival process is the infinite population hypothesis. There exists an intrinsic assumption that the number of subscribers that may offer a call is infinite or unbounded. For practical situations, the Poisson arrival process is reasonably valid in cases where the number of subscribers is much larger than the number of servers available. This condition is satisfied in public telephony systems, where the number of users connected to a line concentrator is much greater than the number of trunks available between the line concentrator and the exchange. Similarly in mobile networks, it is reasonable to expect that, for the purposes of efficiency, the network is operated such that the number of mobile subscribers within the coverage area of a cell is much greater than the number of channels allocated to that cell.

2.3 Unencumbered call holding time

2.3.1 Analytic model for unencumbered call holding time

The unencumbered call holding time is the time interval for which a mobile subscriber, if unhindered, would employ the services of the network. We define a random variable, T_M, which describes the unencumbered call holding time of users in the wireless network. The trend in existing literature has been to assume that the unencumbered call holding time distribution $f_M(t)$ is negative exponentially distributed [2,7,9,15,16,21,27]:

$$f_M(t) = \mu_M e^{-\mu_M t} \tag{2.3}$$

where the mean unencumbered call holding time $E[T_M] = 1/\mu_M$. This is another assumption that has been borrowed from the days of fixed network analysis [13,14]. There are some useful advantages in assuming negative exponential service times in the queuing theory analysis of telecommunications networks. The amazing characteristic of the negative exponential distribution is its memory-less property: the past history of a random variable that is negative exponentially distributed plays no role in predicting its future. The cdf $F_M(t)$ of the negative exponential distribution can be obtained by integrating equation (2.3):

$$F_M(t) = 1 - e^{-\mu_M t} \tag{2.4}$$

Now consider some time s in the future until which no call departure occurs. The conditional probability that a call departure occurs in the interval from s to $s+t$ sec given that no departure occurred in the interval from 0 to s sec is:

$$P[X \le t+s \mid X > s] = \frac{P[s < X \le t+s]}{P[X > s]} = \frac{P[X \le t+s] - P[X \le s]}{P[X > s]} \tag{2.5}$$

Substituting equation (2.4):

$$P[X \le t+s \mid X > s] = \frac{1 - e^{-\mu_M (t+s)} - \left(1 - e^{-\mu_M s}\right)}{1 - \left(1 - e^{-\mu_M s}\right)} = 1 - e^{-\mu_M t} \tag{2.6}$$

This result shows that the distribution of the remaining time until the call departure, given that s sec has elapsed is identically equal to the unconditional distribution of the service time. The exponential distribution is the only continuous distribution with this property. This property makes the solutions of queuing theory models more tractable and manageable.

2.3.2 Field measurements for unencumbered call holding time

Chlebus [34] performed measurements on live networks to determine if the unencumbered call holding time in cellular networks is negative exponential. The author collected two sets of empirical data, corresponding to the busy hour of two different cellular communications systems. Four hypothesised probability distributions, namely, the negative exponential, gamma, Weibull and a mixture of two log-normals were examined and best fitted to the measured data. For both cellular communications systems, the best fit was obtained using the mixture of two log-normals, but, this was obtained at an expense of fitting a five parameter distribution instead of the two-parameter gamma / Weibull or the one parameter negative exponential. Furthermore, the author found that the two-parameter Weibull and gamma distributions do not provide a significant improvement over the single parameter negative exponential distribution. His conclusion was that the negative exponential distribution performed well enough to be considered for modelling purposes.

Poisson new call arrivals and negative exponential unencumbered call holding times have been easily adopted in wireless network analysis because they are generally construed to be wireless network versions of wire-line network parameters. However there are no wire-line equivalents for the remaining two wire-less network parameters, hand-off call arrivals and cell dwell times.

2.4 Hand-off call arrivals

In wireless networks the geographical area is subdivided into cells, each of which is served by a base station. This cellular concept allows for the frequency spectrum to be better utilised under the frequency re-use notion. However, it has an important side effect in that it leads to the occurrence of call hand-offs. It is common for users, whilst in the midst of a conversation, to move out of the coverage area of the base station that is currently serving them and into the coverage area of another base station. When this occurs, it is necessary that a channel be allocated for the user from the available pool of channels at the disposal of the target base station. Thereafter, the control of the call is transferred from the source base station to the target base station via appropriate signalling between the user and the two base stations. A successful hand-off is one where the subscriber successfully seizes the allocated channel in the target base station and relinquishes the channel he/she occupied in the source base station without any discernible break in conversation. Not all hand-offs are successful because, for example, there may not be any free channels available in the target base station and in which case a hand-off failure occurs.

2.4.1 Hand-offs in GSM

In GSM, the hand-off decisions are made by the BSC, based on the measurement results reported by the MS / BTS and on the various hand-off thresholds set for each cell [26]. The trigger for a hand-off occurs when a measured value breaches a hand-off threshold value. The following list contains some of the thresholds that may be set for the purposes of hand-off:

1. Interference (uplink or downlink)
2. Uplink quality (quality is a measure of BER)
3. Downlink quality (quality is a measure of BER)
4. Uplink power level
5. Downlink power level

6. MS-BS distance (maximum or minimum)
7. Fast/Slow-moving MS
8. Better cell (Power budget: path loss to neighbour is better than path loss to serving cell)

If the HO threshold comparison indicates that a hand-off might be required, the BSC examines the potential target cells for the hand-off. The BSC performs intra-BSC hand-offs autonomously. If there is an inter-BSC hand-off to be performed, the BSC sends a list of the preferred cells to the MSC and the MSC performs the hand-off according to the list. The crucial principle for selecting the target cells for the hand-off is that the neighbouring cell must be better than the current serving cell in order for the hand-off to be useful.

2.4.2 Handoff strategies

In cellular networks, two forms of blocking are possible, new call blocking and hand-off call blocking. Although the user perceives both situations as quality degradation, the latter is much more annoying than the former and is set to a much lower level. Various Hand-off strategies are employed to reduce the blocking experienced during hand-off.

2.4.2.1 Channel reservation

To minimise the *forced termination* of hand-off calls, network operators normally employ channel reservation for hand-off calls where a few channels are reserved for the exclusive use of hand-off calls [15]. Specific channels are not reserved but only the number. In essence, channel reservation operates by trading off new call blocking for improved hand-off call blocking. Channel reservation is the most common form of combating hand-off blocking and is the one that we consider in detail in this thesis.

2.4.2.2 Other Hand-off strategies

The handoff queuing scheme [35] allows for hand-off call requests to be queued and then be served in a FIFO manner as channels are released by departing calls. An extension to this was the measurement based priority schemes (MBPS) [36] where, instead of the simple FIFO queuing discipline, priority is given to certain users based on measurement levels.

More sophisticated measurement based schemes have been studied recently and show improved performance [37,38,39,40]. In such methods, there exists a need to transmit measured data between MS's and BTS's and between BTS's through the fixed network to achieve QOS improvement. These measured magnitudes include the MS/BTS power levels (both instantaneous and time evolution), power levels as received in others BTS, time already spent in the handoff dwell area, number of active calls and/or existing MS's in neighbour cells, etc. Although some of these measurements are simple or already available in some cellular systems, there exists an added complexity in such measurement based schemes.

2.4.3 Analytic model for hand-off call arrivals

The earlier papers on wireless network performance analysis simply assumed that the hand-off call arrival process is Poisson [15,16]. In the absence of proof to indicate the contrary, the Poisson assumption for hand-off call arrivals in these early papers was a reasonably useful one since it allowed for simple and tractable models that could be easily implemented on the computers of that era. There are, in fact, very few papers in the literature that examine the true nature of the hand-off call arrival process. In an appropriately titled paper "Is Hand-off traffic really Possonian?" Chlebus and Ludwin [19] show that under Poisson new call arrivals and a *non-blocking* environment (an environment where there exists an infinite number of channels in each cell) hand-off traffic is indeed Poisson. This means that the overall call arrival in each cell is Poisson since the sum of a Poisson new call arrival process and a Poisson hand-off call arrival process is Poisson as well [14]. Chlebus and Ludwin further claim without proof that, under Poisson new call arrivals and a blocking environment (where there are a finite number of channels), the hand-off call arrival process is *smooth*. A smooth traffic process is one where the variance of the offered traffic is less than its mean. In a Poisson arrival process, the variance of the offered traffic equals the mean. Consequently, the overall call arrival process in a blocking environment should be smooth since the sum of a Poisson new call arrival process and a smooth hand-off call arrival process is smooth as well. Such a summation is possible under a loose assumption of independence between the new call arrival process and the hand-off call arrival process in a cell.

2.4.4 Field measurements for overall call arrivals

Field measurements for overall call arrivals are available for a BC Tel Mobility Cellular network in Canada [31] and for a TACs network in Barcelona [32]. Data was measured on the BC Tel Mobility Cellular network for three separate cell sites. One was a highway cell, another was in downtown Vancouver and the last one in a small density area. Data was then cleaned to remove anomalies. For the TACS network in Barcelona, traffic measurements were obtained through a scanning receiver controlled by a PC. Since the TACS standard employs Frequency Modulation (FM), the authors used the detection of the carrier in the downlink to determine channel occupancy. The authors then performed a form of pre-processing and filtered the data to eliminate interference, short transmission cuts and other undesirable affects [32].

The paper based on network measurements in the BC Tel Mobility Cellular network in Canada [31] suggests that the inter-arrival times of cellular calls is negative exponential and therefore the call arrival process is Poisson. The more recent paper based on network measurements in the TACS network in Barcelona [32] disputes the earlier paper's findings and suggests that the call arrival process is non-Poisson and indeed smooth. However, it is not difficult to resolve the conflict between the two papers. Firstly it must be stressed that both papers do not measure the inter-arrival times of hand-off calls and new calls separately. They do not make a distinction between the two and simply measure the inter-arrival times of all calls offered to respective cells. In the measurements in Canada, the authors, due to a shortcoming in their measurement device that led to the inability to monitor blocked calls, simply disregarded high load periods when all channels were busy. This meant that they only monitored inter-arrival times for those periods when at least some channels were not occupied. This is equivalent to a non-blocking environment. Chlebus et al's [19] proven result for non-blocking environments applies and the call arrival process is Poisson. In the measurements in Barcelona, the authors measure inter-arrivals times in a blocking environment (although it was a low blocking environment) and found the traffic to be of the smooth variety. This confirms Chlebus et al's proposition that traffic is smooth in blocking environments.

For the TACS network in Barcelona, the results of the field studies were fitted to popular probability density functions. The parameters of the candidate distributions were determined using maximum likelihood estimation. Thereafter, the suitability of the candidate distributions was tested using Kolmogorov-Smirnov goodness-of-fit test. The authors categorised the offered traffic into three different load types, heavy, medium and light, and found the best fit distribution for the inter-arrival times for each load type. For the heavy load scenario, the best fit was found to be an Erlang-3,8 distribution:

$$f(t) = p\beta^{-n}\frac{t^{n-1}}{(n-1)!}e^{-\frac{t}{\beta}} + (1-p)\beta^{-k}\frac{t^{k-1}}{(k-1)!}e^{-\frac{t}{\beta}} \quad (2.7)$$

where $\beta = 17.84$, $n = 3$, $k = 8$, and $p = 0.875$. For the medium load sample, the best fit was found to be a combination of two erlang-k distributions:

$$f(t) = p\beta_1^{-n}\frac{t^{n-1}}{(n-1)!}e^{-\frac{t}{\beta_1}} + (1-p)\beta_2^{-k}\frac{t^{k-1}}{(k-1)!}e^{-\frac{t}{\beta_2}} \quad (2.8)$$

where $\beta_1 = 12.8$, $\beta_2 = 26.6$, $n = 4$, $k = 2$, and $p = 0.4$. For the light load sample, a mixture of two log-normal distributions was found to be the best fit:

$$f(t) = p\frac{1}{t\sqrt{2\pi\sigma_1^2}}e^{-\frac{[\log(t)-\mu_1]^2}{2\sigma_1^2}} + (1-p)\frac{1}{t\sqrt{2\pi\sigma_2^2}}e^{-\frac{[\log(t)-\mu_2]^2}{2\sigma_2^2}} \quad (2.9)$$

where $p = 0.5$, $\mu_1 = 4.6$, $\sigma_1 = 0.44$, $\mu_2 = 3.94$, $\sigma_2 = 0.63$. It is not advisable to assume generality of the results presented in this paper, because each load sample was collected among 10 different scanned channels of the same cell. It is reasonable to expect that network layout and subscriber mobility varies from place to place and the best fit candidate distributions presented above may not be valid for other possible scenarios. Interestingly, the field measurements in [32] indicate that smooth traffic was offered to the cells under all three-load types. In our work that we present in this thesis, we analytically show that smooth hand-off traffic is generated under quite general network scenarios.

2.5 Cell dwell times and Channel holding times

2.5.1 Analytic model for cell dwell times and channel holding times

The cell dwell time is the time interval for which the mobile user is within communications range of a base station. Due to the statistical nature of the cell dwell times, it is common to use probability density functions to represent them. We define two random variables, T_n and T_h, which describe the cell dwell times of new and previously handed-off calls in a cell. We assume that these random variables have probability density functions (pdfs), $f_n(t)$ and $f_h(t)$, respectively. As can be imagined, the cell dwell times are very much dependent on subscriber and network attributes such as subscriber position, velocity, direction of motion, cell coverage etc. These pdfs vary from mobility model to mobility model.

The channel holding time describes the time spent by a mobile subscriber making use of the resources (channels) within a cell. The channel holding time is determined by both the unencumbered call holding time and the cell dwell times. The channel assigned to a new call will be held until either the call is completed in the cell or the user moves out of the coverage area of the cell before the call is completed. Similarly the channel assigned to a hand-off call offered to a cell will be held until either the call is completed or the users moves out of the coverage area of the cell. Therefore the channel holding times of a new call T_{Hn} and a previously handed-off call T_{Hh} are:

$$T_{Hn} = \min(T_M, T_n)$$
$$T_{Hh} = \min(T_M, T_h) \qquad (2.10)$$

Assuming independence of T_M and T_n, and of T_M and T_h, Hong and Rappaport [15] showed that the cumulative distribution function (cdf) of the channel holding time of new calls $F_{Hn}(t)$ and hand-off calls $F_{Hh}(t)$ are:

$$F_{Hn}(t) = F_M(t) + F_n(t)(1 - F_M(t))$$
$$F_{Hh}(t) = F_M(t) + F_h(t)(1 - F_M(t)) \qquad (2.11)$$

where $F_M(t)$ is the cdf of the unencumbered call holding time. It is cumbersome to model loss systems under two different channel holding times. The approach taken by Hong and Rappaport and various other authors [15-16], has been to define a single channel holding time T_H with cdf $F_H(t)$ that is a proportional sum of the channel holding times of new and hand-off calls:

$$F_H(t) = F_M(t) + \frac{1}{1+\Upsilon_C}[1 - F_M(t)][F_n(t) + \Upsilon_C F_h(t)] \tag{2.12}$$

where Y_C is the ratio of mean carried hand-off calls to the mean carried new calls and we show how to calculate it in the subsequent chapter. We now look at the various proposals in the literature for cell dwell times and channel holding times.

2.5.1.1 Macmillan's mobility model

Macmillan [16] probably employed the simplest model for cell dwell times. No distinction is made between the cell dwell times of new and hand-off calls. Both call types are assumed to have the same cell dwell time distribution, a memory-less negative exponential distribution with mean $1/\mu_d$. In addition, it is assumed that the unencumbered call holding time is negative exponentially distributed with mean $1/\mu_M$. Therefore, the channel holding time T_H is also negative exponentially distributed with mean $E[T_H] = 1/(\mu_M + \mu_d)$. The cdf $F_H(t)$ of the channel holding time is given by:

$$F_H(t) = 1 - e^{-(\mu_M + \mu_d)t} \tag{2.13}$$

The author does not offer any reasons for the use of the negative exponential distribution.

2.5.1.2 Del Re, Fantacci and Giambene's mobility model

Del Re et al. [2] propose a user mobility model that implicitly defines the cell dwell times. Similar to Macmillan [16], these authors assume that both new call and hand-off calls have the same cell dwell time distribution. They assume that:

- All users travel a distance L that is uniformly distributed between 0 and $2R$ where R is a cell side
- Users transit cells at a constant velocity V, uniformly distributed between 0 and V_{max}.

They define the parameter α to characterise user mobility:

$$\alpha = \frac{2R}{V_{\max}\ E[T_M]} \tag{2.14}$$

They show that the cell dwell time pdf $f_d(t)$ is:

$$f_d(t) = \frac{u(t) - u(t - \alpha E[T_M])}{2\alpha E[T_M]} + \frac{\alpha E[T_M] u(t - \alpha E[T_M])}{2t^2} \quad (2.15)$$

where $u(t)$ is:

$$u(t) = \begin{cases} 1, & t \geq 0 \\ 0, & otherwise \end{cases} \quad (2.16)$$

They also show that the mean channel holding time $E[T_H]$ is given by:

$$E[T_H] = E[T_M](1 - P_H) \quad (2.17)$$

where the probability of hand-off P_H is given by:

$$P_H = \frac{1 - e^{-\alpha}(1 - \alpha)}{2\alpha} - \frac{\alpha}{2}\int_{\alpha}^{\infty} \frac{e^{-x}}{x} dx \quad (2.18)$$

2.5.1.3 Hong and Rappaport's mobility model

The following assumptions are made in the mobility model of Hong and Rappaport [15].

- They consider a homogeneous network where coverage is provided by hexagonal cells of radius $\Re$. To simplify analysis, they approximate the hexagonal cell by a circle which has the same area. The equivalent radius R of the approximating circle is:

$$R = \sqrt{\frac{3\sqrt{3}}{2\pi}}\Re \quad (2.19)$$

- They assume that the mobiles are evenly spread around the cell. Therefore, the distribution of the distance r and the direction ϕ (in polar co-ordinates) of the mobile from the centre of the cell is:

$$f_r(r) = \frac{2r}{R^2} \quad for\ 0 \leq r \leq R \quad (2.20)$$

$$f_\phi(\phi) = \frac{1}{2\pi} \quad for\ 0 \leq \phi \leq 2\pi \quad (2.21)$$

- They assume that a mobile travels in any direction with equal probability. They also assume that the mobile's direction remains constant during its travel within the cell.
- The speed V of a mobile is constant during its travel in a cell and is a random variable that is uniformly distributed between 0 and V_m with pdf:

$$f_V(v) = \frac{1}{V_m} \quad for\ 0 \leq v \leq V_m \quad (2.22)$$

They show that the pdf $f_n(t)$ of the new call cell dwell time is:

$$f_n(t) = \begin{cases} \dfrac{8R}{3\pi V_m t^2}\left(1-\left[1-\left(\dfrac{V_m t}{2R}\right)^2\right]^{3/2}\right) & 0 \le t \le \dfrac{2R}{V_m} \\ \dfrac{8R}{3\pi V_m t^2} & t \ge \dfrac{2R}{V_m} \end{cases} \tag{2.23}$$

They show that the pdf $f_h(t)$ of the hand-off call cell dwell time is:

$$f_h(t) = \begin{cases} \dfrac{4R}{\pi V_m t^2}\left(1-\left[1-\left(\dfrac{V_m t}{2R}\right)^2\right]^{1/2}\right) & 0 \le t \le \dfrac{2R}{V_m} \\ \dfrac{4R}{\pi V_m t^2} & t \ge \dfrac{2R}{V_m} \end{cases} \tag{2.24}$$

Hong and Rappaport then substitute the cdfs of the above cell dwell time distributions in equation (2.12) to obtain the cdf of the channel holding time.

2.5.1.4 Rappaport's mobility model

In a later paper [27], Rappaport proposed that most cell dwell times that occur in practice may be represented using a sum of N statistically independent negative exponentially distributed random variables denoted T_i where i = 1, 2, 3,..., N. The mean $E[T_i]=1/\mu_i$ and the variance $V[T_i]=1/(\mu_i)^2$. The mean cell dwell time $E[T_d]$ is:

$$E[T_d] = \sum_{i=1}^{N} E[T_i] = \sum_{i=1}^{N} \frac{1}{\mu_i} \tag{2.25}$$

The variance of the cell dwell time $V[T_d]$ is:

$$V[T_d] = \sum_{i=1}^{N} \frac{1}{(\mu_i)^2} \tag{2.26}$$

An interesting case occurs when μ_i does not change with i. Then T_d has an erlang-k distribution where $k = N$. Rappaport further assumes that the unencumbered call holding time is negative exponentially distributed. Thereafter, the channel holding time pdf may be obtained from these two distributions using equation (2.12).

2.5.1.5 Orlik and Rappaport's mobility model

In a later paper [28], Orlik and Rappaport pointed out the shortcomings of the cell dwell time distribution presented in section 2.5.1.4. Essentially, a random variable that is represented using a sum of N statistically independent negative exponentially distributed random variables will always have a squared coefficient of variation less than or equal to one. In order to accommodate more variable mobilities that have squared coefficient of variation greater than one, they proposed that cell dwell times T_d may be represented using a sum of N statistically independent hyper-exponentially distributed random variables denoted T_i where i = 1,2,3....N. Each random variable T_i is considered a phase of T_d, and each phase can have M_i stages where each stage is a negative exponentially distributed random variable $T_{i,n}$ where n=1,2,3...M_i. The mean $E[T_{i,n}]=1/\mu_{i,n}$ and the variance $V[T_{i,n}]=1/(\mu_{i,n})^2$. Since each phase T_i is a hyper exponential distribution, its pdf. $f_{Ti}(t)$ is as follows:

$$f_{Ti}(t) = \sum_{n=1}^{M_i} a_{i,n}\,\mu_{i,n}\,e^{-\mu_{i,n}t} \tag{2.27}$$

where the probabilities $a_{i,n}$ satisfy the relationship:

$$\sum_{n=1}^{M_i} a_{i,n} = 1 \tag{2.28}$$

The mean of the M_i-stage hyper-exponential random variable T_i is:

$$E[T_i] = \sum_{n=1}^{M_i} \frac{a_{i,n}}{\mu_{i,n}} \tag{2.29}$$

The variance of the M-stage hyper-exponential is:

$$V[T_i] = \left[\sum_{n=1}^{M_i} \frac{2a_{i,n}}{\mu_{i,n}^2}\right] - E[T_i]^2 \tag{2.30}$$

From this the mean $E[T_d]$ and the variance $V[T_d]$ of the cell dwell times that are a sum of N hyper-exponentials is:

$$E[T_d] = \sum_{i=1}^{N} E[T_i] = \sum_{i=1}^{N}\sum_{n=1}^{M_i} \frac{a_{i,n}}{\mu_{i,n}} \tag{2.31}$$

$$V[T_d] = \sum_{i=1}^{N} V[T_i] \tag{2.32}$$

In [28], Orlik and Rappaport assume that the unencumbered call holding time is negative exponentially distributed. In a later paper [30] they assume that the unencumbered call holding time also has a sum of hyper-exponential (SOHYP) pdf. Orlik and Rappaport suggest that the SOHYP pdf is a very general distribution and should be able to approximate all types of cell dwell time pdfs that occur in practice. So far we have discussed five mobility models based on analysis. We now look at some models in the literature based on simulation studies.

2.5.2 Simulation studies for cell dwell times and unencumbered call holding times

2.5.2.1 Guerin's mobility model

Guerin [17] employed a simple simulation to determine the nature of the channel holding time distributions. The assumptions are as follows:

- Cells are assumed to be circular in nature.
- All subscribers maintain the same constant speed for the entire duration of their calls.
- The subscriber's initial position is uniformly distributed over the cell area.
- The subscriber's initial direction of motion is uniformly distributed over $[0, 2\pi]$.
- The unencumbered call holding time is negative exponentially distributed.
- Changes in subscriber's direction of motion occur at random intervals that are negative exponentially distributed with a certain mean $1/\mu_C$. Therefore, the rate μ_C can be interpreted as the average number of changes in direction per unit time.

Guerin's simulation did not investigate cell dwell times directly and concentrated entirely on determining the nature of the channel holding time distributions. For reasonably sized cells, Guerin found that different rates of change in direction led to different types of channel holding time distributions. A very low rate of change in direction introduced large amounts of memory in the user's mobility. A very high rate of change in direction made the subscriber's mobility approximate a random walk (Brownian motion). Both these extreme rates of change in direction resulted in channel holding time distributions that were markedly different from the memory-less negative exponential distribution. However, Guerin found that an "reasonable" rate of change in direction led to channel holding time distributions that could be well approximated by the negative exponential distribution.

2.5.2.2 Zonoozi and Dassanayake's mobility model

In a more recent simulation study [18], Zonoozi and Dassanayake investigated cell dwell time distributions and channel holding time distributions under a more generalized mobility model compared to that of Guerin. The assumptions are as follows:

- Users are independent and uniformly distributed over the entire region served by the cellular network.
- Subscribers are allowed to move away from the starting point in any direction with equal probability.
- The change in the subscriber's direction along his / her path is taken to be from a uniform distribution limited in the range of $\pm\alpha$ degrees with respect to the current direction.
- The initial velocity of the subscriber is assumed to be a random variable that has a Gaussian probability density function truncated in the range [0, 100km/h] and mean of 50 km/h.
- The velocity increment of each mobile is taken to be a uniformly distributed random variable in the range ±10% of the current velocity.

Zonoozi and Dassanayake simulated the above network scenario and best fitted resulting cell dwell time distributions to generalized gamma distributions and then evaluated the agreement between the two using Kolmogorov-Smirnov goodness-of-fit test.

The pdf of the generalized gamma distribution is as follows:

$$f(t) = \frac{c}{b^{ac}\,\Gamma(a)}\, t^{ac-1}\, e^{-(t/b)^c} \qquad \textbf{(2.33)}$$

where Γ(a) is the gamma function and is given by:

$$\Gamma(a) = \int_0^\infty (x^{a-1})\, e^{-x}\, dx \qquad \textbf{(2.34)}$$

Assuming that the gamma distribution is the best-fit distribution, Zonoozi and Dassanayake found values for the parameters *a*,*b* and *c* for both the new call and hand-off call cases such that the maximum deviation between the observed and hypothesized distributions is a minimum. Their results are as follows:

$$
\begin{aligned}
a &= \begin{cases} 0.62, & \textit{new call} \\ 2.31, & \textit{handoff call} \end{cases} \\
b &= \begin{cases} 1.84\,R, & \textit{new call} \\ 1.22\,R, & \textit{handoff call} \end{cases} \\
c &= \begin{cases} 1.88, & \textit{new call} \\ 1.72, & \textit{handoff call} \end{cases}
\end{aligned} \tag{2.35}
$$

where R is the cell radius. Zonoozi and Dassanayake go on to show that the resulting channel holding time distribution is well represented by the negative exponential distribution. The author of this PhD thesis disputes this finding. The reason why the channel holding time distribution appears similar to the negative exponential distribution is that Zonoozi and Dassanayke considered a cell radius R that is reasonably large, namely, 3 km. For a mean unencumbered call holding time of 120 sec, setting the cell radius at 3 km leads to a hand-off probability of approximately 0.25 (see figure 15 in [18]). This means that only 25% of the calls require a hand-off and the remaining 75% of the calls terminate normally after completion of conversation. Therefore, the call departures in the cell are dominated by calls that terminate normally within the cell rather than those that require a hand-off to the neighboring cells. This leads to a channel holding time distribution that is dominated by the negative exponential unencumbered call holding times rather than the gamma cell dwell times. The result is that the channel holding time distribution appears similar to the negative exponential distribution. When the author of this thesis considered smaller cell sizes (eg. R<1km) for Zonoozi and Dassanayake's method, the channel holding time distributions no longer appeared to be negative exponential and, in fact, exhibited characteristics of the gamma distribution. This is to be expected since, a small cell size leads to a larger probability of hand-off and the cell dwell times (which are generalized gamma) dominate over the negative-exponential unencumbered call holding times.

2.5.3 Field measurements for channel holding times

Field measurements on live networks have tended to concentrate on determining the nature of channel holding time distributions rather than cell dwell time distributions. This is because the channel holding times are directly measurable from a live network whereas cell dwell times are more difficult (but not impossible) to measure.

A study of channel holding times for a cellular telephone system in Canada is reported in [31]. Data was measured on the BC Tel Mobility Cellular network for three separate cell sites. One was a highway cell, another was in downtown Vancouver and the last one in a small density area. The measured data was then cleaned to remove anomalies due to hand-off thresholds. The authors suggest that the observed channel holding times can be modeled using a log-normal distribution. A similar study of channel holding times for a TACS system in Barcelona is reported in [33]. The TACS mobile system uses Frequency Modulation (FM) and the FM detection of the downlink carrier frequency is considered sufficient for the knowledge of the channel occupancies. The measured data is then 'cleaned' to eliminate noise, interference etc. The filtered data is then fit to common distributions: the negative exponential, the shifted exponential, the Erlang-jk, the hyper-erlang and log-normal distributions. The parameters of the candidate distributions were determined using Maximum Likelihood Estimation. Thereafter, the Kolgomorov-Smirnov (K-S) goodness-of-fit test was used to evaluate the best-fit distribution amongst the candidate distributions. The authors conclude that the channel holding time distributions are best represented by the log-normal distribution. In fact, they propose a linear composition of three different log-normal distributions as the best fit for their measured channel holding times. However this is a computationally expensive procedure as it involves an eight-parameter estimation. Furthermore, it is difficult to accept the generality of the above results. Even the authors point out that a simple parameter like the average channel holding time varies from place to place and even network to network. They show that their measured average channel holding time was shorter than that of the study in the BC Tel Mobility Cellular network in Canada. The authors suggest that these differences can be "easily explained as being caused by the difference between the life-styles of Canada and Europe".

2.6 Summary

The performance analysis of cellular networks is a study in probability theory. As such, it is imperative that the relevant arrival and service processes are modelled accurately. We considered the following arrival and service processes:

- New call arrivals

- Hand-off call arrivals
- Unencumbered call holding times
- Cell dwell times
- Channel holding times.

In this chapter, we examined the various models proposed in the literature for the above five processes. We considered analytical models, simulation studies and field measurements results. In both simulation studies and field measurements, suitable distributions for the above processes were obtained by best fitting to the results. Although, these distributions are quite useful in their own right, it must be mentioned that they do not naturally lend themselves to elegant and tractable models during performance analysis.

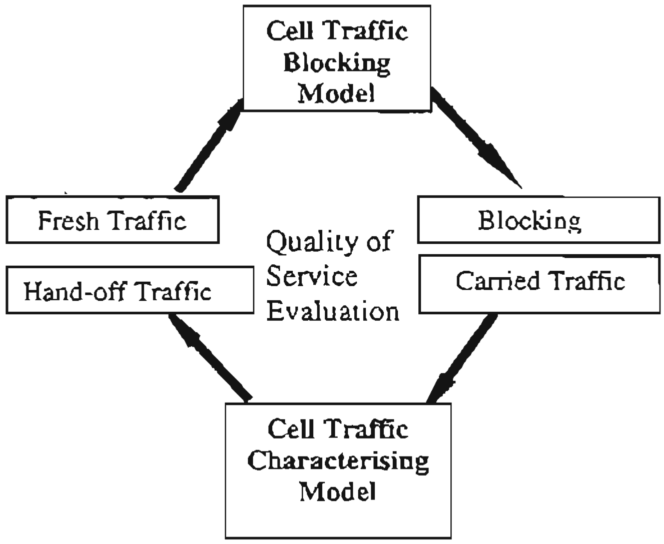
CHAPTER 3

3.1 Introduction

In the previous chapter, we examined the different arrival and service processes that need to be accurately modelled during the performance analysis of cellular networks. We discussed the various models that have been proposed in the literature for these parameters based on analysis, simulation or live network measurements. In this chapter, a survey is performed on the performance analysis models in the literature that make use of these models. As we shall show, almost all of these performance analysis models assume that the various arrival and service processes are memory-less negative exponential distributions or combinations of various negative exponential distributions. The reason for the use of a single or a combination of negative exponential distributions is that it simplifies queuing theory analysis by allowing the use of multi-dimensional Markov models [41,42].

Most performance analysis algorithms in the literature may be loosely decomposed into three parts, a *cell traffic characterising model*, a *cell traffic blocking model* and a *quality of service evaluation* model as shown in figure 3-1. In the cell traffic characterising model the new call and hand-off call traffic offered to each cell are characterised in some manner. In the cell traffic blocking model the blocking experienced by the various traffic streams offered to each cell is determined. The quality of service evaluation part is essentially a fixed point iteration (successive substitution) of the cell traffic characterising and cell traffic blocking parts, where, the outputs of the one are fed in as the inputs of the other and vice versa, until convergence is obtained. Once convergence is obtained, the customer orientated grade-of-service parameters such as forced termination probability may be determined. We now consider the various performance analysis algorithms proposed in the literature. Note that although the authors of the following performance analysis algorithms do not decompose their models into the three specific portions discussed above, we do so to simplify explanation.

Figure 3-1: Performance Analysis



3.2 Performance analysis algorithm by Macmillan [16], Chlebus [43] and Hong and Rappaport [15].

The performance analysis algorithms by Macmillan [16], Chlebus [43] and Hong & Rappaport [15] are very similar in that they employ the same assumptions for characterising new calls, hand-off calls, unencumbered call holding times and channel holding times. We examine the model by Hong and Rappaport [15] in greater detail here. We first consider the assumptions made in Hong and Rappaport's performance analysis algorithm and then discuss it in terms of a cell traffic characterising model, a cell traffic blocking model and a quality of service evaluation model.

3.2.1 Assumptions in Hong and Rappaport's model

- Hong and Rappaport assume a homogeneous network where each cell is identical to its neighbours in every aspect such as traffic statistics, user mobility, cell coverage etc.
- Each cell is assumed to have N neighbours. Given that a call is to be handed off, it has a fixed probability $R_j = 1/N$ of being handed-off to neighbour j.
- Each cell has C channels under a fixed channel allocation (FCA) method.
- Hong and Rappaport assume that *channel reservation* of level r is implemented in each cell. It is considered better to block a mobile user at call initiation rather than grant him/her service and subsequently drop a call during hand-off when the subscriber moves between different cells. To minimise this *forced termination* of hand-off calls, network operators normally employ channel reservation for hand-off calls where a few channels are reserved for the exclusive use of hand-off calls.
- Hong and Rappaport define the new call blocking in a cell, under channel reservation of level r, to be p_{nb}. They assume that all hand-off calls offered to a cell suffer the same hand-off call blocking p_{hb}.

3.2.2 Cell Traffic characterising model in Hong and Rappaport's model

Hong and Rappaport's elegant model was one of the earliest works on the performance analysis of cellular networks. They made the following assumptions about the various arrival and service processes:

- New calls arrive in each cell with a Poisson arrival rate of λ_n. This parameter is assumed to be known at the time of executing the performance analysis.
- The unencumbered call holding time, T_M, is negative exponentially distributed with mean $1/\mu_M$. See equations (2.3) and (2.4).
- The hand-off call arrival process is assumed to be Poisson with mean arrival rate per cell λ_h.
- The mobility model employed by Hong and Rappaport was discussed in section 2.5.1.3 of the previous chapter. For this mobility model, the new call cell dwell pdf $f_n(t)$ and

hand-off call cell dwell time pdf $f_h(t)$ are given by equations (2.23) and (2.24) respectively.

- The channel holding time distribution is negative exponentially distributed with cdf given by equation (2.12). Thereafter Hong and Rappaport best fit the resulting channel holding time distribution with a suitable member from the family of negative exponential distributions. Let us define the mean of this negative exponential distribution to be $1/\mu_H$.
- The mean new call offered traffic A_n and the mean hand-off offered traffic A_h are as follows:

$$A_n = \frac{\lambda_n}{\mu_H} \quad ; \quad A_h = \frac{\lambda_h}{\mu_H} \tag{3.1}$$

- Hong and Rappaport define Y_c to be the ratio of the mean carried hand-off attempt rate to the mean carried new call attempt rate in a cell:

$$Y_c = \frac{\lambda_h(1 - P_{hb})}{\lambda_n(1 - P_{nb})} \tag{3.2}$$

- Hong and Rappaport define P_N and P_H to be the probabilities with which a new call and a previously handed-off call would require a new hand-off. Therefore,

$$\begin{aligned} P_N &= Prob[T_M > T_n] = \int_0^\infty [1 - F_M(t)]\, f_n(t)\, dt \\ P_H &= Prob[T_M > T_h] = \int_0^\infty [1 - F_M(t)]\, f_h(t)\, dt \end{aligned} \tag{3.3}$$

where $F_M(t)$ is the cumulative distribution function (cdf) of T_M.

3.2.3 Cell Traffic Blocking Model

The cell traffic blocking model in Hong and Rappaport's paper is used to determine the blocking experienced by the new and hand-off call offered traffic. The cell traffic blocking model was originally presented by Akinpelu [44] for fixed network analysis. Assume that two Poisson streams having arrival rates λ_n and λ_h occur in a cell L of C servers. Assume also that the first arrival stream λ_n experiences channel reservation of level r. This implies that stream λ_h has access to all C channels in cell L whereas stream λ_n has access to only $C - r$. It is also assumed that the call departures from state k in cell L occur at the call departure rate $\mu_k = k\mu_H$.

Akinpelu modelled the above scenario using a state-dependent Birth-Death model where the arrival rates in states $k = 0....C\text{-}r\text{-}1$ is a Poisson arrival rate of intensity λ_T where $\lambda_T = \lambda_n + \lambda_h$ and the arrival rates in states $k = C\text{-}r, C\text{-}r+1,.....,C$ is a Poisson arrival rate of intensity λ_R where $\lambda_R = \lambda_h$. The state-dependent arrival rate is used to model the effect of channel reservation. The equilibrium equations for this scenario may be easily derived by equating probability transitional rates to and from each state:

$$\begin{aligned}
&\text{for } k = 0:\ \lambda_T\, p_0 = \mu_H\, p_1 \\
&\text{for } 0 < k < C-r:\ (\lambda_T + k\mu_H)\, p_k = (k+1)\mu_H\, p_{k+1} + \lambda_T\, p_{k-1} \\
&\text{for } k = C-r:\ (\lambda_R + (C-r)\mu_H)\, p_{C-r} = (C-r+1)\mu_H\, p_{C-r+1} + \lambda_T\, p_{C-r-1} \\
&\text{for } C-r < k < C:\ (\lambda_R + k\mu_H)\, p_k = (k+1)\mu_H\, p_{k+1} + \lambda_R\, p_{k-1} \\
&\text{for } k = C:\ C\mu_H\, p_C = \lambda_R\, p_{C-1}
\end{aligned} \tag{3.4}$$

Re-arranging the above equation gives:

$$p_k = \frac{\lambda_T^{\,k}}{\mu_H^k\, k!}\, p_0 \quad \text{for } k = 0,1,....C-r \tag{3.5}$$

and:

$$p_k = \frac{\lambda_R^{\,k-C+r}}{\mu_H^{k-C+r}\,\dfrac{k!}{(C-r)!}}\, p_{C-r} \quad \text{for } k = C-r, C-r+1,......C \tag{3.6}$$

The state probabilities p_k may be solved for using the following initial condition:

$$\sum_{k=0}^{C} p_k = 1 \tag{3.7}$$

There are two call congestion probabilities of interest, the call congestion p_{nb} experienced by stream λ_n and the call congestion p_{hb} experienced by stream λ_h. Note that the time congestion probability E which is, by definition, the fraction of time when all C channels are busy, is the same as the call congestion probability p_{hb} of stream λ_h, since, stream λ_h sees congestion only when cell L is in its final state C.

The call congestion p_{hb} experienced by the stream λ_h that is not subjected to channel reservation is:

$$p_{hb} = p_C \tag{3.8}$$

The call congestion p_{nb} experienced by the stream λ_n that is subjected to channel reservation from states *C-r* to *C* is given by:

$$p_{nb} = \sum_{k=C-r}^{C} p_k \qquad (3.9)$$

3.2.4 Quality of service evaluation

Under the network homogeneous assumption, each cell is identical to the other and therefore it is sufficient to model a single cell. For this cell, the outputs of the cell traffic characterising model are fed in as the inputs of the cell traffic blocking model and vice versa in what is known as a fixed-point iteration. This fixed-point iteration is started with the assumption that only new calls are offered to each cell. The iteration is stopped when sufficient convergence is obtained. The fixed-point iteration is a useful and efficient way to solve the *non-linear* set of equations that describe the system. It is not difficult to show that the system equations are non-linear: In the cell traffic blocking model it can be shown that the hand-off call blocking p_{hb} is dependent on the mean hand-off arrival rate λ_h. In the cell traffic characterising model, it can be shown that the mean hand-off arrival rate λ_h is in turn dependent on the hand-off call blocking p_{hb} (see equation 3.10).

3.2.4.1 Application of the cell traffic blocking model

Given that the mean hand-off arrival rate λ_h and the mean channel holding time are known from the last iteration of the cell traffic characterising model, the new call blocking p_{nb} and hand-off call blocking p_{hb} may be determined for a finite cell *L* of *C* channels under channel reservation of level *r* using Akinpelu's model.

3.2.4.2 Application of the cell traffic characterising model

Once the blocking values are known, new values for the hand-off arrival rate λ_h and mean $1/\mu_H$ of the negative exponential distribution may be determined as follows:

Hong and Rappaport employ an elaborate argument, involving the determination of the mean number of hand-off calls that each new call may undertake without experiencing blocking and prior to call termination, to determine the hand-off arrival rate λ_h. A simpler but equivalent method was proposed by Chlebus [43] using balance of flow. Chlebus shows that if there are N neighbours for each cell, then under the network homogenous assumption, each of the N neighbours of cell L will offer $1/N$ of their carried calls that which requests for a further hand-off to cell L. Therefore, the hand-off arrival rate λ_h^{j+1} as determined by the $j+1^{th}$ iteration is simply:

$$\lambda_h^{j+1} = N\left(\frac{P_N\,\lambda_n\,(1-p_{nb}) + P_H\,\lambda_h^j\,(1-p_{hb})}{N}\right) = P_N\,\lambda_n\,(1-p_{nb}) + P_H\,\lambda_h^j\,(1-p_{hb}) \quad (3.10)$$

Note that Chlebus does not accommodate channel reservation in his model in [43], however, it is elementary to extend his model as shown in the above equation to allow for the fact that new calls and hand-off calls experience different blocking under channel reservation of level r.

The cdf of the channel holding time (equation 2.12) is dependent on the ratio of the new hand-off call carried rate to the ratio of the new call carried rate Y_c (equation 3.2), which in turn is dependent on the mean hand-off arrival rate λ_h. Since the mean hand-off arrival rate is re-calculated at every iteration, it is necessary that the cdf of the channel holding time is re-calculated as well. Thereafter, the best fit member from the family of negative exponential distributions is chosen to represent the actual channel holding time distribution. For this purpose, Hong and Rappaport suggest that the mean $1/\mu_H$ of the negative exponential distribution may be chosen such that the following expression is minimised:

$$G = \frac{\int_0^\infty \left| F_H^C(t) - e^{-\mu_H t} \right| dt}{2\int_0^\infty F_H^C(t)\; dt} \quad (3.11)$$

where $F_H^C(t)$ is the complementary cumulative distribution function (survivor function) of the channel holding time: $F_H^C(t)=1 - F_H(t)$. A value of $G=0$ specifies an exact fit and a value of $G=1$ indicates a bad fit.

Once new values for the hand-off arrival rate λ_h and mean $1/\mu_H$ of the negative exponential distribution are known they are used in the cell traffic blocking model to obtain new values for the new call blocking and hand-off blocking. The fixed-point iteration is continued until the difference between the values obtained from successive iterations is smaller than a pre-set value. Once convergence is obtained, the forced termination probability, F_T, which is the probability with which a call is dropped by an unsuccessful handover, may be calculated as:

$$F_T = \sum_{i=0}^{\infty} \text{Prob}\{\textit{Call dropping after i successful handovers}\}.\text{Prob}\{\textit{i successful handovers}\} \qquad (3.12)$$

$$F_T = \sum_{i=0}^{\infty} P_N\, p_{hb} [P_H (1-p_{hb})]^i = \frac{P_N\; p_{hb}}{1-P_H(1-p_{hb})} \qquad (3.13)$$

3.3 Performance analysis algorithm by Rappaport using sum of exponential cell dwell time distributions.

In a later paper, Rappaport [27] proposed a performance analysis algorithm that uses a sum of independent negative exponential random variables. The main driver for this generalisation was, in Rappaport's view, that the classical simple negative exponential distribution, although very useful in tele-traffic analysis, was limited when it comes to modelling different types of cell dwell time distributions. He suggests the highway scenario as a prime example where the negative exponential distribution fails. In highway vehicular traffic streams, the subscribers, who are occupants in moving vehicles, tend to move at fairly constant speeds, along essentially fixed paths. Because of the memory aspect of such mobility, the cell dwell times are likely to be highly concentrated about some mean. The generalisation to a sum of independent negative exponential variables as we discussed in section 2.5.1.4 accommodates this type of subscriber mobility.

To simplify explanation of Rappaport's model, we decompose it into three parts, namely, a cell traffic characterising model, a cell traffic-blocking model and a quality of service evaluation model. Moreover, Rappaport's model is an intricate one that allows for various other generalisations. For example, it allows for a number of different *platforms*. A platform, in a sense, is a mode of transport for a person requiring mobile services. Platform types include pedestrians with handsets, mobile subscribers in moving cars and mobile subscribers in high-speed trains etc. To simplify the discussion on Rappaport's model we concentrate on the scenario where there is only one type of platform, be it pedestrians or cars. Similar to Rappaport we assume that a single platform can only support a single call. The reader is directed to the original work [27] for the extension to multiple platforms.

3.3.1 Assumptions

- A homogeneous network is considered. Each cell is identical to the other in every respect.
- Each cell can accommodate C channels.
- Channel reservation of level r is implemented.
- Each cell is assumed to have M neighbours. Given that a call is to be handed off, it has a fixed probability equal to $1/M$ of being handed-off to neighbour j.

3.3.2 Cell traffic characterising model

The various arrival and service processes are defined as follows:

- The new call origination rate for a single platform is assumed to be δ. The interval between successive call offerings by the same platform is implicitly assumed to be negative exponentially distributed.
- The number of non-communicating platforms in any cell is denoted ν.
- The total rate at which new calls are generated in a cell is denoted λ_n. Thus, $\lambda_n = \delta\nu$. Strictly speaking, the new call arrival rate in a cell λ_n should be state dependent and decrease as more and more platforms start communicating (as is the case in an Engset arrival process [45,46]). However, Rappaport assumes that $\nu >> C$ and under an "infinite" population model (with the word "infinite" being used in the loosest of terms) only a small fraction of platforms will be served at any one time and therefore the above approximation is reasonably valid. The new call arrival rate, λ_n, is assumed to be known at the time of effecting the performance analysis.
- The unencumbered call holding time T_M is negative exponentially distributed with mean $E[T_M]=1/\mu_M$.
- The cell dwell times are a sum of N independent negative exponentially distributed random variables and are described in section 2.5.1.4. Rappaport proposed that most cell dwell times that occur in practice may be represented using a sum of N statistically

independent negative exponentially distributed random variables denoted T_i where i = 1, 2, 3,..., N. Each NED random variable is considered a phase of the total cell dwell time and has mean mean $E[T_i]=1/\mu_i$.

- The hand-off call arrival process is implicitly defined and will be discussed later. However, it is important to point out that it is assumed to be a Poisson process.
- The probability P_N that a new call, which has been accepted into service, would require a hand-off may be calculated in the following manner. Since the new call arrival rate is Poisson, the fraction of new call arrivals in a cell that arise when the platform is in phase i is:

$$\rho_n(i) = \frac{E[T_i]}{E[T_d]} \tag{3.14}$$

Let π_i be the probability that a platform completes its current dwell time phase before its call is satisfactorily completed. Since the dwell times are negative exponential, we have,

$$\pi_i = \frac{\mu_i}{\mu_i + \mu_M} \tag{3.15}$$

For a call that is being served on a platform in phase i, N-i+1 dwell time phases must be completed by the platform for a hand-off attempt to be generated. Therefore the probability that a platform in phase i requires a hand-off is

$$P_N^i = \prod_{j=i}^{N} \pi_j = \prod_{j=i}^{N} \frac{\mu_j}{\mu_j + \mu_M} \tag{3.16}$$

The overall probability P_N that a new call (which is not blocked) requires a hand-off is therefore:

$$P_N = \sum_{i=1}^{N} \rho_n(i)\, P_N^i = \sum_{i=1}^{N} \rho_n(i) \prod_{j=i}^{N} \pi_j = \sum_{i=1}^{N} \rho_n(i) \prod_{j=i}^{N} \frac{\mu_j}{\mu_j + \mu_M} \tag{3.17}$$

- When calls are successfully handed off to a target cell, the platforms begin their dwell times from the first phase. The probability P_H that a previously handed-off call requires a new hand-off is:

$$P_H = P_N^1 = \prod_{j=1}^{N} \frac{\mu_j}{\mu_j + \mu_M} \tag{3.18}$$

3.3.3 Cell traffic blocking model

Rappaport defines the state of a cell at an instant in time by a sequence of N non-negative numbers:

$$v_1, v_2, v_3, v_4, \ldots\ldots\ldots\ldots\ldots v_N \quad (3.19)$$

where v_i $\{i=1,2,3.....N\}$ is the number of platforms that are in phase i of their cell dwell time. Essentially, the entire sequence $v_1,v_2,......v_N$ defines the state of a cell. Each phase of the cell dwell time is negative exponentially distributed with mean $E[T_i]=1/\mu_i$. For convenience sake, Rappaport orders the various possible states (or sequences of numbers) using an index $s = 0,1,2,....s_{max}$. Then the state variable v_i can be shown to be explicitly dependent on the state. That is, $v_i=v(s,i)$. A state is a permissible ordering of sequence numbers of the type shown above. When the cell is in state s, the number of channels being used by the platforms is.

$$j(s) = \sum_{i=1}^{N} v(s,i) \quad (3.20)$$

The channel size constraint requires that the number of calls that can be supported in state s is:

$$j(s) \le C \quad (3.21)$$

Rappaport identifies five relevant driving processes:

- {n} the generation of new calls in the cell of interest;
- {c} the completion of calls in the cell of interest;
- {h} the arrival of communicating vehicles at the cell of interest;
- {d} the departure of communicating calls from the cell of interest;
- {ϕ} the transition between dwell times;

Under Poisson new call arrivals, Poisson hand-off call arrivals, negative exponential phases in the cell dwell time and negative exponential unencumbered call holding times, the problem is amenable to solution using multi-dimensional birth-death processes. For each state, Rappaport identifies the possible predecessor states, which can give rise to the current step with just one step transition. In addition, the state probability transitions are found. For convenience, in the following paragraphs, we employ the dummy index z where $z=1,2,\ldots N$.

3.3.3.1 New call arrivals

A single step transition into state s, due to a new call arrival when a platform is in phase i when the cell is in state x_n, will cause the state variable $v(x_n,i)$ to be incremented by 1. Because of channel reservation, a new call can be served if and only if the number of channels in use is less than C-r. Therefore a permissible state x_n is a predecessor state of s for a new call arrival for platforms in phase i if $j(x_n)<C-r$ and the state variables are related by

$$\begin{aligned} v(x_n,i) &= v(s,i)-1 \\ v(x_n,z) &= v(s,z) \qquad for\ z\neq i \end{aligned} \tag{3.22}$$

Let $\lambda_n(i)$ denote the average arrival rate per cell of new calls for the platforms in phase i of the dwell time. Since the new call arrival rate is Poisson, the fraction of new call arrivals in a cell that arise when the platform is in phase i is:

$$\rho_n(i) = \frac{E[T_i]}{E[T_d]} \tag{3.23}$$

Thus:

$$\lambda_n(i) = \rho_n(i)\,\lambda_n = \lambda_n \frac{E[T_i]}{E[T_d]} \tag{3.24}$$

Let $\gamma_n(s,x_n)$ denote the flow rate into state s from state x_n when new call arrivals occur. Therefore:

$$\gamma_n(s,x_n) = \lambda_n(i) = \rho_n(i)\,\lambda_n = \lambda_n \frac{E[T_i]}{E[T_d]} \tag{3.25}$$

3.3.3.2 Call completions

A single step transition into state s, due to a call completion when a platform is in phase i when the cell is in state x_c, will cause the state variable $v(x_c,i)$ to be decremented by 1. Therefore a permissible state x_c is a predecessor state of s for a call completion for platforms in phase i if the state variables are related by

$$\begin{aligned} v(x_c,i) &= v(s,i)+1 \\ v(x_c,z) &= v(s,z) \qquad for\ z\neq i \end{aligned} \tag{3.26}$$

Let $\gamma_c(s,x_c)$ denote the flow rate into state s from state x_c when a call departure occurs. Therefore:

$$\gamma_c(s,x_c) = \mu_M \, v(x_c,i) \tag{3.27}$$

3.3.3.3 Hand-off arrivals

A hand-off arrival is always due to a platform entering a cell and therefore it corresponds to a platform in phase 1. Therefore a transition into state s due to a hand-off arrival when the cell is in state x_h will cause the state variable $v(x_h,1)$ to be incremented by 1. Moreover, a permissible state x_h is a predecessor state of s for hand-off arrivals if $j(x_h)<C$ and the state variables are related by

$$\begin{aligned} v(x_h,1) &= v(s,1) - 1 \\ v(x_h,z) &= v(s,z) \qquad for \;\; z \neq 1 \end{aligned} \tag{3.28}$$

Let $\gamma_h(s,x_h)$ denote the flow rate into state s from state x_h when hand-off call arrivals occur. Therefore:

$$\gamma_h(s,x_h) = \lambda_h \tag{3.29}$$

3.3.3.4 Hand-off departure

A hand-off departure is always due to a platform completing its last phase. Therefore a transition into state s due to a hand-off departure when the cell is in state x_d will cause the state variable $v(x_d,N)$ to be decreased by 1. Moreover, a permissible state x_d is a predecessor state of s for hand-off departures if the state variables are related by

$$\begin{aligned} v(x_d,N) &= v(s,N) + 1 \\ v(x_d,z) &= v(s,z) \qquad for \;\; z \neq N \end{aligned} \tag{3.30}$$

Let $\gamma_d(s,x_d)$ denote the flow rate into state s from state x_d when hand-off call departures occur. Therefore:

$$\gamma_d(s,x_d) = \mu_N \, v(x_d,N) \tag{3.31}$$

3.3.3.5 Dwell time phase transitions

A transition into state s due to the completion of a dwell time phase of a platform in phase i when the cell is in state x_ϕ will cause change in two state variables simultaneously. The state variable $v(x_\phi,i)$ will be decreased by 1 and the state variable $v(x_\phi,i+i)$ will be increased by 1. This corresponds to a phase transition from i to $i+1$. Therefore, a permissible state x_ϕ is a predecessor state of s for phase transitions of platforms in phase i if the state variables are related by

$$\begin{aligned} v(x_\phi, i+1) &= v(s, i+1) - 1 \\ v(x_\phi, i) &= v(s, i) + 1 \\ v(x_\phi, z) &= v(s, z) \quad for \;\; z \neq i \end{aligned} \tag{3.32}$$

where $i=1,2,3....N-1$. Let $\gamma_\phi(s, x_\phi)$ denote the flow rate into state s from state x_ϕ when dwell time phase transitions occur. Therefore:

$$\gamma_\phi(s, x_\phi) = \mu_i \, v(x_\phi, i) \tag{3.33}$$

3.3.3.6 Flow balance equations

From the above equations, the total transitional flow rate into state s from any permissible state x can be found using:

$$q(s,x) = \gamma_n(s,x) + \gamma_c(s,x) + \gamma_h(s,x) + \gamma_d(s,x) + \gamma_\phi(s,x) \tag{3.34}$$

in which $s \neq x$. The total flow out of state s is denoted $q(s,s)$ and is given by:

$$q(s,s) = -\sum_{\substack{k=0 \\ k \neq s}}^{S_{max}} q(k,s) \tag{3.35}$$

To find the statistical equilibrium state probabilities for a cell, the flow balance equations for the states are written. This results in a set of $s_{max}+1$ simultaneous equations for the unknown state probability $p(s)$. They are:

$$\begin{aligned} &\sum_{j=0}^{S_{max}} q(i,j)\, p(j) = 0 \quad where \;\; i = 0,1,2,3........S_{max} - 1 \\ &\sum_{j=0}^{S_{max}} p(j) = 1 \end{aligned} \tag{3.36}$$

in which, for $i \neq j$, $q(i,j)$ represents the net transition flow into state i from state j, and $q(i,i)$ is the total transitional rate out of state i. State probabilities $p(j)$ can be solved using the above conditions. The new call blocking and hand-off blocking may be determined as follows.

3.3.3.7 New call blocking

The new call blocking probability is the average fraction of new calls that are denied access to a channel. Under channel reservation, blocking of new calls occur when the number of occupied channels is equal to or greater than *C-r*. Rappaport defines the following set of states:

$$B_0 = \{s: \quad C - r \leq j(s) \leq C\} \tag{3.37}$$

Then the blocking probability for a new call is:

$$p_{nb} = \sum_{s \in B_0} p(s) \tag{3.38}$$

3.3.3.8 Hand-off call blocking

The hand-off call blocking is the average fraction of hand-off attempts that are denied a channel. Since hand-off traffic benefits from channel reservation, the blocking of hand-off calls occur when the number of occupied channels equals *C*. Rappaport defines the following set of states:

$$H_0 = \{s: \quad j(s) = C\} \tag{3.39}$$

Then the blocking probability for a hand-off call is:

$$p_{hb} = \sum_{s \in H_0} p(s) \tag{3.40}$$

3.3.4 Quality of service evaluation

Under the network homogeneous assumption, each cell is identical to the other and therefore it is sufficient to model a single cell. For this cell, the outputs of the cell traffic characterising model are fed in as the inputs of the cell traffic blocking model and vice versa in a fixed point iteration. The iteration is stopped when sufficient convergence is obtained.

3.3.4.1 Application of the cell traffic blocking model

Given that the mean hand-off arrival rate λ_h is known from the last iteration of the cell traffic characterising model, the new call blocking p_{nb} and hand-off call blocking p_{hb} may be determined in a finite cell L of C channels under channel reservation of level r using the cell traffic blocking model presented above.

3.3.4.2 Application of the cell traffic characterising model

Once the blocking values are known, new values for the hand-off arrival rate λ_h may be determined. It can be shown that if there are M neighbours for each cell, then under the network homogenous assumption, each of the M neighbours of a cell L will offer $1/M$ of their carried calls that which requests for a further hand-off to cell L. Therefore, the hand-off arrival rate λ_h^{j+1} as determined by the $j+1^{\text{th}}$ iteration is simply:

$$\lambda_h^{j+1} = M\left(\frac{P_N\,\lambda_n(1-p_{nb}) + P_H\,\lambda_h^j(1-p_{hb})}{M}\right) = P_N\,\lambda_n(1-p_{nb}) + P_H\,\lambda_h^j(1-p_{hb}) \qquad (3.41)$$

Once new values for the hand-off arrival rate λ_h are determined they are used in the cell traffic blocking model to obtain new values for the new call blocking and hand-off blocking. The fixed point iteration is continued until the difference between the values obtained from successive iterations is smaller than a pre-set value. Once convergence is obtained, the forced termination probability, F_T, which is the probability with which a call is dropped by an unsuccessful handover, may be calculated as:

$$F_T = \sum_{i=0}^{\infty} \text{Prob}\{\textit{Call dropping after i successful handovers}\}.\text{Prob}\{\textit{i successful handovers}\} \qquad (3.42)$$

$$F_T = \sum_{i=0}^{\infty} P_N\, p_{hb}[P_H(1-p_{hb})]^i = \frac{P_N\ p_{hb}}{1-P_H(1-p_{hb})} \qquad (3.43)$$

In the paper, Rappaport presents extensions to the model discussed above to allow for multiple platforms, for non-homogenous networks and for the imperfect detection of the need for hand-offs.

Although Rappaport's model makes very useful extensions, it has two important shortcomings. The first shortcoming, to which the author alludes, is that the performance analysis algorithm suffers from *state space explosion*. Consider a scenario where there are two types of platforms and that the dwell times of each type of platforms may be described using the sum of exactly two exponentials. The following table [27] provides the number of states required to analyse a homogeneous network where there are *C* channels in each cell.

Table 3-1: Number of states needed for a particular cell size

C	Number of states needed
15	3876
20	10626
25	23751
30	46376
36	91390

The other shortcoming is that Rappaports's model, even with all the interesting extensions presented, is essentially a single moment method. All call arrival processes (new and hand-off) are represented using only the mean of the processes, as Poisson processes. In practice, hand-off traffic is generated as a result of the mobility of the users. There is no evidence to suggest that hand-off traffic is Poisson in real-life. Since Rappaport does not validate his results with, for example, simulation results it is unclear what is the effect of ignoring the higher moments. As we shall show in the subsequent chapters, higher moments are very important for characterising hand-off traffic, but the trick is avoid the excessive state space explosion that occurs when higher moments are modelled.

3.4 Performance analysis algorithm by Orlik and Rappaport using sum of hyper-exponential cell dwell time distributions

In other work [28], Orlik and Rappaport present a performance analysis model for cellular networks using a sum of hyper-exponential (SOHYP) distributions to model cell dwell times. This model makes the exact same assumptions as the model we discussed in section 3.3 (eg. NED unencumbered call holding time, homogeneous network, Poisson call arrivals etc), except that now the cell dwell times are SOHYP distributed rather than a sum of exponentials. In a later paper [30], they suggest that the cell dwell time pdf and unencumbered pdf can both be modelled using two separate SOHYP distributions.

We do not delve into great detail about these models because they are very much in a similar vein to the model we discussed in the previous section. Admittedly, the use of the SOHYP distribution allows a wider ranger of cell dwell times to be accommodated than the simple sum of exponentials. The SOHYP can model cell dwell time distributions that are highly variable. It also has the added advantage that its pdf $f(t)$ has the useful and practical property, $f(t=0)=0$. The use of SOHYP cell dwell times and SOHYP unencumbered call holding times allow the resulting performance analysis algorithm to be amenable to solution using Markovian multi-dimensional birth-death models.

Similar to the earlier model, there exist two shortcomings in these performance analysis algorithms. Firstly they suffer from state space explosion, in fact, to a greater degree than the case where the cell dwell times are a sum of exponentials. Secondly, both models ignore the higher moments of the hand-off arrival process and assume it to be Poisson.

3.5 Performance analysis algorithm by Rappaport for the multiple call hand-off problem

In another paper [29], Rappaport considers the multiple hand-off scenario. Consider a train with mobile subscribers travelling on it. When the train moves out of the coverage area of a cell and into the coverage area of another cell, the target base station will be inundated with multiple requests for hand-off since all the communicating subscribers in the train would request for a hand-off. Rappaports' solution for this scenario is a model similar to that we described in section 3.3. The assumptions are that the new and hand-off arrivals are both Poisson and the unencumbered call holding time pdf and cell dwell time pdf are both negative exponentially distributed. The multiple hand-off requests are dealt with using an arrival process similar to *bulk* arrival processes [47]. The advantage is that this approach to the multiple hand-off scenario still leaves the problem amenable to solution using multi-dimensional birth-death models. The disadvantages, as before, are that there occurs a state space explosion even for reasonably sized channel sizes and the hand-off arrival processes are modelling using only their respective means.

3.6 Performance analysis algorithm by Long-Rong Hu and Rappaport for the multiple hierarchical cellular overlays

Hierarchical cellular overlays are a relatively recent proposition in wireless networks. Service areas are assumed to be covered by micro-cells, with overlaying macro cells covering spots that are difficult in radio propagation for microcells and thus providing overflow groups for clusters of microcells. At the highest hierarchical level, satellite beams [48,49,50] may be used as overflow groups for clusters of macro-cells. Hierarchical cellular networks are useful for serving users with different mobilities: users with low mobilities may be served by a lower level microcell whereas users with high mobility may be served by a geographically larger macro-cell. This enables the number of requests for hand-off by the high mobility group to be kept at a manageable level.

Long-Rong Hu and Rappaport [9] consider a three level hierarchy with two types of user mobility: low speed and high speed. Users in each mobility group has a preferred layer of use and if no channels are available in the appropriate cell in the preferred layer, calls are attempted to be handed off to a higher layer. In this paper the authors assume that calls may be not offered by a higher layer to a lower layer.

The authors assume Poisson new and hand-off call arrivals and consider NED unencumbered call holding times and cell dwell times. Users in different mobility groups are assumed to have dwell times with different means. The offering of a call from a lower layer to an upper layer is similar to the overflow scenario in the fixed or wireline networks. The authors borrow a popular model from the wireline case to model the overflow process, namely, the Interrupted Poisson process [51]. The authors go on to show how customer orientated grade of service parameters such as new call blocking and forced termination probability may be calculated using the above model.

3.7 Performance analysis algorithm by S.A. El-dolil, WC Chong and R. Steele for the highway microcells with macro cellular overlays

El-dolil, Chong and Steele [7] present a performance analysis algorithm for highway cellular networks serving vehicular traffic. They propose a vehicular mobility model and then represent the cell dwell times, for the purposes of analysis, using a negative exponential distribution. This is one of the few papers in the literature that deviates from Poisson new call arrivals. The authors assume a finite population model and employ the Engset arrival process [45,46] to model new call arrivals. They also model the hand-off arrival process using a modified Engset process where the parameters of this Engset process are network dependent. Unencumbered call holding times, cell dwell times and channel holding times are all assumed to be negative exponentially distributed. Using these arrival and service processes the authors show how customer orientated grade of service parameters such as new call blocking and forced termination probability may be determined. The authors then go on to describe an extension to their highway microcellular network scenario by allowing for a overlaying macro-cellular network that receives calls that could not find spare channels in the micro-cellular layer.

3.8 Performance analysis algorithm by M. Sidi and D. Starobinski for homogenous cellular networks.

Sidi and Starobinski [21] attempt to model large-scale homogeneous networks using building blocks of simple two and three cell scenarios. The new call arrival is considered to be Poisson and the unencumbered call holding time, cell dwell times, and channel holding times are all assumed to be negative exponentially distributed. This is one of the few models that attempt to model hand-off arrival process as an arrival process other than Poisson. For simple two cell / three cell scenarios, the inter-hand off traffic between the cells in question is implicitly modelled as a general arrival process that is determined by the network parameters and the mobility of the users. However, the hand-off traffic from cells outside the two-cell or three-cell scenarios are assumed to be Poisson in nature. This model is, in a way, a compromise on modelling all the hand-off traffic generated by all the cells in a network as a general arrival processes. The advantages of this model is that it is one of the few models to show that, under the condition of zero channel reservation where both hand-off traffic and fresh traffic are treated equally, hand-off calls and new calls may experience different blocking levels because they are different types of arrival processes. The disadvantages are that the model suffers from state-space explosion and also cannot accommodate channel reservation in the exact form proposed by the authors. The latter is a serious limitation in this imaginative performance analysis model since channel reservation is an all important control measure in cellular networks.

3.9 Extensions to existing performance analysis algorithms

We now look at some performance analysis algorithms, which model the effect of a performance enhancing feature or a previously ignored characteristic of a cellular network. These models generally make very simple assumptions regarding the arrival and service processes in the cellular network.

3.9.1 Speed sensitive hand-off strategy in hierarchical cellular networks

Jabbari and Fuhrman [24] present a performance analysis method for a two-tier hierarchical network consisting of a layer of micro-cells and an over-sailing layer of macro-cells. The authors assume that there are two types of mobile subscribers, slow moving subscribers (pedestrians) and fast moving subscribers (occupants in vehicles). During new call attempts and hand-off attempts, based on speed sensitive measurements, the network tries to allocate channels from the macro-cell layer for fast moving mobiles and channels from the micro-cell layer for slow moving mobiles. However, if there is no capacity in the macro-cell layer, fast moving mobiles are allowed to overflow to the micro-cell layer and similarly the slow moving mobile subscribers are allowed to overflow to the macro-cell layer during congestion in the target cell. The authors also consider a take-back strategy where a fast moving subscriber who overflowed to a micro-cell will attempt at the next hand-off to move back to the macro-cell layer. A similar take-back strategy is assumed for slow moving subscribers who overflowed to the macro-cell layer. The authors analysed the above scenario under Poisson new call arrivals, Poisson hand-off call arrivals, Poisson overflow calls, negative exponential dwell times, negative exponential unencumbered call holding times and a fluid flow mobility model. The authors do not consider any channel reservation. Under the above assumptions, this performance analysis algorithm can be considered to be an extension of the classical methods by Hong and Rappaport [15], Mcmillan [16] or Chlebus [43] to apply for the specific case of speed sensitive hand-off strategy. The authors show that the Erlang-B model gives the new call blocking under the Poisson call arrival assumptions. They also present explicit expressions for the forced termination probability of slow moving and fast moving mobiles.

3.9.2 Model for Directed Retry in cellular mobile networks

Yum and Leung [52] present an analytical model for the performance analysis of cellular networks under the mechanism of directed retry. Due to irregular terrain and interference of the buildings, the radio coverage of a cell is often irregular. In order to ensure a network with full radio coverage over a given area, at least 30% overlap of cells is required. Mobiles in the overlap regions can generally maintain calls of reasonable quality with any of the base stations that contribute to the overlap region.

In directed retry, a mobile unit in cell i initiates a call request on a signalling channel to its dominant base station. After the reception of the request, the base station checks to see if there is a free channel in cell i. If there is, the base station will allocate a free channel to the mobile; if there is no free channel, the base station will provide the mobile all its neighbouring cells' information. The mobile checks the quality of the links to the neighbouring cells and if the best one of them exceeds a pre-set threshold, the call is set-up in the neighbouring cell. If none of the links are of sufficient quality, the call attempt is rejected.

Yum and Leung analyse the mechanism of directed retry using Poisson call arrivals and negative exponential call durations. Consequently the blocking B_i experienced in a cell i of C_i channels and being offered total traffic A_i, is given by the Erlang-B formula (equation 2.2). Assuming that cell i has S_i neighbours and assuming that f_i is the fraction of coverage overlap, Yum and Leung derive the overall blocking probability in the system in terms of the above parameters. They also present an analytical model for directed retry for hand-off calls.

3.9.3 Performance analysis algorithm for customer retrial phenomenon

The work by Tran-Gia and Mandes [22] considers the modelling of the customer retrial phenomenon in cellular networks. They present two models in this paper. Their first model uses Markov modelling to analyse the repeated call attempt pheoomena in a cell with a finite customer population. Under this finite population model, they assume that a user, who finds his call attempt rejected by the network, will wait for a finite interval that is negative exponentially distributed and re-try with a finite probability. In their second model, they consider an infinite population model where there are two arrival processes to a cell: fresh calls and hand-off calls. Blocked fresh calls can cause retrials but hand-off calls do not have re-attempts although they enjoy a better protection than new calls due to the implementation of channel reservation. Using Markov modelling the authors show how new call blocking and hand-off call blocking are affected by the re-trial phenomena.

3.9.4 Evaluation of call completion Probability by Fang, Chlamtac and Lin

The work by Fang, Chlamtac and Lin [23], considers call performance for a PCS network. Under the assumptions of Poisson call arrivals and very general cell dwell time distributions and very general unencumbered call holding time distributions, these authors derive expressions for the call completion probability (ie. 1 – Forced Termination probability). They distinguish between a call forced to terminate (an incomplete call) and a call that is not forced to terminate (a completed call) and determine the effective network holding time pdfs for both complete and incomplete calls under the above assumptions. However, the paper does not examine performance analysis in the classical sense as it assumes that a performance analysis model already exists and that it can be used to provide the inputs for the analysis presented in the paper. The authors assume that such a performance analysis model will provide the new call blocking values and hand-off call blocking values for use in their analysis. The analysis presented by the authors may be used in establishing the appropriate criteria for user-billing in cellular networks. Work of a similar vein can also found in other work by the same authors [53,54].

3.9.5 Performance analysis algorithm for call re-establishment

The paper by Lin, Lin and Jeng [55] consider the performance analysis of Call Re-establishment. Call interruption is the phenomenon whereby an active radio link between a mobile subscriber and the Base station is interrupted due to temporary propagation losses for example, due to structures such as bridges and tunnels. GSM has a Call Re-establishment process, which attempts to avoid the forced termination of calls due to call interruption. In this mechanism, if a communication channel is interrupted, the network reserves the channel for the interrupted call and sets off an interruption timer. If the timer expires or the remote party hangs up the phone before the interruption period is over, the interrupted call is forced to terminate. Otherwise, the interrupted call is resumed by the call re-establishment mechanism. Under the assumption of Poisson new and hand-off call arrivals as well as negative exponential conversation and interruption periods, the authors present analytic models to evaluate the performance of a GSM cellular system with the call re-establishment mechanism.

3.9.6 Performance analysis algorithm for Channel splitting

Li and Alfa [56] examine the phenomenon of Channel Splitting. Under high traffic conditions and good propagation conditions, the different cellular network standards such as GSM and PACS allow for a full-rate traffic channel to be split into two possibly equal half-rate channels. This, in essence, can double the capacity of a cell. Li and Alfa analyse a network where an original l kb/s channel may be split into two channels of l_1 and l_2 kb/s. They also allow for hand-off calls to be queued in a finite buffer if there are no resources available even after channel splitting. They model the call arrival process as general Markov Arrival Processes [57] and employ negative exponentially distributed unencumbered call holding times and cell dwell times. Under these assumptions, the authors derive expressions for the forced termination probability of a hand-off call and the queuing time of a hand-off call in the buffer. They also obtain expressions for some interesting distributions of the network, such as, the busy period time in a cell and the first time to split a certain channel.

3.10 Summary

In this chapter, we examined the various performance analysis algorithms proposed in the literature. These performance analysis algorithms generally differ from one another in the manner they treat the various arrival and service processes relevant to cellular networks. It is clear that the simple negative exponential distribution and mathematical combinations of negative exponential distributions are the most favoured distributions to model call arrival and call service processes. None of the performance analysis algorithms that we considered in this chapter make use of arrival and service processes obtained directly from simulation studies or field measurement studies. The reason is that such distributions do not lend themselves to elegant and tractable solutions.

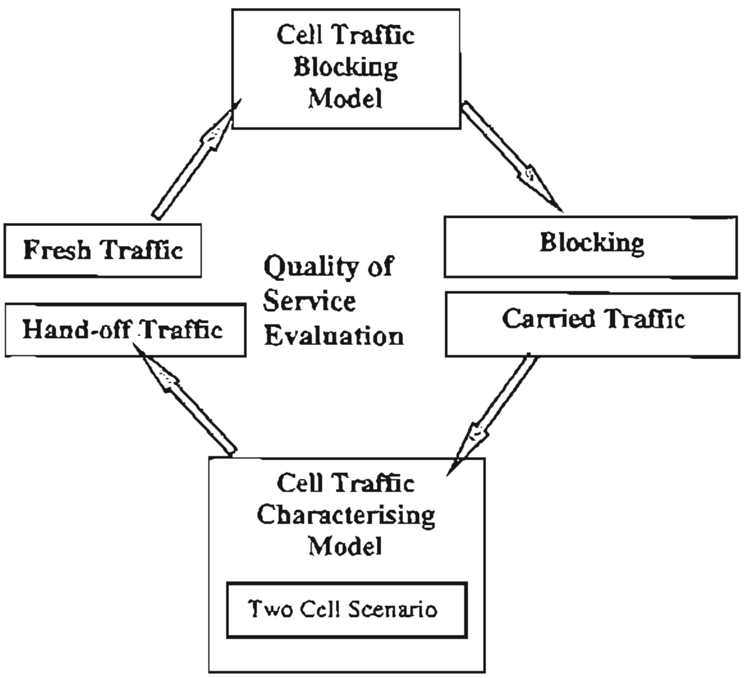
Of all the methods discussed in this chapter, the work by Professor Rappaport and co-authors easily stand out as the most comprehensive ones for use in the performance analysis of cellular networks. However, these models suffer from some limitations as well. The main limitations are that of state space explosion and the assumption of Poisson-type hand-off processes. In subsequent chapters, we propose a different approach, which we think, addresses the limitations that we have discussed so far. We take our cue from the moment-based approaches that are quite popular in the analysis of fixed or wireline networks. These moment-based approaches have to be extended to accommodate the unique characteristic of cellular networks, in that the subscribers have mobility. To ensure that the moment-based approach does not suffer from tractability problems, we follow the approach in fixed network analysis and represent the relevant traffic processes by only the first two moments, the mean and the variance. In this manner we avoid the state space explosion that is peculiar to most of Professor Rappaport and co-author's models. By characterising the traffic moments using their mean and variance, we are able to address the other limitation in Professor Rappaport et al.'s various models in that some of the higher moments of the hand-off traffic process are modelled. As we shall show in the subsequent chapters, this leads to more accurate results than existing methods when compared to simulation.

CHAPTER 4

4.1 Introduction

In the next three chapters we consider our proposed performance analysis algorithm for cellular networks. Our approach is a moment-based one that avoids the state space explosion of existing methods but still allows the various traffic processes to be modelled beyond the mean. For simplification, we loosely decompose the performance analysis of a cellular network into three parts, a generic *cell traffic characterising model*, a generic *cell traffic blocking model* and a *quality of service*

Figure 4-1: Performance Analysis Algorithm



evaluation model. See figure 4-1. In the cell traffic characterising model we determine the mean and variance of traffic offered to each cell. In the cell traffic blocking model we determine the blocking experienced by the various traffic streams offered to each cell. The quality of service evaluation part is essentially a fixed point iteration of the cell traffic characterising and cell traffic blocking parts, where, the outputs of the one are fed in as the inputs of the other and vice versa. As can be seen in figure 4-1, at the heart of our cell traffic characterising model, there lies the stochastic theory analysis of a simple *two-cell scenario*. The two-cell scenario is fundamental to our performance analysis algorithm and is used as "building blocks" to analyse large-scale networks. Typically, we determine the mean and variance of hand-off traffic offered by a cell to its neighbour in our two-cell scenario, and then extend the results to apply for multi-cellular networks. In this chapter, we present results for the two-cell scenario. In subsequent chapters, we show the application of the two-cell scenario in the performance analysis of large-scale networks.

4.2 Assumptions in our model

- Similar to existing studies [7,9,15,27,28,29,30], we assume a homogeneous network where each cell is identical to its neighbours in every aspect such as traffic statistics, user mobility, cell coverage etc.
- Each cell is assumed to have N neighbours. Given that a call is to be handed off, it has a random probability $R_j = 1/N$ of being handed-off to neighbour j.
- We assume that each cell has C channels under a fixed channel allocation (FCA) method. We define a channel to be a single call carrying resource be it a fixed frequency slot (FDMA), time slot (TDMA) or a spread spectrum code (CDMA).
- We assume that *channel reservation* of level r is implemented in each cell. It is considered better to block a mobile user at call initiation rather than grant him service and subsequently drop his call during hand-off when he moves between different cells. Under channel reservation for hand-off calls, a few channels are reserved for the exclusive use of hand-off calls.
- Similar to existing studies [7,9,15,27,29], we assume that new calls arrive in each cell with a Poisson arrival rate of λ_n. The unencumbered call holding time, T_M, is negative exponentially distributed with mean $1/\mu_M$ (see equations 2.3–2.4). The

unencumbered call holding is the time interval for which a mobile user, if unhindered, would employ the services of the network.

- The cell dwell time is the time interval for which the mobile user is within communications distance of a base station. We define two random variables, T_n and T_h, which describe the cell dwell times of new and previously handed-off calls in a cell. We then assume that these random variables have probability density functions (pdfs), $f_n(t)$ and $f_h(t)$ respectively. As shown in chapter 2, these pdfs vary from mobility model to mobility model.
- We define P_N and P_H to be the probabilities with which a new call and a previously handed-off call would require a new hand-off. Therefore,

$$\begin{aligned} P_N &= Prob[T_M > T_n] = \int_0^\infty [1 - F_M(t)]\, f_n(t)\ dt \\ P_H &= Prob[T_M > T_h] = \int_0^\infty [1 - F_M(t)]\, f_h(t)\ dt \end{aligned} \tag{4.1}$$

where $F_M(t)$ is the cumulative distribution function (cdf) of T_M.

- We define the new call blocking in a cell to be p_{nb}. We assume that all hand-off calls offered to a cell suffer the same hand-off call blocking p_{hb}. This implies that all hand-off calls offered to a particular cell i, irrespective of the nature of their origination or the path they undertook to reach cell i, experience the same hand-off call blocking p_{hb} in cell i. This assumption is not quite straightforward since the arrival statistics of hand-off traffic generated by new calls in one of the immediate neighbours of cell i could be different from the arrival statistics of hand-off traffic generated by new calls in cells far afield that had to undergo successive hand-offs to reach cell i. However we employ the assumption that all the different types of hand-off traffic experience the same blocking since it simplifies network analysis. Furthermore, the effect of this assumption can be easily gauged from comparison with our simulator, which is free of such restrictions.
- We define the *total* mean hand-off call arrival rate into a cell to be λ_h. This parameter is obviously network dependent and we shall show later how to calculate it. Also note that we do NOT assume the hand-off process to be Poisson.
- We define Y_c to be the ratio of the mean carried hand-off attempt rate to the mean carried new call attempt rate in a cell:

$$\Upsilon_c = \frac{\lambda_h(1 - p_{hb})}{\lambda_n(1 - p_{nb})} \tag{4.2}$$

- As shown in [15], the cdf, $F_H(t)$, of the channel holding time T_H may be written as:

$$F_H(t) = F_M(t) + \frac{1}{1 + \Upsilon_C}[1 - F_M(t)][F_n(t) + \Upsilon_C F_h(t)] \tag{4.3}$$

- In chapter 7, we prove that for a negative exponential pdf for the unencumbered call holding time and arbitrary cell dwell time cdfs, $F_n(t)$ and $F_h(t)$, the mean channel holding time $E[T_H]$ is given by the following simple expression:

$$E[T_H] = \frac{1}{\mu_M}\left(1 - p_N \frac{1}{1+\Upsilon_c} - p_H \frac{\Upsilon_c}{1+\Upsilon_c}\right) \tag{4.4}$$

4.3 Candidate Channel Holding Time Distributions

The channel holding time distribution is of great importance in cellular network analysis. It describes the distribution of the time spent by a mobile subscriber making use of the resources (channels) within a cell. The channel holding time distribution plays an important role in determining the distribution of the offered traffic in a cell. The channel holding time distribution is determined by various network and subscriber attributes such as unencumbered call holding time, mobile velocity, mobile's direction of motion, changes in motion, geometry of the cell etc.

The stochastic theory techniques of the present day do not lend themselves to easy and tractable models when modelling the channel holding time exactly. Consequently, candidate distributions are generally proposed and used in their place. The parameters of the candidate distribution are adjusted to mimic certain characteristics of the actual channel holding time distribution and it is the candidate distribution that is then used in the tele-traffic analysis of cellular networks.

4.3.1 Negative Exponential Distribution

The most famous and most common candidate distribution is the simple negative exponential distribution. It has the following pdf:

$$f(t) = \mu e^{-\mu t} \tag{4.5}$$

In the negative exponential distribution there is only one parameter that may be adjusted to fit an actual channel holding time distribution, namely, the mean parameter $E[T] = 1/\mu$. This restricts the level of approximation that the negative exponential distribution may offer when modelling an actual channel holding time distribution. However, the usefulness of the negative exponential distribution may not be under stated because of the ease with which it lends itself to modelling. This attribute, which we loosely term as "ease of application", is in fact a good measure as to whether other candidate distributions will enjoy some of the popularity of the negative exponential distribution. In this thesis, we put forward two candidate distributions that we feel satisfy this characteristic and are, in their own right, suitable models for the actual channel holding time distribution. They are the det-neg distribution and the gamma distribution.

4.3.2 Det-neg distribution

We define the pdf of the det-neg distribution as follows:

$$f_{DT-NG}(t) = p\delta\left(t - \frac{1}{\mu}\right) + (1-p)\mu e^{-\mu t} \tag{4.6}$$

where $\delta(.)$ is the Dirac-delta function. The parameter p is a probability parameter where $0 \leq p \leq 1$. The parameter μ is a service rate and has units of departures per unit time. The det-neg distribution is a linear composition of two common distributions, namely, the deterministic distribution with probability p and the negative exponential distribution with probability $1-p$. The mean $E[T_{DT-NG}]$ and variance $V[T_{DT-NG}]$ of the det-neg distribution are easy to obtain and are:

$$E[T_{DT-NG}] = \frac{1}{\mu}\ ;\ V[T_{DT-NG}] = \frac{(1-p)}{\mu^2} \tag{4.7}$$

4.3.3 Gamma distribution

The pdf of the gamma distribution [41] is as follows:

$$f_{GAM}(t) = \frac{(\mu ct)^{c-1}\mu c e^{-\mu ct}}{\Gamma(c)} \tag{4.8}$$

where $\Gamma(c)$ is the gamma function with parameter c and is given by:

$$\Gamma(c)=\int_0^{\infty} x^{c-1}e^{-x}\,dx \tag{4.9}$$

The parameter c is often termed the shape parameter and the parameter μ is a service rate and has units of departures per unit time. The mean $E[T_{GAM}]$ and variance $V[T_{GAM}]$ of the gamma distribution are:

$$E[T_{GAM}]=\frac{1}{\mu}\ ;\ V[T_{GAM}]=\frac{1}{\mu^2 c} \tag{4.10}$$

4.3.4 Justification for the choice of candidate channel holding time distributions

There lie many advantages in the use of the det-neg or gamma distributions to model the actual channel holding time distributions found in cellular networks. First and foremost, the det-neg and gamma distributions are two-parameter distributions and therefore offer a second degree of flexibility, when modelling the actual channel holding time distribution, compared to the single parameter negative exponential distribution. Secondly, since the det-neg and gamma distributions are easily reduced to the negative exponential distribution (p=0 or c=1), we can ensure that the det-neg and gamma distributions will be at least as good as, if not better than, the negative exponential distribution when best fitting an actual channel holding time distribution. Furthermore, the det-neg and gamma distributions are simple to use. By manipulating the two parameters, p and μ, the proposed det-neg distribution can be used to fit a large number of common channel holding time distributions – including the negative exponential (p=0) and the deterministic (p=1). Similarly by manipulating the two parameters, c and μ, the gamma distribution can be used to fit a large number of popular channel holding time distributions – including the negative exponential (c=1), the erlang-k (c=k) and the deterministic ($c=\infty$). Also, the det-neg and gamma distributions are simple to use in our moment-based performance analysis algorithm and it is this simplicity that enables us to bypass the state space explosion associated with other more elaborate performance analysis models in the literature [27,28,30]. The gamma distribution has an added advantage over the det-neg distribution in that it may be used to model channel holding time distributions where the square of the mean of the distribution is less than the variance (i.e. c<1). On the other hand, the det-neg distribution is much more suited to environments such as the highway cellular network and train-line networks where the negative exponential distribution is clearly inadequate [27, 58]. Finally, the use of the gamma distribution is nothing new in the performance analysis of cellular networks: Rappaport [27] employed the erlang-k distribution in the performance analysis of cellular networks and the erlang-k is no more than a special case of the gamma distribution.

4.4 Stochastic Theory Analysis Of The Two-Cell Scenario

Consider the scenario shown in figure 4-2a, consisting of two cells in series, the first one having *C* servers and the second having infinite servers. New calls occur in cell L_1 with Poisson arrival rate λ and are served for a random time interval *T* with a general pdf *f(t)* and mean of *E[T]* minutes. We assume that each carried call in the first cell, after completion of service, is either handed-off to the second cell L_2 with a probability Q_h or is terminated within the first cell with probability $1-Q_h$. Calls are serviced in the second cell for time intervals that are random variables with the *same* pdf *f(t)* and mean of *E[T]* minutes. Note that for simplicity we assume that the second cell receives no new calls and also generates no hand-off calls to be given to the first cell. We are interested in the mean and variance of hand-off traffic offered by the first cell to the second cell. The distribution of carried traffic in a loss server system is the distribution of the number of occupied servers in that system for a given arrival and service process. For mathematical analysis, it is convenient to define what is known as an offered traffic

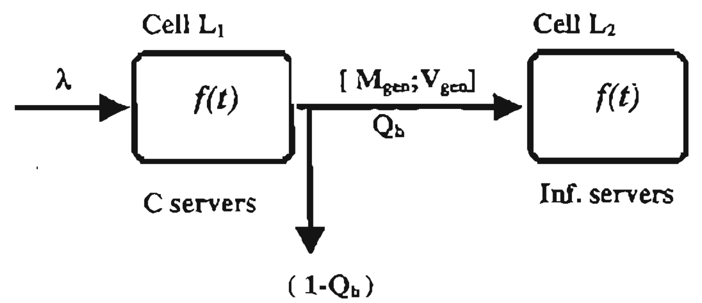


Figure 4-2a: Two-Cell Scenario

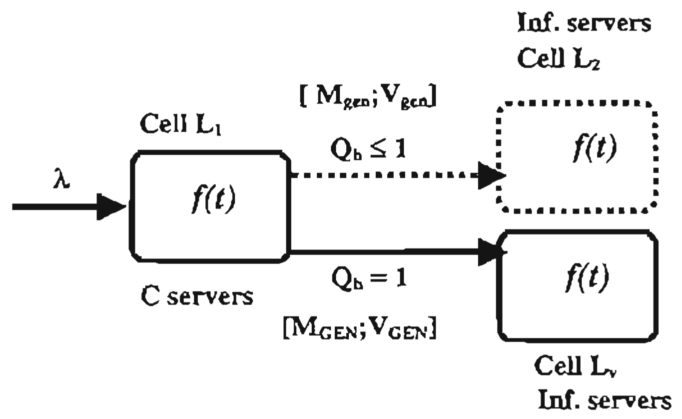


Figure 4-2b: Virtual Cell Scenario

distribution. Under classical definition [14], an offered traffic distribution due to an arrival and service process, is equivalent to the carried traffic distribution in an imaginary infinite (non-blocking) server system where the same arrival and service processes are assumed to occur. Using this definition, we may determine the mean and variance of hand-off offered traffic by calculating the mean and variance of carried traffic in the infinite sized second cell. It is exactly for this reason that we set the channel size in the second cell to be infinite.

The mean offered traffic in the first cell due to the Poisson arrival process is $A = \lambda E[T]$ (Little's result [59]). A useful attribute of Poisson arrival processes in loss systems is that they are insensitive to service time distributions beyond the mean $E[T]$ of the distribution [59]. What we mean by this is that, given the same Poisson arrival rate, the M/M/C and M/G/C loss systems have the same solution provided the mean service times are the same. Therefore, the standard results for the M/M/C loss systems apply for the first cell under arbitrary channel holding time distributions. For ease of representation we define the mean service rate $\mu = 1 / E[T]$. The state probability p_j of finding the first cell in state j is [14,59]:

$$p_j = \frac{\frac{A^j}{j!}}{\sum_{k=0}^{C} \frac{A^k}{k!}} \tag{4.11}$$

Time congestion E is the fraction of time for which the first cell is in state C and is given by the Erlang-B formula:

$$E = p_C = \frac{\frac{A^C}{C!}}{\sum_{k=0}^{C} \frac{A^k}{k!}} \tag{4.12}$$

The mean carried traffic M_C in the first cell is:

$$M_C = \sum_{j=0}^{C} j p_j = A[1 - E] \tag{4.13}$$

The variance of carried traffic V_C in the first cell is:

$$V_C = \sum_{j=0}^{C} j^2 p_j - M_C^2 = M_C - (C - M_C)AE \tag{4.14}$$

It is a mathematically open question whether there exists explicit solution for the moments of hand-off traffic offered by the first cell to the second cell for arbitrary channel holding time distributions *f(t)*. Let M_{gen} & V_{gen} be the mean and variance of hand-off traffic offered by the first cell to the second cell, for hand-off probability $Q_h \leq 1$, under the Poisson new call arrival process in the first cell, and channel holding time pdfs *f(t)* that are generally (arbitrarily) distributed. We now employ an abstraction called the *Virtual Arrival Process* to solve the above system. As shown in figure 4-2b we define a virtual cell L_v of infinite servers in parallel to cell L_2. The virtual arrival process, which is assumed to occur in the virtual cell L_v, is purely a mathematical construction whereby calls from cell L_1 are handed off to cell L_v with *hand-off probability* $Q_h=1$. When hand-off probability $Q_h=1$, every call carried in the first cell is offered to the virtual cell. These calls are then assumed to be serviced in the virtual cell for time intervals that are random variables with pdf *f(t)* and mean of *E[T]* minutes. We define the *Virtual Arrival Traffic (VAT)* to be the traffic generated by this virtual arrival process under the generalised service times in the virtual cell. We determine the mean M_{GEN} and variance V_{GEN} of this virtual arrival traffic and then use them to determine the mean M_{gen} and variance V_{gen} of hand-off traffic offered by the first cell to the second cell for $Q_h \leq 1$. Note the difference in the usage of the lowercase and uppercase subscripts. Consider a single call in progress in the virtual cell: under the simplistic assumption of a fixed (or time-invariant) hand-off probability Q_h, the call that is in progress in the virtual cell L_v, may be considered to be involved in a simple Bernoulli trial, whereby, with a probability of success equal to Q_h it will also be present in the second cell L_2. We assume that such hand-off calls have the same service time in both cells L_2 and L_v.

For derivation purposes, we now assume that we know the stationary state probability *p(m)* of an external observer finding *m* calls in the virtual arrival traffic in the virtual cell. The conditional probability, *p(k|m)*, of having exactly *k* hand-off calls in cell L_2 out of a total of *m* calls in the virtual cell is binomially distributed:

$$p(k \mid m) = \frac{m!}{k!(m-k)!} Q_h^{\,k} (1 - Q_h)^{m-k} \qquad (4.15)$$

The unconditional probability *p(k)* of having *k* hand-off calls is the following:

$$p(k) = \sum_{m=k}^{\infty} \frac{m!}{k!\,(m-k)!} Q_h^k (1 - Q_h)^{m-k} p(m) \tag{4.16}$$

The j[th] moment of the hand-off offered traffic stream G_j from cell L_1 to cell L_2 may be determined to be:

$$G_j = \sum_{k=0}^{\infty} k^j\, p(k) = \sum_{k=0}^{\infty} k^j \sum_{m=k}^{\infty} \frac{m!}{k!\,(m-k)!} Q_h^k (1 - Q_h)^{m-k} p(m) \tag{4.17}$$

$$G_j = \sum_{m=0}^{\infty} \sum_{k=0}^{m} k^j \frac{m!}{k!\,(m-k)!} Q_h^k (1 - Q_h)^{m-k} p(m) \tag{4.18}$$

Thereafter, it can be easily shown that the mean M_{gen} and variance V_{gen} of hand-off traffic offered by the first cell to the second cell is related to the mean M_{GEN} and variance V_{GEN} of virtual arrival traffic offered by the first cell to the virtual cell as follows:

$$M_{gen} = G_1 = M_{GEN}\, Q_h \qquad for\ Q_h \le 1 \tag{4.19}$$

$$V_{gen} = G_2 - G_1^2 = M_{gen} \left(1 + Q_h \cdot \left(\frac{V_{GEN}}{M_{GEN}} - 1 \right) \right) \quad for\ Q_h \le 1 \tag{4.20}$$

The usefulness of the above equations is that, provided we find the mean M_{GEN} and variance V_{GEN} of the virtual arrival traffic offered by the first cell to the virtual cell for hand-off probability Q_h=1 for a particular channel holding time distribution, we may then apply the above equations to determine the mean M_{gen} and variance V_{gen} of hand-off traffic offered by the first cell to the second cell for any other hand-off probability $Q_h \le 1$ under the same channel holding time distribution [60,61,62].

The derivation of equation (4.19) and (4.20) relies on the assumption that the hand-off probability Q_h is time-invariant. When the cell dwell times and the unencumbered call holding times are both negative exponentially distributed, the hand-off probability is fixed or time-invariant. Put more formally, the age of the interval since the call arrival does not affect the hand-off probability. Therefore the above equations are strictly correct for the case where both the dwell times and unencumbered call holding times are negative exponential. In our assumption section (section 4.2), we have considered the unencumbered call holding time to be negative exponential but the cell dwell times to be arbitrarily distributed. In such a scenario, the age of the interval since the call arrival affects the hand-off probability and therefore the hand-off probability is not fixed or time-invariant.

However, to simplify analysis, we assume that hand-off probability is fixed or time-invariant. Although strictly untrue we feel that the assumption of a fixed hand-off probability is a reasonable assumption regarding the nature of hand-off traffic and one that allows us to derive a performance analysis algorithm that is quite accurate. In comparison, the more popular assumption in literature, that hand-off traffic is Poisson, is a much poorer assumption and one that provides more conservative results as we shall show in chapter 6.

The mean offered traffic M_{GEN} of the virtual arrival traffic is the same as the mean carried traffic M_C in the first cell because every carried call in the first cell is offered to the virtual cell when hand-off probability Q_h is equal to 1. Hence equation (4.19) reduces to:

$$M_{gen} = M_C Q_h = A(1-E)Q_h \quad for\ Q_h \leq 1 \tag{4.21}$$

The variance V_{GEN} of the virtual arrival traffic offered by the first cell to the virtual cell, when Q_h=1, is a little more difficult to obtain for arbitrary channel holding time distributions. However, it is possible to solve for it for specific channel holding time distributions. We now present exact results for V_{NEG} and V_{DET}, which are the variances of the virtual arrival traffic obtained when $f(t)$ has a negative exponential and deterministic distribution respectively. We also analyse the simple two-cell scenario under the det-neg and gamma distributions. The exact theoretical analysis of the two-cell scenario, under det-neg or gamma channel holding times, is far too cumbersome to be practical from an industrial usage point of view. In the analysis that we present in subsequent sections, we set our primary goal to be one of engineering ease-of-application. Therefore, we make necessary approximations, wherever applicable, to ensure that the results we present for both the det-neg and gamma distributions are simple to apply and easy to implement. However, as we shall show in the results section, we have not sacrificed accuracy in any manner.

4.4.1 The Virtual Arrival Traffic Under Negative Exponential Distributions [60,61,62]

We define a joint probability distribution $p_{n,m}$ of having n busy channels in cell L_1 and m busy channels in the virtual cell of infinite size. The Markov state diagram for the virtual call arrival process is shown in figure 4-3. Note that we have shown only the transitional rates to and from state (n,m) in figure 4-3. The Birth-Death equations for the above Markov state diagram may be derived under the assumption that the state transitional rates to and from a particular state (n,m) are equal: (NB. $0 \leq n \leq C$, $0 \leq m < \infty$)

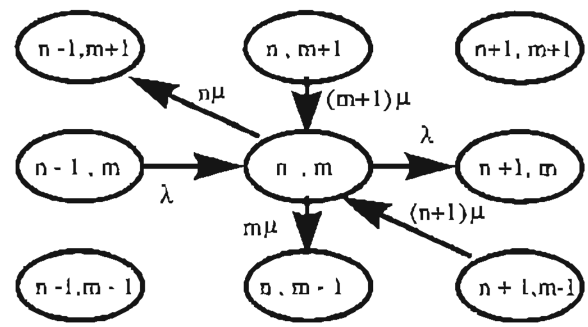


Fig 4-3: Exact Birth-Death Model for the virtual arrival process

$$\text{for } n = 0 ; (\lambda + m\mu) p_{0,m} = (m+1)\mu \ p_{0,m+1} + \mu \ p_{1,m-1} \tag{4.22}$$

$$\text{for } 0 < n < C; (\lambda + m\mu + n\mu) p_{n,m} = (m+1)\mu \ p_{n,m+1} + \lambda \ p_{n-1,m} + (n+1)\mu \ p_{n+1,m-1} \tag{4.23}$$

$$\text{for } n = C \ ; (m\mu + C\mu) p_{C,m} = (m+1)\mu \ p_{C,m+1} + \lambda \ p_{C-1,m} \tag{4.24}$$

Note that in the above equations we assume that $p_{n,m}$ equals zero for all $n<0$ and/or $m<0$. Using the probability generating function $G_n(z)$, where,

$$G_n(z) = \sum_{m=0}^{\infty} p_{n,m} \ z^m \tag{4.25}$$

we may convert equation (4.22) to the following form:

$$\mu(z-1) G'_0(z) + \lambda \ G_0(z) = \mu \ z G_1(z) \tag{4.26}$$

using a change of variable, x, where x = z-1;

$$x\mu \ G'_0(x+1) + \lambda G_0(x+1) = \mu(x+1) G_1(x+1) \tag{4.27}$$

Since the PGF is related to the Binomial Moment Generating function $F_n(x)$ as follows [14, pp. 461]:

$$F_n(x) = G_n(x+1) \tag{4.28}$$

$$\text{where } F_n(x) = \sum_{j=0}^{\infty} \beta_{n,j} \ x^j \tag{4.29}$$

and $\beta_{n,j}$ is the j^{th} Binomial moment of the virtual arrival traffic when n calls are in the originating cell L_1. Equation (4.27) becomes:

$$\mu x \ F'_0(x) + \lambda \ F_0(x) = \mu(x+1) F_1(x) \tag{4.30}$$

Similar manipulations of equations (4.23) & (4.24) give equations (4.31) and (4.32) respectively:

$$\mu x F'_n(x) + (\lambda + n\mu) F_n(x) = (n+1)\mu(x+1) F_{n+1}(x) + \lambda F_{n-1}(x) \quad (4.31)$$

$$\mu x F'_C(x) + C\mu F_C(x) = \lambda F_{C-1}(x) \quad (4.32)$$

Substituting (4.29) into equations (4.30-4.32) and equating coefficients of equal powers of x give different equations relating the different Binomial moments. Equating coefficients of equal powers of x, where x is raised to 1, give equations (4.33-4.35).

$$for \ i=0: \quad (\lambda+\mu)\beta_{0,1} = \mu\beta_{1,1} + \mu p_1 \quad (4.33)$$

$$for \ 0<i<C: \quad (\lambda+\mu+i\mu)\beta_{i,1} = (i+1)(\mu\beta_{i+1,1} + \mu p_{i+1}) + \lambda \beta_{i-1,1} \quad (4.34)$$

$$for \ i=C: \quad (\mu+C\mu)\beta_{C,1} = \lambda \ \beta_{C-1,1} \quad (4.35)$$

where $\beta_{i,1}$ is the first Binomial moment of the virtual arrival traffic when i customers are in cell L_1. The probabilities p_i are the state probabilities in the Erlang-B distribution (equation (4.11)). Adding these equations together gives the expression for the mean of the virtual arrival traffic:

$$M_{GEN} = \frac{\lambda}{\mu} . (1 - E) \quad (4.36)$$

Equating coefficients of equal powers of x for all x raised to 2, and adding the C+1 equations gives the relationship between the second complete Binomial moment β_2 and the partial first Binomial moments $\beta_{i,1}$ as shown below:

$$2\beta_2 = \sum_{i=0}^{C} i\beta_{i,1} \quad (4.37)$$

Thereafter the variance V_{NEG} is:

$$V_{NEG} = 2\beta_2 - M_{NEG}^2 + M_{NEG} \quad (4.38)$$

Note that the partial first Binomial moments $\beta_{i,1}$ are the solutions of the (C+1) simultaneous equations (4.33-4.35). The solution of these simultaneous equations is quite simple since they form a matrix that only has non-zero values in the tri-diagonals. The number of row substitutions required in a Gaussian elimination method for solving the above system is exactly 2xC.

4.4.2 The Virtual Arrival Traffic Under Deterministic Distributions [58]

Now consider the case where the channel holding times in both cells are deterministic or fixed at a value of $E[T]$ minutes. We define M_{DET} and V_{DET} as the mean and variance of the virtual arrival traffic offered by the first cell to the virtual cell when probability Q_h=1.

For the purpose of illustration, we assume that a call arrives to find a free channel available in the first cell and is, consequently, accepted into the first cell at time S. Due to the deterministic channel holding time, this call will then terminate its service in the first cell at time $S+ E[T]$ minutes and thereupon be offered to and accepted in the virtual cell of infinite channels. Due to the deterministic channel holding time in the virtual cell, the call will then terminate its service in the virtual cell at exactly $S+2\ E[T]$ minutes. This means that every call that is accepted in the first cell will be accepted in the infinite sized virtual cell exactly $E[T]$ minutes later and every call that departs the first cell will subsequently depart the virtual cell $E[T]$ minutes later. Since every event, be it an arrival or a departure, that occurs in the first cell will occur in the virtual cell $E[T]$ minutes later, the number of calls in progress in the first cell at an arbitrary time S would be the same as the number of calls that will be in progress in the virtual cell at time $S+ E[T]$. The state probability distribution as seen by an external observer at arbitrary time S, in the first cell is simply p_i (defined in equation 4.11). Consequently, this is the same state probability distribution that will be seen by the external observer in the virtual cell at time $S+ E[T]$. Since $E[T]$ minutes after an arbitrary time S is in itself an arbitrary time, the state probability distribution that will be observed by the external observer in the virtual cell at arbitrary time is also p_i. By classical definition [14], the offered traffic distribution for the virtual arrival traffic offered from the first cell to the virtual cell is simply the carried traffic distribution in the infinite sized virtual cell. Therefore, the mean M_{DET} and variance V_{DET} of virtual arrival traffic offered by the first cell to the virtual cell, for Q_h=1, is the same as the mean and variance of carried traffic in the virtual cell (which is the same as the mean M_C and variance V_C of carried traffic in the first cell):

$$M_{DET} = \sum_{i=0}^{C} i\,p_i = M_C = A(1-E); \qquad V_{DET} = \left(\sum_{i=0}^{C} i^2 p_i\right) - M_{DET}^2 = V_C = M_C - AE(C-M_C) \qquad (4.39)$$

Interestingly, the virtual cell, although infinite in size, will never carry more than C calls at any one time under deterministic holding times, because the first cell (which is finite) is not able to carry more than C calls. In addition to the above equations for the mean and the variance, it can be shown that all the moments of the virtual arrival traffic are actually the same as the corresponding moments of carried traffic in the first cell under deterministic channel holding times and hand-off probability $Q_h = 1$.

4.4.3 The Virtual Arrival Traffic Under Arbitrary Channel Holding Times [63]

The solution for the variance of the virtual arrival traffic V_{GEN} offered by the first cell to the virtual cell, under arbitrary channel holding times, is quite difficult. The arrival process in the first cell is renewal, and more specifically, Poisson. However, because of the effects of blocking, the call departure process from the first cell is non-Poisson and also non-renewal. The virtual cell scenario that we employ is in fact *two servers in series*, where, the departure process of the first server is fed in as the arrival process of the second server. As such, the two-cell scenario constitutes a generalised form of what is known as Jackson networks [64]. As shown in [64], Jackson networks have been solved under simple assumptions such as Poisson arrivals, negative exponential service times and infinite server or infinite waiting room (i.e. cases of no blocking). The above reference discusses other special cases where Jackson type networks have been solved. However, the difficulty with our virtual cell scenario is that there exists blocking in the finite server first cell and therefore the general results of Jackson networks, such as product form solutions, do not apply. The solution of the queuing process in the virtual cell where the arrival process is non-renewal and the service times are arbitrarily distributed is very cumbersome. In fact, it is unsure whether tractable and explicit results are possible. In the spirit of our primary goal of ensuring simple and easy to implement results, we abandon exact analysis and present approximate methods for simplifying derivation.

4.4.4 Approximation For The Variance Of The Virtual Arrival Traffic Using Asymptotic Analysis

An accurate expression for the variance V_{GEN} of the virtual arrival traffic may be obtained via asymptotic analysis. In our asymptotic analysis we consider the behaviour of our virtual cell scenario of figure 4-2b when the arrival rate λ in the first cell increases without bound. Essentially, we solve for the variance of the virtual arrival traffic in the asymptotic case when the Poisson arrival rate $\lambda \to \infty$ and then borrow the structure of the results and apply them for finite values of λ. We show in the result section that such a methodology produces very accurate expressions for the variance of the virtual arrival traffic under finite values of λ.

Theoretically it is possible to allow the arrival rate λ to tend to infinity even though practically it might be difficult to conceive a case where an arrival process offers infinite calls per unit time. To distinguish between the normal case where the Poisson parameter λ is finite and the asymptotic case where the Poisson parameter $\lambda \to \infty$, we use the infinity symbol in the superscript to characterise the various parameters of the asymptotic scenario. Now consider figure 4-2b, when the limit $\lambda \to \infty$ is taken and the service rate μ and the channel size C are constant values. The limit $A \to \infty$ occurs as a result of the limit $\lambda \to \infty$, since the offered traffic $A = \lambda/\mu$ and the channel holding time $1/\mu$ is constant. Under the limit $A \to \infty$, the time congestion in the first cell E^{∞} (equation 4.12) becomes the following:

$$E^{\infty} = \lim_{A \to \infty} p_C = \lim_{A \to \infty} \frac{\frac{A^C}{C!}}{\sum_{k=0}^{C} \frac{A^k}{k!}} = 1 \tag{4.40}$$

The mean carried traffic M^{∞}_C in the first cell (equation 4.13) under the limit $A \to \infty$ is:

$$M^{\infty}_C = \lim_{A \to \infty} \sum_{j=1}^{C} j p_j = C \tag{4.41}$$

Similarly, the variance of carried traffic V^{∞}_C in the first cell (equation 4.14) in the limit $A \to \infty$ becomes:

$$V^{\infty}_C = \lim_{A \to \infty} [\sum_{j=1}^{C} j^2 p_j] - [M^{\infty}_C]^2 = C^2 - C^2 = 0 \tag{4.42}$$

The limits $M^{\infty}_C = C$ and $V^{\infty}_C = 0$ have a nice physical interpretation in that, as $\lambda \to \infty$, calls arrive at an infinite rate, and as each call finishes its service in the first cell, there is always a new arrival (because of the infinite arrival rate) to utilise the freed up channel in the first cell. This leads to all *C* channels being occupied all the time. This results in a mean carried traffic $M^{\infty}_C = C$ Erlangs. Since all *C* channel are occupied all the time, there is also little variation in the number of occupied channels from the mean value of $M^{\infty}_C = C$ and therefore the variance of carried traffic in the first cell $V^{\infty}_C = 0$.

4.4.4.1 The Effect of an Infinite Arrival Rate on the Hunting Process in the First Cell

Hunting is the name given to the process of selecting a free channel from the pool of free channels when faced with a new call arrival. For example, sequential hunting is the name given to the process where each channel in the pool is sampled sequentially until one that is free is found. Under finite arrival rates in the first cell, the consequence of the hunting process is that the order in which calls are accepted, in the C different channels of the first cell, are *inter-dependent.* Put more formally, the distribution of the time interval between the acceptance of two calls in a particular channel is dependent on the nature of the hunting process and the state of occupancy – whether free or not – of the remaining C-1 channels. However, under an infinite arrival rate in the first cell, each time a call finishes its service in a particular channel and leaves, there will always be a new arrival to take up occupancy of this channel immediately afterwards. The infinite arrival rate has an important consequence in that the process of calls being accepted in each of the *C* channels in the first cell is now independent of the hunting process and the occupancy status of the remaining *C-1* channels in the first cell. Furthermore, by definition, the service process in each of the *C* channels in the first cell is independent of the service process in the remaining *C-1* channels. Therefore, the infinite arrival rate makes the stochastic process in progress in any single channel in the first cell become independent of the stochastic processes in progress in the remaining *C-1* channels of the first cell. We may therefore consider a single channel in the first cell in isolation.

Figure 4-4 illustrates the arrival and service processes in a single channel of the first cell as well as the corresponding arrival and service processes that occur in the infinite sized virtual cell as a result of departures from our single channel. The random variable, Y,

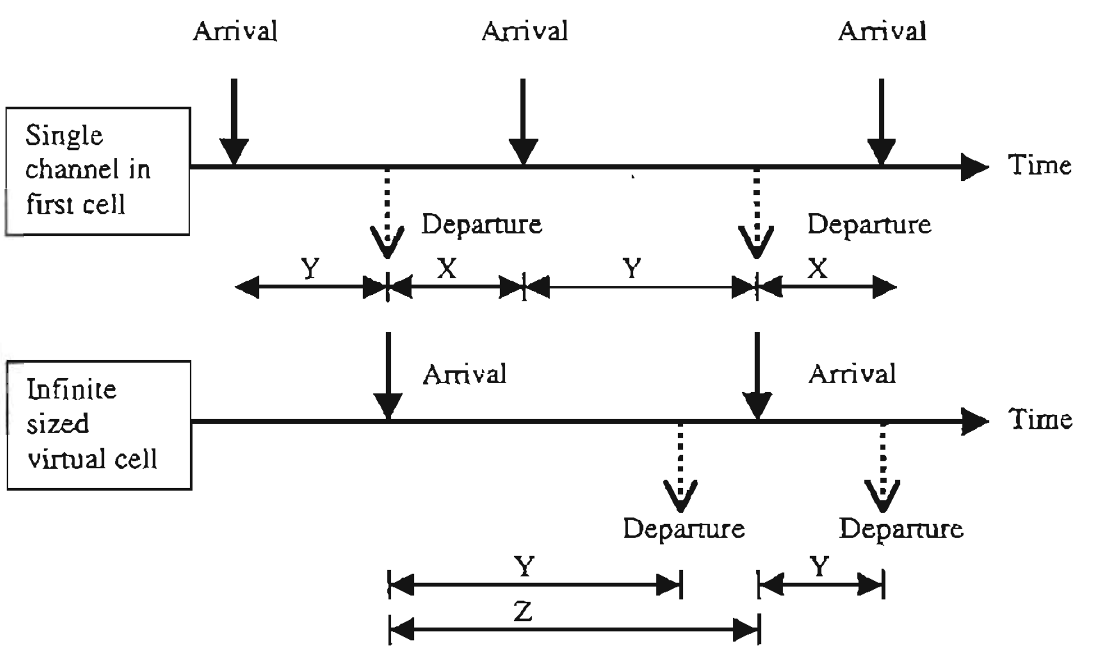


Figure 4-4: The arrival process in the virtual cell due to departures from a single channel in the first cell

describes the service processes in both our single channel and the virtual cell, and has pdf $f(t)$. The random variable, X, represents the time interval between the departure of a call and the acceptance of a subsequent call in our single channel. Under a finite arrival rate, the random variable X is dependent on many factors such as the Poisson arrival process in the first cell, the hunting process in the first cell and the state of occupancy of the C channels in the first cell. The random variable Z describes the interval between two successive call arrivals in the virtual cell as a result of call departures from our single channel and therefore, $Z = X + Y$.

Under an infinite arrival rate, the time interval, X, between a call departure from our single channel and a subsequent arrival in the same channel is infinitesimally small, ie., $X \to 0$. Therefore, the distribution of the interval between successive departures, X+Y, from this channel tends towards the distribution of the service time, Y, in the same channel, ie., X+Y→Y. This means that the inter-arrival distribution of call arrivals in the virtual cell that occur solely as a result of call departures from this particular channel tends towards $f(t)$ and Z→Y. By definition, all calls in the virtual cell have a service distribution with pdf $f(t)$. Therefore, the resulting queue in the virtual cell, as a result of our single channel, is a G/G/∞ queuing system where the inter-arrival and service distribution have identical pdfs $f(t)$. Because each of the C channels in the first cell is independent from the others under an infinite arrival rate, there are in effect C independent G/G/∞ queuing systems in the virtual cell. The virtual arrival traffic in the virtual cell is in fact the cumulative effect of these C independent G/G/∞ queuing systems. Therefore, the mean M^{∞}_{GEN} of the virtual arrival traffic offered by the first cell to the virtual cell, under the infinite arrival rate in the first cell and arbitrary service time pdf $f(t)$, is the same as the mean carried traffic in C independent G/G/∞ queuing systems. Firstly, let us define $M_{G=GEN}$ and $V_{G=GEN}$ as the mean and variance of carried traffic in the above-mentioned G/G/∞ queuing system. Hence,

$$M^{\infty}_{GEN} = C\, M_{G=GEN} = C \tag{4.43}$$

Similarly, the variance V^{∞}_{GEN} of the virtual arrival traffic offered by the first cell to the virtual cell, under the infinite arrival rate in the first cell and arbitrary service time pdf $f(t)$, is the same as the variance of carried traffic in C independent G/G/∞ queuing systems. Hence,

$$V^{\infty}_{GEN} = C\, V_{G=GEN} \tag{4.44}$$

In this manner, we reduce the calculation of the variance of the virtual arrival traffic V^{∞}_{GEN} under the limit $\lambda\rightarrow\infty$, to the calculation of carried traffic in C independent G/G/∞ queuing systems. However, the solution of a G/G/∞ queuing system is no trivial matter. We now consider four special cases: the negative exponential, the deterministic, the det-neg and the gamma distributions.

4.4.4.2 The Limiting Virtual Arrival Traffic Under Negative Exponential Distributions

In figure 4-2b, when the channel holding time distribution $f(t)$ is negative exponential and the limit $\lambda\rightarrow\infty$ is taken, the queuing scenario in the virtual cell reduces to a C number of independent and identical M/M/∞ queuing systems. The M/M/∞ queuing system is a Poisson process and the variance of carried traffic $V_{G=NEG}$ is the same as the mean carried traffic, namely 1 Erlang. Therefore, the variance of the virtual arrival traffic V^{∞}_{NEG} under negative exponential channel holding times is:

$$V^{\infty}_{NEG} = C\, V_{G=NEG} = C \tag{4.45}$$

4.4.4.3 The Limiting Virtual Arrival Traffic Under Deterministic Distributions

In figure 4-2b, when the channel holding time distribution $f(t)$ is deterministic and the limit $\lambda \rightarrow \infty$ is taken, the queuing scenario in the virtual cell reduces to a C number of independent and identical D/D/∞ queuing systems. The D/D/∞ queuing system is one where the inter-arrival and service distribution have the same deterministic pdf and this results in a constant traffic process. Consequently, the variance of carried traffic $V_{G=DET}$ is 0. Therefore, the variance of the virtual arrival traffic V^{∞}_{DET} under deterministic channel holding times is:

$$V^{\infty}_{DET} = C\, V_{G=DET} = 0 \tag{4.46}$$

4.4.4.4 The Limiting Virtual Arrival Traffic Under Det-Neg Distributions [65]

In figure 4-2b, when the channel holding time distribution $f(t)$ is det-neg and the limit $\lambda \rightarrow \infty$ is taken, the queuing scenario in the virtual cell reduces to a C number of independent and identical DN/DN/∞ queuing systems. The DN/DN/∞ queuing system is one where the inter-arrival and service distributions have the same det-neg pdf. The solution of the DN/DN/∞ queuing system is cumbersome even with the added simplification that the arrival process is renewal and that the inter-arrival and service distributions are identical det-neg distributions. In line with our initially stated objective of ensuring ease-of-application and tractability of our results we bypass exact analysis using the following simple approximation.

Firstly we know the variance of carried traffic $V_{G=DT\text{-}NG}$ in the DN/DN/∞ queuing system at the two endmost points. The variance $V_{G=DT\text{-}NG} = 1$ for det-neg probability parameter $p=0$ and the variance $V_{G=DT\text{-}NG} = 0$ for det-neg probability parameter $p=1$. Using linear interpolation between these two points we obtain a crude approximation for the variance, essentially, $V_{G=DT\text{-}NG} \approx 1 - p$ for $0 \leq p \leq 1$.

As it turns out, a remarkably accurate expression for the variance $V_{G=DT\text{-}NG}$ may be obtained using a quadratic approximation. A quadratic approximation requires three points and we only have two. The simplest way to obtain a third point is via simulation. We found that variance $V_{G=DT\text{-}NG} \approx 5/8$ for det-neg probability parameter $p=1/2$. The resulting quadratic expression is:

$$V_{G=DT-NG} \approx 1 - \frac{p}{2} - \frac{p^2}{2} \quad (4.47)$$

In the results section we illustrate the remarkable accuracy of this quadratic expression. Using this result in equation 4.44, the variance of the virtual arrival traffic $V^{\infty}_{DT\text{-}NG}$ under det-neg channel holding times and in the limit $\lambda \to \infty$ is:

$$V^{\infty}_{DT-NG} = C\,V_{G=DT-NG} \approx C\left(1 - \frac{p}{2} - \frac{p^2}{2}\right) \quad (4.48)$$

4.4.4.5 The Virtual Arrival Traffic Under Det-Neg Channel Holding Times And Finite Arrival Rates

Let us now consider the variance of the virtual arrival traffic under finite arrival rates in the first cell, channel holding times that are det-neg distributed with mean $1/\mu$ and probability parameter p where $0 \le p \le 1$. For a fixed channel size $C=C_1$ in the first cell and a fixed service rate $\mu = \mu_1$, the variance $V_{DT\text{-}NG}(\lambda, p)$ is a curvilinear surface that is a function of the mean arrival rate λ and the det-neg probability parameter p. We know the two endmost curves that lie on this curvilinear surface.

1. We know that the curve $V_{DT\text{-}NG}(\lambda, 0) = V_{NEG}$ describes the curvilinear surface when the probability parameter $p=0$ and may be obtained using equations 4.33-4.38.
2. We also know that the curve $V_{DT\text{-}NG}(\lambda, 1) = V_{DET}$ describes the curvilinear surface when the probability parameter $p=1$ and is given by equations 4.39.

Interpolation between the curves $V_{DT\text{-}NG}(\lambda, 0) = V_{NEG}$ and the curve $V_{DT\text{-}NG}(\lambda, 1) = V_{DET}$ gives a very simple approximation for the curvilinear surface $V_{DT\text{-}NG}(\lambda, p)$ under arbitrary probability parameter p as shown below: (We employed the same approximation to analyse highway cellular networks in [58]).

$$V_{DT-NG}(\lambda, p) = pV_{DET} + (1-p)V_{NEG} \qquad (4.49)$$

However, we can obtain a further improvement by noting the following:

3. The curve $V_{DT-NG}(\infty, p) = C(1 - p/2 - p^2/2)$ is an accurate approximation for the curvilinear surface when the limit $\lambda \rightarrow \infty$ is taken.

The simplest curvilinear surface that satisfies all three conditions is:

$$V_{DT-NG}(\lambda, p) \approx V_{DET} + [V_{NEG} - V_{DET}]\left[1 - \frac{p}{2} - \frac{p^2}{2}\right] \qquad (4.50)$$

In the absence of exact results, we propose the above expression as a suitable approximation for the variance $V_{DT\text{-}NG}$ of the virtual arrival traffic offered by the first cell to the virtual cell under det-neg channel holding time distributions and finite arrival rates. In the results section, we illustrate the remarkable accuracy of the above approximate expression

4.4.4.6 The Limiting Virtual Arrival Traffic Under Gamma Distributions [66]

In figure 4-2b, when the channel holding time distribution $f(t)$ is gamma distributed and the limit $\lambda \rightarrow \infty$ is taken, the queuing scenario in the virtual cell reduces to a C number of independent and identical GAM/GAM/∞ queuing systems. The GAM/GAM/∞ queuing system is one where the inter-arrival and service distributions have the same gamma pdf. The solution of the GAM/GAM/∞ queuing system is cumbersome even with the added simplification that the arrival process is renewal and that the inter-arrival and service distributions are identical gamma distributions. To ensure ease-of-application and tractability of our results, we bypass exact analysis using the following simple but novel method.

Let us define $V_{G=GAM}(c)$ as the variance of carried traffic in a GAM/GAM/∞ queue for shape parameter c. Firstly, we know two points on the curve $V_{G=GAM}(c)$, namely that $V_{G=GAM}(1) = 1$ for shape parameter $c=1$ and $V_{G=GAM}(\infty) = 0$ for shape parameter $c=\infty$. However, it is possible to obtain more points on the curve $V_{G=GAM}(c)$, by considering the points where the shape parameter c is integral in nature. When the shape parameter c takes on values of k, where k is a positive integer, the gamma distribution reduces to an Erlang-k

distribution. The queuing system reduces to an $E_k/E_k/\infty$ system. However, even the $E_k/E_k/\infty$ system is quite cumbersome to solve unless k takes on small values. In table I, we present numerical solutions for the variance $V_{G=GAM}(c)$ when c takes on small integral values. These variance values were obtained by solving specific $E_k/E_k/\infty$ systems using multi-dimensional Birth-Death models: We derived the state equations by equating probability transitional rates to and from each state and phase of the $E_k/E_k/\infty$ system (similar to the work in pages 114-160 of [47]). Then using probability generating functions and thereafter moment generating functions we determined the variance of carried traffic in the $E_k/E_k/\infty$ system. For conciseness sake we do not present the derivation here but simply the results. In Appendix A we present the derivation for the case where $c = 2$. The difficulty of solving for $V_{G=GAM}(c)$ for large c, is that the dimensionality of the problem increases proportionally with the square of the shape parameter c when using the method of memory-less stages and phases.

Table I: The variance of carried traffic in a G/G/∞ queuing system

Shape parameter, c	Variance, $V_{G=GAM}(c)$	Alternate representation of Variance, $V_{G=GAM}(c)$
1	1	$\frac{2x(1)}{(2x1)}$
2	0.75	$\frac{2\,(1x3)}{(2x1)\,(2x2)}$
3	0.625	$\frac{2\,(1x3x5)}{(2x1)\,(2x2)\,(2x3)}$
4	0.546875	$\frac{2\,(1x3x5x7)}{(2x1)\,(2x2)\,(2x3)(2x4)}$
∞	0	0

As can be seen from the second column the variance $V_{G=GAM}(c)$ decreases with increasing shape parameter c. Interestingly, there appears to be a proper mathematical structure in this progression as the shape parameter takes integral steps from $c = 1,2,3$ onwards. It is exactly the structure of this mathematical progression that we have presented in the third column. The mathematical progression was obtained via simple inspection.

Assuming our proposed mathematical progression in the third column is true, for any shape parameter $c=k$, where k is a positive integer, the variance $V_{G=GAM}(c)$ becomes:

$$V_{G=GAM}(c) = \frac{2\,(1x3x5x7\,.....\,[2c-3]x[2c-1])}{(2x1)\,(2x2)\,(2x3)(2x4).........(2x[c-1])\,(2xc)} \qquad (4.51)$$

We are interested in non-integral values of c as well. We can rewrite the above expression in the following form and thus extend it to non-integral values of c.

$$V_{G=GAM}(c) = \frac{2\,(1x3x5x7\,.....\,[2c-3]x[2c-1])}{(2x1)\,(2x2)\,(2x3)(2x4).........(2x[c-1])\,(2xc)}\;\frac{(2x4x6x8\,.....\,[2c-2]x2c)}{(2x4x6x8\,.....\,[2c-2]x2c)} \qquad (4.52)$$

$$V_{G=GAM}(c) = \frac{2\Gamma(2c+1)}{2^c\,2^c\,\Gamma(c+1)\Gamma(c+1)} \qquad (4.53)$$

where $\Gamma(c+1)$ is the gamma function (equation 4.9) and for integral values of c, $\Gamma(c+1) = c!$

We now propose that the above expression gives the variance of carried traffic $V_{G=GAM}(c)$ in a GAM/GAM/∞ queue where the inter-arrival and service distributions are identical gamma pdf with mean $1/\mu$. Comparison with simulation results (see results section) shows this expression to be very accurate. However, analytical proof using present day techniques seem very difficult. The difficulty lies in the fact that a gamma distribution with a non-integral shape parameter c no longer lends itself to the method of stages and phases. Since it is quite difficult to prove the above expression for $V_{G=GAM}(c)$ for non-integral values of c, we leave it as a conjecture.

4.4.4.7 The Virtual Arrival Traffic Under Gamma Channel Holding Times And Finite Arrival Rates

Let us consider the variance of the virtual arrival traffic under finite arrival rates in the first cell, for channel holding times that are gamma distributed with mean $1/\mu$ and shape parameter c where $0 \leq c < \infty$. For a fixed channel size $C=C_1$ in the first cell and a fixed service rate $\mu = \mu_1$, the variance $V_{GAM}(\lambda, c)$ is a curvilinear surface that is a function of the mean arrival rate λ and the gamma shape parameter c. We know the two curves that lie on this curvilinear surface for finite arrival rate λ. We also know the curve that occurs when the limit $\lambda \rightarrow \infty$ is taken. [Please note the difference in our usage of the uppercase "C " and the lowercase "c "].

1. We know that the curve $V_{GAM}(\lambda, 1) = V_{NEG}$ describes the curvilinear surface when the shape parameter $c=1$.
2. We also know that the curve $V_{GAM}(\lambda, \infty) = V_{DET}$ describes the curvilinear surface when the shape parameter c grows without bound.
3. We know that the curve $V_{GAM}(\infty, c) \approx C_1 \dfrac{2\Gamma(2c+1)}{2^c\, 2^c\, \Gamma(c+1)\Gamma(c+1)}$ is an accurate approximation for the curvilinear surface when the limit $\lambda \to \infty$ is taken (applying equation (4.53) in equation (4.44)).

The simplest curvilinear surface that satisfies all three conditions is:

$$V_{GAM}(\lambda, c) \approx V_{DET} + [V_{NEG} - V_{DET}]\left[\frac{2\Gamma(2c+1)}{2^c\, 2^c\, \Gamma(c+1)\Gamma(c+1)}\right] \qquad \textbf{(4.54)}$$

In the results section, we will illustrate the accuracy of the above approximate expression for the variance of the virtual arrival traffic offered by the first cell to the virtual cell under gamma channel holding time distributions.

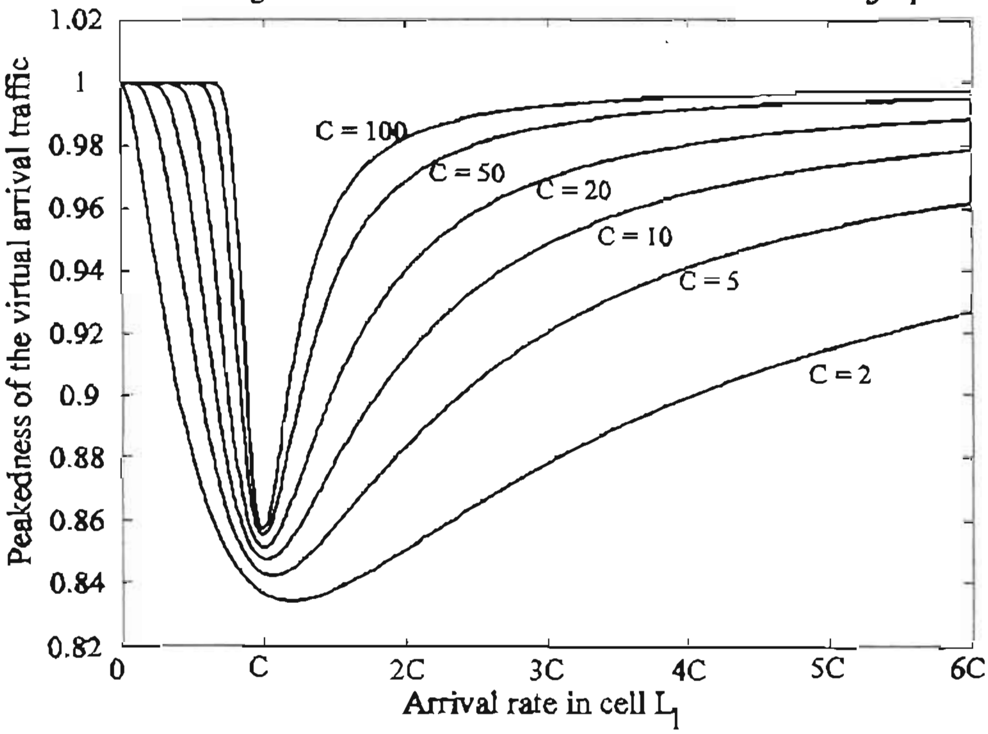
So far we have examined our cell traffic characterising model for determining the mean and variance of hand-off traffic offered by a cell to its neighbour in a simple two-cell scenario. We need two other models to be able use our proposed two-cell model to analyse a complete cellular network. We need a model to determine the blocking experienced by each traffic stream in each cell. We also need a method to extend our cell traffic characterising model from a simple two-cell model to one that is applicable for a multi-cell case where a cell has many neighbours and is able to receive and offer hand-off traffic to each one of its neighbours. We look at models for these purposes in subsequent chapters. Now we look at the accuracy of our expressions for characterising the traffic offered in our two-cell scenario.

4.5 RESULTS

4.5.1 Virtual Arrival Traffic under negative exponential channel holding times

In our cell traffic characterising model, we have shown that the virtual arrival traffic is fundamental to calculating hand-off traffic. Equations (4.19) and (4.20) are of such a nature that it is sufficient to show that the virtual arrival traffic [M_{GEN} ; V_{GEN}] is smooth (or peak), to prove that the respective hand-off arrival traffic [M_{gen} ; V_{gen}] is smooth (or peak). Simple rearrangement of equations (4.19) and (4.20) show that the peakedness, where peakedness =variance/mean, of the hand-off arrival traffic is less than one (or greater than one), if and only if, the peakedness of the virtual arrival traffic is less than one (or greater than one). In this section, we empirically show that the virtual arrival traffic under negative exponential channel holding times has a peakedness Z_O, where $Z_O = V_{NEG}/M_{NEG}$, in the range (0,1). For this purpose we chose the channel size in cell L_I to be one of the following values [2,5,10,20,50,100]. We normalised the service rates μ in cell L_I and μ in cell L_v to be one call per unit of time. For each possible value of channel size C, we varied the arrival rate λ from 0 to 6C. The peakedness of the virtual arrival traffic was determined using equations

Figure 4-5: Peakedness of the Virtual Arrival Traffic under neg-exp CHT



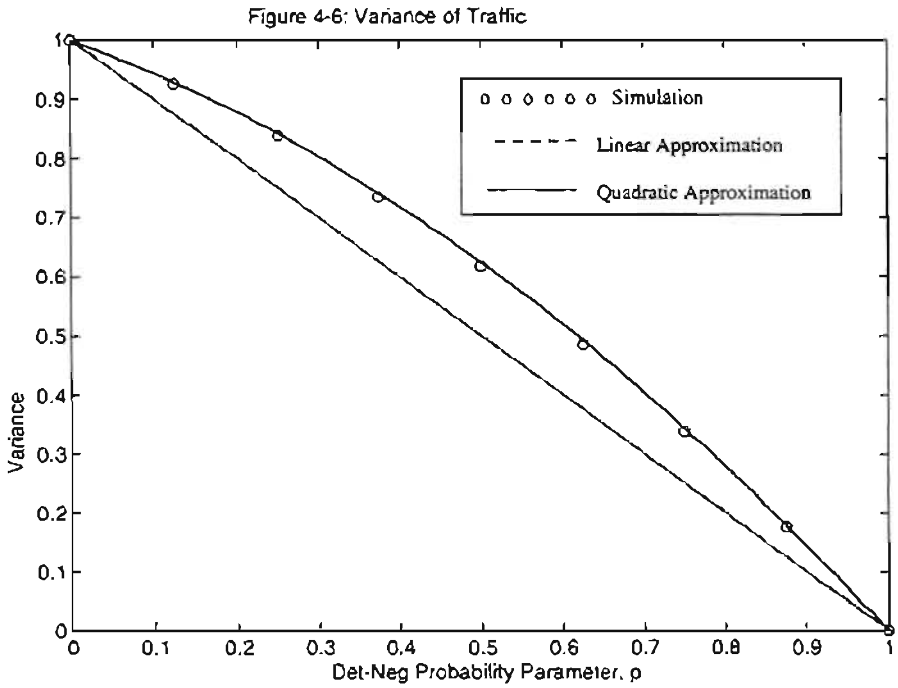
(4.33-4.38) and plotted on figure 4-5 for the various channel sizes. The vertical axis represents the peakedness of the virtual arrival traffic and the horizontal axis represents the arrival rate in cell L_1 as a multiple of the channel size C. The following conclusions may be inferred from figure 4-5. The virtual arrival traffic is a smooth process for a large range of channel sizes and arrival rates for negative exponential channel holding times. A more rigorous statement would require the analytical proof that the virtual arrival traffic is a smooth process and this formal mathematical proof is beyond the scope of this work. The peakedness Z_O of the virtual arrival traffic, for the different channel sizes examined, is a concave function with the following asymptotic behaviour: Z_O tends to 1 as A tends to 0, and Z_O tends to 1 as A tends to ∞. The peakedness Z_O of the virtual arrival traffic, for the different channel sizes examined, has a minimum approximately in the region where the offered traffic $A=\lambda/\mu\approx C$. The region $A\approx C$ is the optimal operating region for a cellular network; it would be a waste of resources to operate a cellular network in the region $A<<C$ and it would be bad service to the customers to operate in the region $A>>C$.

4.5.2 Virtual Arrival Traffic under det-neg and gamma channel holding times

In the first subsection, we show that the expressions that we derived for the variance of carried traffic in the infinite queues, DN/DN/∞ and GAM/GAM/∞, are accurate. In the second subsection, we illustrate the accuracy of the expressions for the variance of the virtual arrival traffic, under det-neg and gamma channel holding time distributions (equations 4.50 & 4.54), for finite arrival rates λ and different channel sizes C in our virtual cell scenario.

4.5.2.1 Variance of carried traffic in the DN/DN/∞ queue.

We simulated a DN/DN/∞ queue, having identical det-neg inter-arrival and service distributions. We simulated an infinite server system by having a system with a very large number of servers. The mean offered traffic, as can be expected under identical inter-arrival and service distributions, is 1 Erlang. From our simulation, we determined the variance of carried traffic in the "infinite" server system for different det-neg probability parameters, p. We then analytically evaluated the variance of carried traffic in the infinite server system for different det-neg probability parameters, p, using the linear

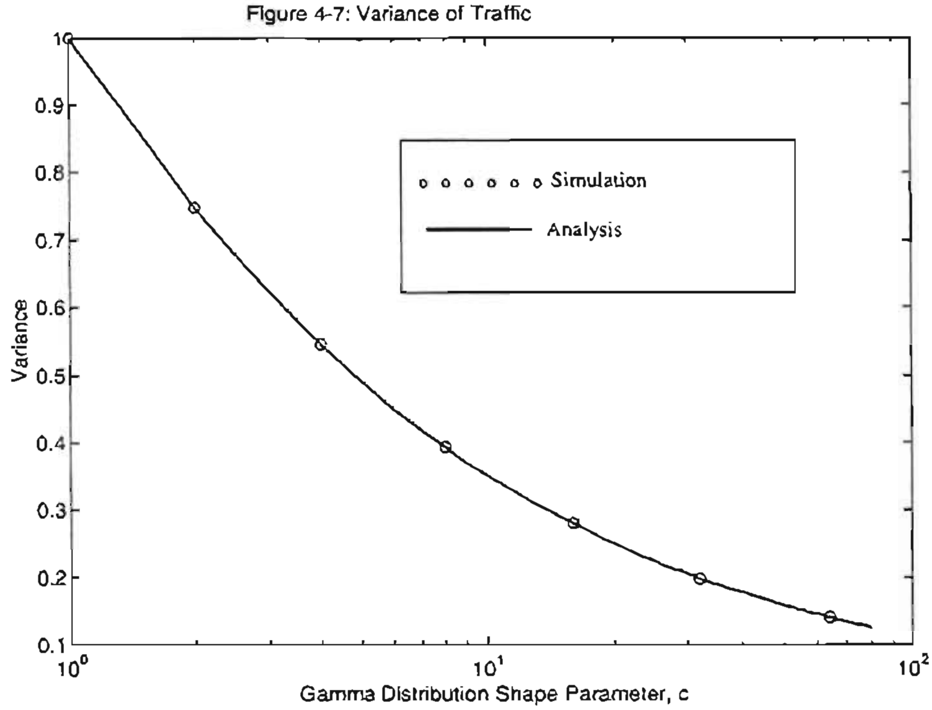


approximation $V_{G=DT-NG} \approx (1-p)$ and the quadratic approximation $V_{G=DT-NG} = (1-\frac{p}{2}-\frac{p^2}{2})$. Figure 4-6 illustrates these results. As can be seen, the quadratic approximation is superior to the linear approximation when compared to the simulation results. However, the true value of this quadratic approximation is that it enables us to bypass the exact and probably intractable analysis of a DN/DN/∞ queue by way of a very simple result that is easy to apply and remarkably accurate.

4.5.2.2 Variance of carried traffic in the GAM/GAM/∞ queue.

We also simulated a GAM/GAM/∞ queue, having identical gamma inter-arrival and service distributions. The mean offered traffic, under identical inter-arrival and service distributions, is 1 Erlang. From our simulation, we determined the variance of carried traffic in the infinite server system for different gamma shape parameters, c. We then analytically evaluated the variance of carried traffic in the infinite server system for different gamma shape parameters, c, using our conjecture $V_{G=GAM}(c) = \frac{2\Gamma(2c+1)}{2^c 2^c \Gamma(c+1)\Gamma(c+1)}$ from earlier. Figure 4-7 illustrates these results. As can be

Figure 4-7: Variance of Traffic



seen, our conjecture is very accurate when compared to the simulation results. However, due to the difficulties associated with proving the above result, especially for cases where the shape parameter c is non-integral, the proof seems far from feasible and our result remains a conjecture.

4.5.3 Variance of the virtual arrival traffic.

In our cell traffic characterising model, we have shown that the virtual arrival traffic is fundamental to calculating hand-off traffic. Provided, we know the mean and variance of the virtual arrival traffic, we are able to determine the mean and variance of any hand-off arrival traffic where hand-off probability $Q_h \leq 1$ (equations 4.19 & 4.20). In this section, we evaluate the accuracy of our expression for the variance of the virtual arrival traffic offered by the first cell to the virtual cell (in figure 4-2b) under det-neg as well as gamma channel holding times. In both cases, we set channel size C to be one of the following values: $C \in \{3, 30\}$. For each channel size C, we varied the offered traffic in the first cell, $A = \lambda/\mu$, from $A = 0$ Erlangs up to $A = 2.666 \times C$ Erlangs. We analytically determined the peakedness, Z = Variance / Mean, of the various virtual arrival traffic for the above parameters. For comparison purposes we also simulated the simple virtual cell scenario under the same conditions. To *imitate* the infinite sized virtual cell of figure 4-2b, we set the channel size in the virtual cell in our simulator to be very large, namely 400 channels.

4.5.3.1 Variance of the virtual arrival traffic under det-neg channel holding time distributions

We set the channel holding time distributions in both cells to be identical det-neg distributions with a normalised mean of $E[T]=1/\mu=1$ unit time. We considered four det-neg distributions with det-neg probability parameter $p \in \{0, 1/3, 2/3, 1\}$. We plotted the simulation and analytical results in figures 4-8 & 4-9.

For the two different channel sizes that we considered, the peakedness of the virtual arrival traffic as determined by our quadratic approximation (equation 4.50) ties up extremely well with simulation results when compared to the linear approximation. Similar results were obtained for various other channel sizes that we considered. However, to avoid repetition,

Figure 4-8: Peakedness of the Virtual Arrival Traffic for channel size C = 3

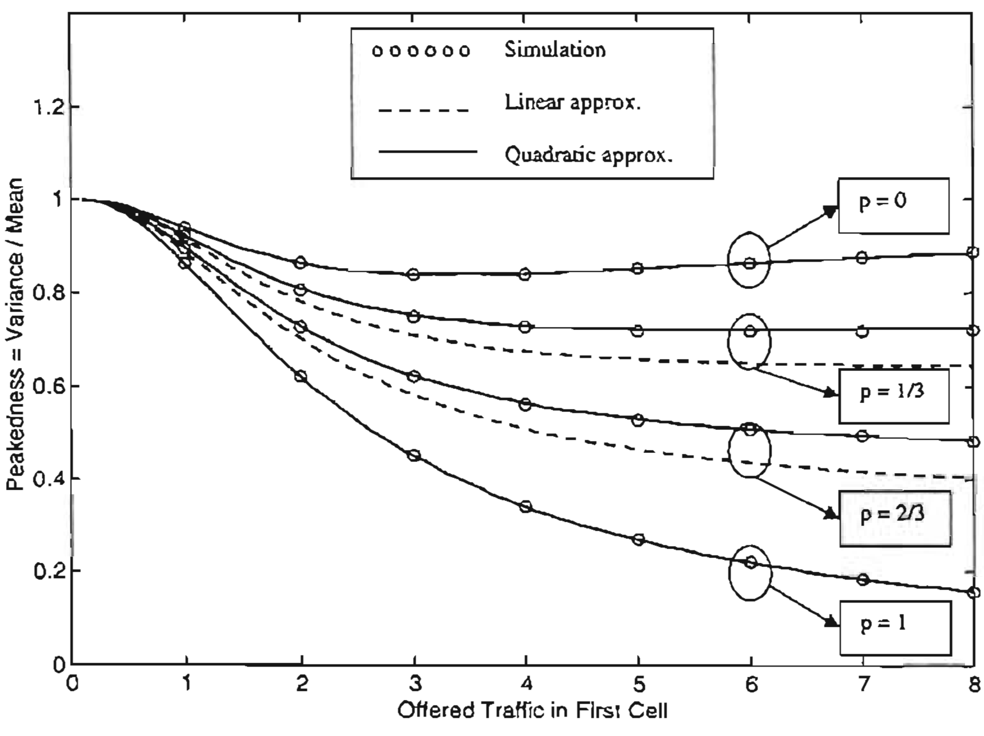
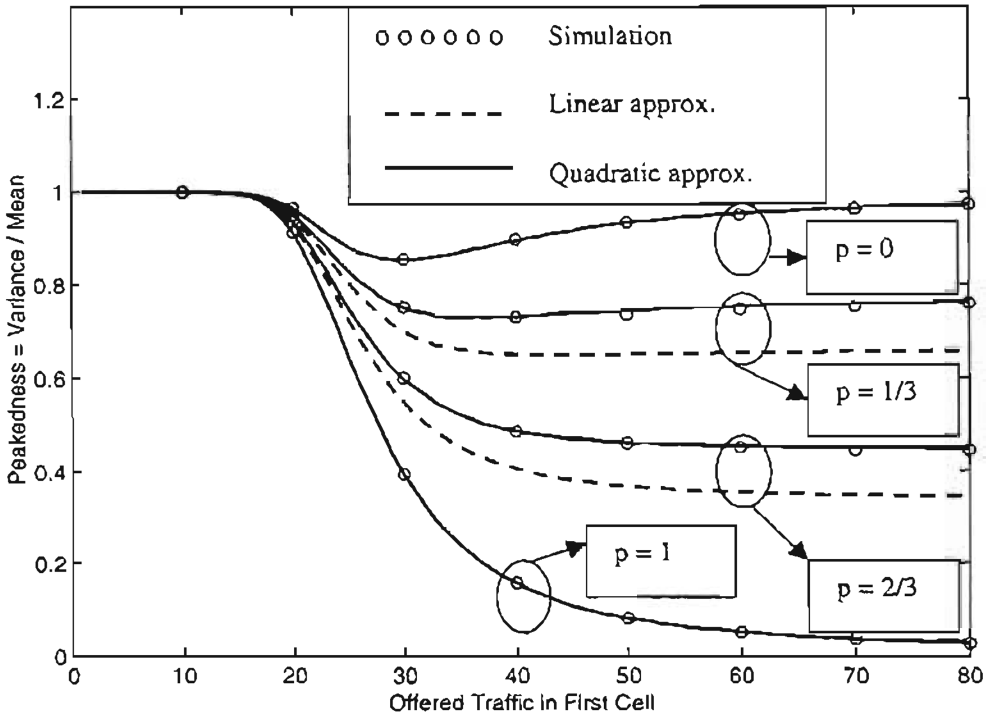


Figure 4-9: Peakedness of the Virtual Arrival Traffic for channel size C = 30



we have not presented those results here. The remarkable accuracy of the quadratic approximation for the variance of the virtual arrival traffic (equation 4.50) is a testament to the validity of the asymptotic analysis that we employed, where, we considered the behaviour of the virtual cell system when the arrival rate λ increased without bound and then borrowed the structure of the results to apply for finite arrival rates.

4.5.3.2 Variance of the virtual arrival traffic under gamma channel holding time distributions

In this section, we set the channel holding time distributions in both cells to be identical gamma distributions with a normalised mean of $E[T]=1/\mu=1$ unit time. We considered four gamma distributions with shape parameter $c \in \{1, 3, 20, \infty\}$. We plotted the simulation and analytical results in figures 4-10 & 4-11.

Figure 4-10: Peakedness of the Virtual Arrival Traffic for channel size C = 3

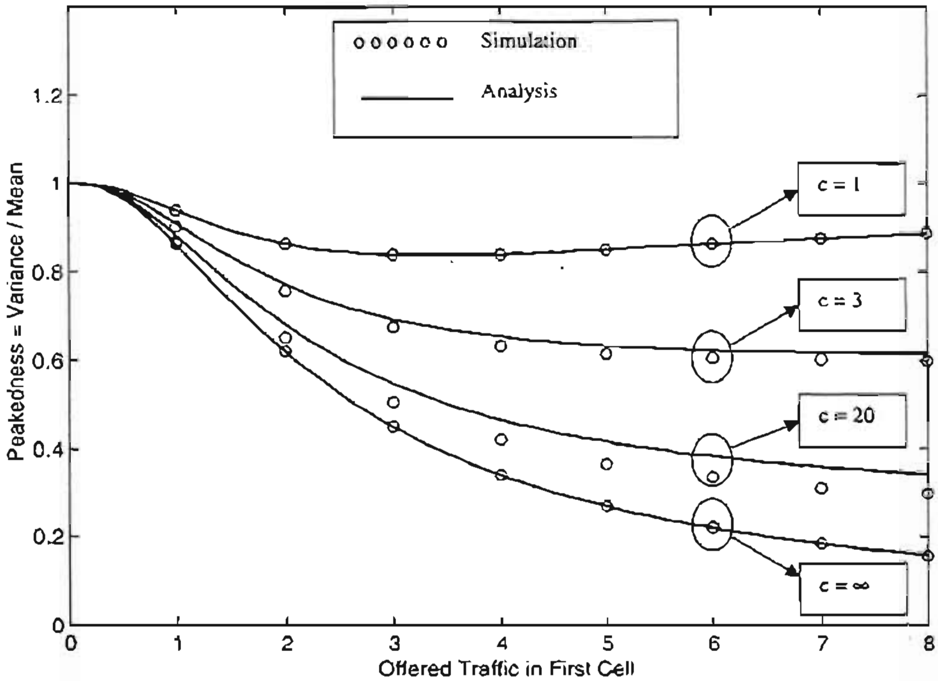
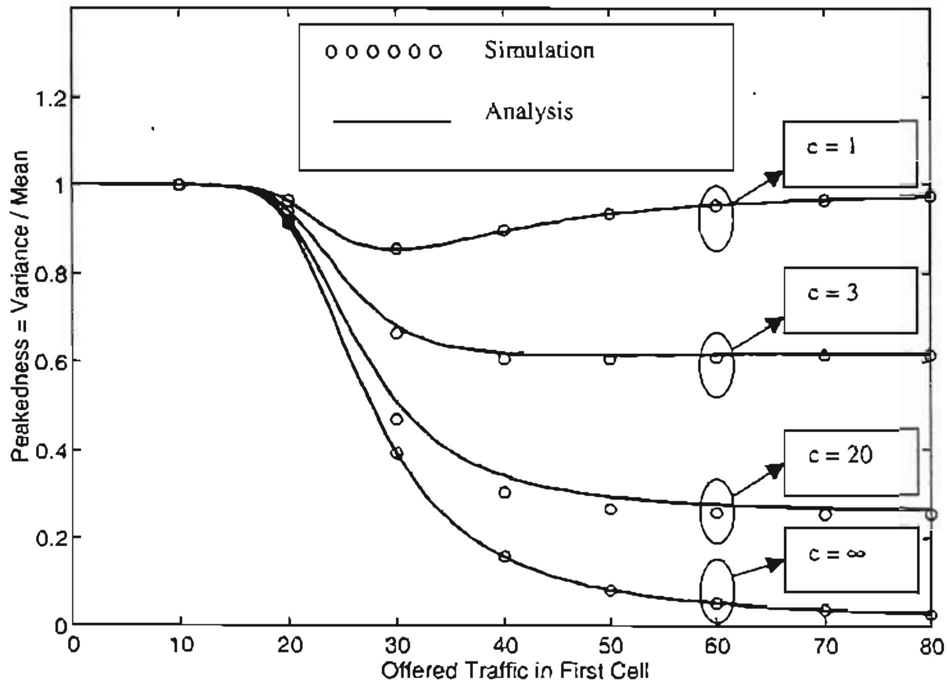


Figure 4-11: Peakedness of the Virtual Arrival Traffic for channel size C = 30



For the two different channel sizes that we considered, the peakedness of the virtual arrival traffic as determined by our analysis (equation 4.54) ties up reasonably well with simulation results. As can be seen, agreement between simulation and analysis was strongest for $c=1$ (where the channel holding times are negative exponential) and $c=\infty$ (where channel holding times are deterministic). This is to be expected since we have exact results for these shape parameter values (equations 4.33-4.38 and equation 4.39). For other values of c, our expression (equation 4.54) is a reasonable approximation and one that we show in the next section to work well in practice. We obtained similar results for various other channel sizes that we considered.

4.6 Summary

Our proposed performance analysis algorithm has three parts, a cell traffic characterising model, a cell traffic blocking model and a quality of service evaluation model. At the heart of our cell traffic characterising model is a simple two-cell scenario. In this chapter we analysed this two-cell scenario and presented results for the mean and variance of traffic offered by a cell to its neighbour in the two-cell scenario. We considered the following four different channel holding time distributions in our two-cell scenario:

- Negative exponential distribution
- Deterministic distribution
- Det-neg distribution
- Gamma distribution

The results that we presented for the neg-exp and deterministic distributions are exact whereas the results that we presented for the det-neg and gamma distribution are approximate, based on asymptotic analysis. We considered the behaviour of our two-cell / virtual cell system when the arrival rate λ increased without bound and then borrowed the structure of the results to apply for finite arrival rates. We have validated the accuracy of all our derivations using comparison with simulation results.

CHAPTER 5

5.1 Introduction

Our proposed approach to the performance analysis of cellular networks is a moment-based one that avoids the state space explosion of existing methods but still allows the various traffic processes to be modelled beyond the mean. As shown earlier, for simplification, we loosely decomposed the performance analysis of a cellular network into three parts, a generic *cell traffic characterising model*, a generic *cell traffic blocking model* and a *quality of service evaluation* model. In the previous chapter we considered the basic essence of our cell traffic characterising model, namely, the simple two-cell scenario. For the simple two-cell scenario, we determined the mean and variance of hand-off traffic offered by a cell to its neighbour under various channel holding time distributions. In this chapter we consider various cell traffic blocking methods that are suitable for use in the performance analysis of cellular networks.

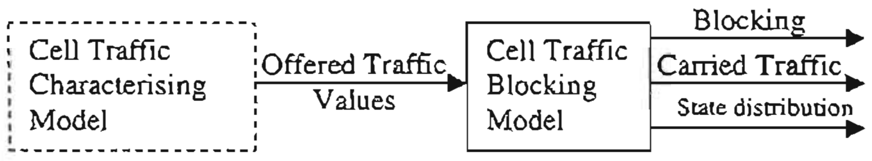


Figure 5-1: Cell Traffic Blocking Model

Figure 5-1 illustrates the function of a cell traffic blocking model and how it inter-relates to a cell traffic characterising model. Essentially, the cell traffic characterising model provides the inputs for the cell traffic blocking model, in terms of offered traffic values. The cell traffic blocking model uses these offered traffic values to determine blocking probabilities, carried traffic values and state probability distributions in a finite server system. Since existing single moment methods [15,16] model new and hand-off call arrival processes as Poisson, they employ the Erlang-B formula and its state-dependant versions as suitable cell traffic blocking methods. In our two moment analysis, the new call arrival process is Poisson but the hand-off arrival process is normally a general arrival process. It might be possible to analyse both the new and hand-off call traffic processes exactly using non/semi-Markov processes and thus determine the blocking experienced by the two streams. However, such an approach is too cumbersome and probably very time consuming from a computational point of view. We therefore follow the *moment matching* approach used in fixed network analysis ([14] pp. 112-124), whereby, a general traffic process is replaced by a simpler traffic process having identical first few moments. It is for the simpler traffic processes that tele-traffic parameters such as blocking, carried traffic etc are determined. The general traffic process is then assumed to suffer the same fate as its equivalent process.

5.1.1 Time congestion and call congestion

Consider an arbitrary traffic process of mean M_G and variance V_G that is offered to a finite system of K servers. The cell traffic blocking method may be used to determine two important blocking quantities for this system, namely, time congestion, E, and call congestion, B. Time congestion, E, is the fraction of time when all servers are busy. Call congestion, B, is the probability that an arriving call will find all trunks busy. Note that for a general arrival process, E ≠ B, time congestion does not equal call congestion. However, as noted by Girard [14], for the Poisson arrival process, the time congestion probability is the same as the call congestion probability. Girard also showed that, given that the carried traffic in the finite server system is M_C, the call congestion B is:

$$B = \frac{M_G - M_C}{M_G} = \frac{M_L}{M_G} \tag{5.1}$$

where M_L is the lost or blocked traffic.

In this chapter, we look at two suitable traffic blocking models from the fixed network arena. We then extend these models for the cellular network scenario where two distinct traffic streams, namely, fresh and hand-off traffic, may be offered to a cell. We further extend the existing models to accommodate the case where channel reservation of level r is implemented to protect hand-off calls at the expense of new calls. In the previous chapter we showed that the hand-off traffic is smooth (or peaked) if and only if the virtual arrival traffic is smooth (or peaked). In the results section of the previous chapter, we have shown that the virtual arrival traffic is smooth for a large number of arrival rates, channel holding time distributions and channel sizes. However, we have chosen cell traffic blocking models that are equally applicable for both smooth and peaked traffic.

5.2 Existing two-moment blocking models from the fixed network arena.

5.2.1 Delbrouck's BPP blocking model

Delbrouck suggested that a general arrival process may be moment matched using an arrival process based on the BPP or Bernoulli-Poisson-Pascal distributions [45]. Consider an arrival process where the arrival rate, λ_j, in each state, j, is dependant on the state j. Let us assume that this arrival rate is $\lambda_j = q(n+j)$. Assume that calls are served with time intervals that are negative exponentially distributed and that each call departs at a mean call departure rate of μ. Equivalently the departure rate μ_j from each state j, is $\mu_j = j\mu$.

The equilibrium equations that describe this arrival process are easily derived and may be used to determine the offered traffic in a fictitious system of infinite servers. The equilibrium equation is as follows:

$$(j\mu + (n+j)q)p_j = (j+1)\mu p_{j+1} + q(j-1+n)p_{j-1} \quad (5.2)$$

The above equations may be solved under the assumption that the sum of the state probabilities equal one:

$$\sum_{j=0}^{\infty} p_j = 1 \quad (5.3)$$

Delbrouck derived the state probabilities as defined by the above system to be:

$$p_j = (1-\frac{q}{\mu})^n (\frac{-q}{\mu})^j \binom{-n}{j} \quad (5.4)$$

Thereafter, the mean M_P and variance V_P of offered traffic (or equivalently carried traffic in an infinite server system) for this arrival process may be determined:

$$M_P = \frac{qn}{\mu - q} \; ; \; V_P = \frac{M_P}{1-\frac{q}{\mu}} \quad (5.5)$$

It can be seen that the peakedness, $Z_P=V_P/M_P$ of the above distribution is greater than 1, when $0<q<\mu$ and $n>0$, which makes the above distribution ideal for matching peaked traffic. Alternatively, the peakedness, Z_P of the above distribution is less than 1, when $q<0$ and $n<0$, which makes the above distribution ideal for matching smooth traffic.

5.2.1.1 Moment Matching using the BPP Process

The BPP traffic process may also be used as part of a moment-matching procedure to determine parameters of interest, when an arbitrary (general) traffic process of mean M_G and variance V_G is offered to a cell L of K servers. The parameters q and n of the BPP process are chosen such that the mean M_P and the variance V_P produced by the BPP process are the same as that of the arbitrary traffic process that we are trying to model. Parameters q and n may be derived by inverting equation (5.5):

$$q = \mu(1 - \frac{M_G}{V_G}) \quad (5.6)$$

$$n = M_G \left(\frac{\mu - q}{q} \right) \tag{5.7}$$

The parameters of interest in cell L, such as the state probability distribution as well as the time congestion and the call congestion experienced by the arbitrary traffic process of mean M_G and variance V_G are then assumed to be the same as those experienced by the BPP process in the same cell. When applying the BPP process to a finite sized cell of K channels, the trend in existing studies has been to truncate the *infinite* state distribution of equation (5.4) at the boundary point K. The truncated BPP distribution, denoted by q_j for the sake of distinction between it and the state probability p_j from the infinite distribution, is given by:

$$q_j = \frac{p_j}{\sum_{i=0}^{K} p_i}, \quad for\ j = 0 \ldots\ldots K \tag{5.8}$$

5.2.1.2 Time Congestion for the truncated BPP process

The time congestion E, experienced in cell L of size K, may be obtained using the recursion [45]:

$$E_0 = 1 \tag{5.9}$$

$$E_k = \frac{q(n+k-1)E_{k-1}}{k + q(n+k-1)E_{k-1}} \tag{5.10}$$

Where E_k is the time congestion observed when the number of servers in the system equals k = 1,2.......K.

5.2.1.3 Call Congestion for the truncated BPP process

Delbrouck [45] determined that the call congestion B experienced by the BPP arrival process:

$$B = E_K \left(1 + \frac{K}{M_G} (Z_G - 1) \right) \tag{5.11}$$

The general traffic process [M_G;V_G] is then assumed to suffer the same blocking. Delbrouck's model is only one of many models that have been suggested in the literature for modelling peaked and smooth traffic. Another useful model is that proposed by Sanders, Haemers and Wilcke.

5.2.2 Two moment traffic blocking model by Sanders, Haemers and Wilcke

Sanders et al's paper [67] presents a two moment model for telephone traffic and discusses how to determine the congestion experienced by the two moment traffic stream when offered to a finite system of K servers. In following work we denote every traffic stream by the pair (M,V) where M is the mean and V is the variance. We use the triplet (M,V,K) to denote the system where the traffic stream (M,V) is offered to a finite system of K servers. Following the fixed network terminology of the above authors we use the term "trunks" as opposed to "channels" in the following work. In the description we present below we consider smooth traffic where $V < M$. Sander's et al's model for peaked traffic can be found in [67].

The authors suggest that any smooth traffic process of mean and variance (M_G,V_G) may be represented as the sum of two independent but stochastically different traffic streams (V_G,V_G) and (M_G-V_G,0). The first stream (V_G,V_G) is a Poisson traffic source with mean and variance V_G. The second stream is a *constant* traffic stream of mean M_G-V_G and zero variance. The total mean and variance of the occupancy distribution of an infinite server system where the two *independent* streams (V_G,V_G) and (M_G-V_G,0) are assumed to occur is simply M_G and V_G respectively.

Consider now the occupancy distribution of a finite server system where the above two streams are assumed to occur. Consider for arguments sake that M_G-V_G is integral in value and that the number of servers $K>M_G-V_G$. Then the constant traffic stream (M_G-V_G ,0) will *permanently* occupy precisely M_G-V_G trunks [67}. The Poisson source (V_G,V_G) ends up having access to only the remaining trunks of size [K-(M_G-V_G)}. In this manner Sanders et al suggest the transformation of a (M_G,V_G,K) system into a (V_G,V_G, (K+V_G-M_G)) system. The latter system is simply a Poisson traffic process V_G in a finite server system of size K+V_G-M_G, and therefore the Erlang-B formula, $E[\cdot,\cdot]$, applies. Assuming that the mean blocked traffic from the original traffic stream (M_G,V_G) is the sum of the mean blocked traffic from its two constituent streams, the mean blocked traffic $M_L=V_G .E[V_G, (K+V_G-M_G)]$ since only the Poisson component and not the constant traffic component experiences blocking. Thereupon the blocking B_G (using equation 5.1):

$$B_G = \frac{V_G\, E[\, V_G \,.\, (\, K + V_G - M_G\,)\,]}{M_G} \tag{5.12}$$

The mean carried traffic M_{car} of the original smooth traffic stream (M_G, V_G) is

$$M_{car} = (M_G - V_G\,) \;+\; V_G \;.\; (\, 1 - E\,[\, V_G \,.\, (\, K + V_G - M_G\,)] \tag{5.13}$$

The variance of the carried traffic V_{car} of the original smooth traffic stream (M_G, V_G) is simply the sum of the two individual variances. This summation is valid due to the lack of interaction between the two constituents streams due to the manner in which they are defined above: they are uncorrelated. This lack of correlation arises out of the fact that the constant traffic stream occupies one set of trunks whilst the Poisson component has access to the remaining set. Since the variance of carried traffic of the constant traffic stream $(M_G - V_G, 0)$ is zero, the variance of carried traffic of the original stream $(M_G; V_G)$ is solely due to the Poisson component. Using equation (4.14) for the variance of carried traffic in a M/M/C system:

$$V_{car} = \; V_G\; (1 - E_G\,) \;-\; E_G\, V_G\; (K + V_G - M_G - V_G\; (1 - E_G\,)) \tag{5.14}$$

where $E_G = E[V_G, (K + V_G - M_G)]$. There is no problem in extending these results to the case where $M_G - V_G$ is *not* integral in nature. One may then assume that the constant traffic stream *permanently* occupies a non-integral number of $M_G - V_G$ trunks. Then the Poisson source (V_G, V_G) ends up with a reduced equivalent set of trunks of size $[K-(M_G - V_G)]$ which is non-integral in value as well. Theoretically there is no problem with a non-integral number of trunks [67] and furthermore the Erlang-B formula has been generalised to non-integral trunk values as shown below.

5.2.3 Erlang-B formula generalised for non-integral values of trunk K

For the Sanders et al's model to be valid for all values of offered traffic, it is necessary that the Erlang-B model be extended for non-integral values of channel sizes as well. The theoretical generalisation of the Erlang-B formula E[A,K] where K is non-integral is as follows:

$$E[A,K]^{-1} = A\int_0^{\infty} e^{-Ay}(1+y)^K dy \tag{5.15}$$

The mathematical properties of the above function has been studied extensively by Jagermann [68]. However, in a network performance analysis algorithm that normally involves repetitive application of the cell traffic blocking model it is necessary that the algorithm for calculating the Erlang-B formula is efficient and numerically stable. The above integration is quite slow to compute. A useful recursion that is numerically stable and reasonably efficient is the following:

$$\frac{1}{E[A,K]} = 1 + \frac{K}{A}\frac{1}{E[A,K-1]} \tag{5.16}$$

The Erlang-B formula for integral values of K can be computed recursively with the starting point:

$$E[A,0] = 1 \tag{5.17}$$

It turns out that the above recursion is valid for positive non-integral values of K as well. The starting point for K where $0<K<1$ is given by a quadratic approximation obtained by Krupp [69]:

$$E[A,K] = 1 - \frac{(A+2)K}{A^2+3A+1} + \frac{K^2}{(A+1)(A^2+3A+1)} \tag{5.18}$$

Krupp goes on to show that the error produced by the recursion for the Erlang-B formula E[A,K] for non-integral values of K (equation 5.16) that uses his starting point as given by equation (5.18) reduces with every recursive application of equation (5.16). Therefore the larger the trunk size the smaller the error. We now examine a model for channel (trunk) reservation under non-integral channel (trunk) sizes.

5.2.4 Girard's Model for channel reservation under non-integral trunk sizes

Girard [70] suggested a model for calculating the blocking experienced by two Poisson traffic streams A_F and A_H when offered to a finite server system of size K where channel reservation (or trunk reservation) of level r is implemented. We assume that the traffic of mean $A_N = A_F + A_H$ is offered to the system in states k where $0 \le k < K-r$ and that traffic of mean $A_R = A_H$ is offered in states k where $K-r \le k < K$. For ease of notation, we set $T=K-r$ and we define the probability $p(A_N, A_R, T, K)$ as the probability that all K trunks are busy and $g(A_N,A_R,T,K)$ as the probability that less than T trunks are busy. Girard derived the following recursions:

$$\frac{1}{p(A_N,A_R,T,k+1)} = \frac{k+1}{A_R}\frac{1}{p(A_N,A_R,T,k)}+1 \tag{5.19}$$

$$\frac{1}{g(A_N,A_R,T,k+1)} = \frac{1}{g(A_N,A_R,T,k)}+\frac{A_R^{k+1-T}}{(k+1)k\bullet\bullet\bullet(T+1)}\frac{E[A_N,T]}{1-E[A_N,T]} \tag{5.20}$$

where k=T,T+1....K-2,K-1 and where the initial values are given using the Erlang-B function $E[\cdot,\cdot]$:

$$g(A_N,A_R,T,T) = 1-E[A_N,T] \tag{5.21}$$

$$p(A_N,A_R,T,T) = E[A_N,T] \tag{5.22}$$

The following recursions for the first moment $M_C(A_N,A_R,T,K)$ and second moment $X_C(A_N,A_R,T,K)$ of carried traffic in the K trunks may also be obtained:

$$M_C(A_N,A_R,T,k+1) = M_C(A_N,A_R,T,k)[1-p(A_N,A_R,T,k+1)]+[k+1]p(A_N,A_R,T,k+1) \tag{5.23}$$

$$X_C(A_N,A_R,T,k+1) = X_C(A_N,A_R,T,k)[1-p(A_N,A_R,T,k+1)]+[k+1]^2 p(A_N,A_R,T,k+1) \tag{5.24}$$

for k=T,T+1,...K-2,K-1 and where the initial values are given in terms of Erlang-B function $E[\cdot,\cdot]$:

$$M_C(A_N,A_R,T,T) = A_N(1-E[A_N,T]) \tag{5.25}$$

$$X_C(A_N,A_R,T,T) = M_C(A_N,A_R,T,T)[1+A_N+T] - A_N T \tag{5.26}$$

The variance $V_C(A_N,A_R,T,K)$ of carried traffic in the K trunks is simply $X_C(A_N,A_R,T,K)-[M_C(A_N,A_R,T,K)]^2$. The crucial feature about Girard's recursions are that they are valid for non-integral values of trunk size K and T provided that the difference K-T = r is integral in value.

5.3 Cell Traffic Blocking model

Delbrouck's and Sanders et al's blocking models in the form discussed before are applicable for the case where a single traffic stream is offered to a finite server system. In cellular networks, there are two types of traffic streams, fresh and hand-off traffic that need to be modelled. Channel reservation is a further complication that needs to be accommodated in a cell traffic blocking model. In this section we extend the previously discussed models from the fixed network arena to accommodate the two different traffic streams common to cellular network analysis and as well as to include channel reservation.

5.3.1 Assumptions of our Cell Traffic Blocking Model

We assume that a cell L is offered new call arrivals with mean A_n where $A_n=\lambda_n/\mu_H$ and N independent hand-off traffic streams of mean M_{ho} and variance V_{ho} from its immediate neighbours. We assume that the mean and variance of the hand-off traffic streams were calculated using the cell traffic characterising model. Following the usual practice in fixed network analysis [14], we represent the N hand-off traffic streams by a single aggregate traffic stream of mean M_Σ and variance V_Σ where

$$M_\Sigma = N\,M_{ho} \;;\; V_\Sigma = N\,V_{ho} \tag{5.27}$$

We assume that new call arrivals are subject to channel reservation of level r for the benefit of hand-off call arrivals.

5.3.2 Application of the Delbrouck's model

In our cell traffic blocking method, we model the new call arrival process as a Poisson arrival process having the same mean A_n. We model the aggregate hand-off traffic stream $\{M_\Sigma; V_\Sigma\}$ using Delbrouck's BPP (Bernoulli-Poisson-Pascal) process [45].

We have to perform a few modifications to Delbrouck's BPP method in order to model the case where Poisson traffic $A_n = \lambda_n/\mu_H$ and aggregate hand-off traffic $[M_\Sigma;V_\Sigma]$ are offered to a cell of C channels where r channels are reserved for hand-off calls. In order to incorporate the Poisson new call arrival

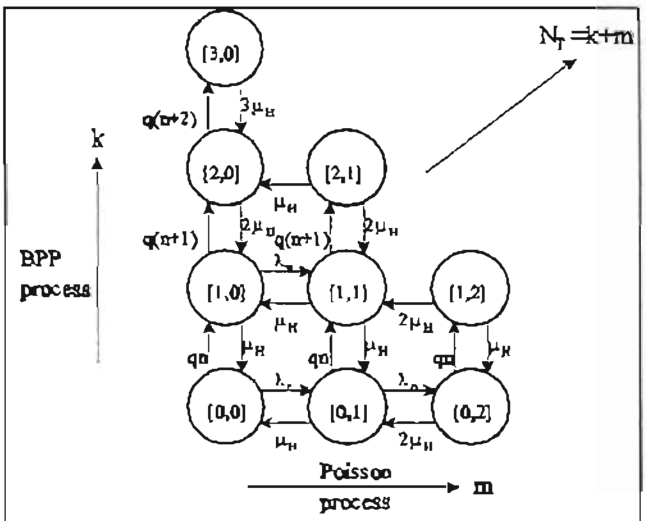


Fig. 5-2: Proposed Two Dimensional BPP process

process as well as the effect of channel reservation, we have to extend the single dimensional Birth-Death process as defined by the standard BPP process to the two-dimensional Birth-Death process shown in figure 5-2. For illustration purposes, figure 5-2 describes the case specific scenario where C=3 channels and channel reservation r=1. As can be seen, the horizontal leg of the two dimensional process models the Poisson arrival process of rate λ_n whilst the vertical leg models the BPP arrival process. The total number of calls in progress, N_T, in the three server system, is simply the sum of the number of Poisson calls in progress, m, and the number of BPP calls in progress, k. As can be seen from figure 5-2, when the number of calls in progress is less than 2, ie., $N_T<2$, both Poisson and the BPP processes compete for the free channels. In other words, a call arrival is possible due to the Poisson arrival rate λ_n or due to the BPP arrival rate $\lambda_j=q(n+j)$. However, when the number of calls in progress, N_T, is equal to 2, channel reservation comes into play, and only the BPP arrival process can effect a call arrival. Due to the effects of channel reservation, state [0,3] does not exist in figure 5-2 and its probability $p_{0,3} = 0$.

The Birth-Death equations for *arbitrary* channel size C and reservation parameter r are as follows.

For $k + m < C - r$: **(5.28)**

$$(q(n+k)+\lambda_n+k\mu_H+m\mu_H)p_{k,m} = q(n+k-1)p_{k-1,m}+\lambda_n p_{k,m-1}+(k+1)\mu_H p_{k+1,m}+(m+1)\mu_H p_{k,m+1}$$

For $k + m = C - r$: **(5.29)**

$$(q(n+k)+k\mu_H+m\mu_H)p_{k,m} = q(n+k-1)p_{k-1,m}+\lambda_n p_{k,m-1}+(k+1)\mu_H p_{k+1,m}+(m+1)\mu_H p_{k,m+1}$$

For $C - r < k + m < C$: **(5.30)**

$$(q(n+k)+k\mu_H+m\mu_H)p_{k,m} = q(n+k-1)p_{k-1,m}+(k+1)\mu_H p_{k+1,m}+(m+1)\mu_H p_{k,m+1}$$

For $k + m = C$: **(5. 31)**

$$(k\mu_H+m\mu_H)p_{k,m} = q(n+k-1)p_{k-1,m}$$

where $p_{k,m}$ is the probability of having k BPP calls and m Poisson calls. Note that in equations (5.28-5.31), probability $p_{k,m}$=0 for k<0 and/or m<0. These equations may be solved using Gauss-Seidel iteration [71] to determine individual state probabilities, $p_{k,m}$ under the initial condition that the sum of the state probabilities add up to 1. The solution of equations (5.28-5.31) results in state probabilities, $p_{k,m}$=0 for values of m>C-

r, because these states do not exist under channel reservation of level r. The mean M_{car} and variance V_{car} of carried traffic in the two dimensional Birth-Death model is as follows:

$$M_{car} = \sum_{k=0}^{C} \sum_{m=0}^{C-k} (k+m) p_{k,m} ; \quad V_{car} = \left(\sum_{k=0}^{C} \sum_{m=0}^{C-k} (k+m)^2 \; p_{k,m} \right) - M^2_{car} \tag{5.32}$$

The blocking p_{nb} experienced by the Poisson arrival process and its moment matching equivalent the new call arrival process A_n due to channel reservation is as follows:

$$p_{nb} = \sum_{k=0}^{C} \sum_{m=\max(0,C-r-k)}^{C-k} p_{k,m} \tag{5.33}$$

The call congestion p_{hb} experienced by the BPP process and hence its moment matching equivalent, the aggregate hand-off offered traffic stream $[M_\Sigma;V_\Sigma]$, is simply the mean aggregate hand-off offered traffic minus the mean aggregate hand-off carried traffic divided by the mean aggregate hand-off offered traffic (see equation (5.1)):

$$p_{hb} = \frac{M_\Sigma - (M_{car} - A_n(1 - p_{nb}))}{M_\Sigma} \tag{5.34}$$

Finally, the individual hand-off offered traffic stream $[M_{ho}; V_{ho}]$ is assumed to experience the same blocking p_{hb} as the aggregate hand-off traffic stream $[M_\Sigma;V_\Sigma]$ since they both have the same peakedness (=variance/mean). This is an assumption borrowed from fixed network analysis [14]. Although, Delbrouck derived his BPP process under negative exponential service time distributions, we shall show in the results section that Delbrouck's BPP process works well for the any type of service time distributions. As stated before, the BPP arrival process is a state dependent Poisson arrival process. The reason for the BPP process' suitability, when moment matching aggregate hand-off traffic streams under any type of service distributions, may be attributed to the fact that the state-dependent Poisson arrival rate in the BPP process is insensitive to service distributions beyond the mean [59].

5.3.3 Application of the Sanders et al's model

We use Sanders et al.'s method and Girard's method from earlier to determine new call blocking p_{nb} and hand-off call blocking p_{hb}, by *moment matching* the various traffic streams offered to a cell with equivalent processes. The fresh traffic offered to a cell is replaced by a Poisson process of the same mean. The aggregate hand-off traffic stream is represented using two component streams. Using Sanders et al's method, we split the aggregate hand-off offered traffic stream (M_Σ,V_Σ) into a Poisson traffic stream (V_Σ,V_Σ) and a constant traffic stream (M_Σ-V_Σ , 0). From the two traffic streams, namely fresh traffic of mean A_n and aggregate hand-off traffic (M_Σ,V_Σ), offered to a cell, we obtain, via moment matching, three traffic streams: two Poisson streams (A_n,A_n) & (V_Σ,V_Σ) and a constant traffic stream (M_Σ-V_Σ , 0).

Following Sanders et al, the constant traffic stream is assumed to occupy permanently M_Σ-V_Σ out of the total C servers available in a cell. The Poisson streams (V_Σ,V_Σ) and (A_n,A_n) have access to the remaining servers. Since the aggregate hand-off traffic is not subjected to channel reservation, its Poisson component (V_Σ,V_Σ) has access to all of C-(M_Σ-V_Σ) channels. But, since the fresh traffic (A_n,A_n) is subject to channel reservation of level *r*, it has access to only C-(M_Σ-V_Σ)-*r* channels. We may now use Girard's model for trunk reservation to determine the state probabilities. Using the above deliberations we may determine the various parameters K,T,A_N and A_R of Girard's model as follows:

$$K = C + V_\Sigma - M_\Sigma \; ; \quad T = C + V_\Sigma - M_\Sigma - r \; ; \quad A_N = V_\Sigma + A_n \; ; \quad A_R = V_\Sigma \tag{5.35}$$

Then applying the recursive equations (5.19-5.26) of Girard's model, we determine the probability p(A_N,A_R,T,K) when all K channels are busy, probability g(A_N,A_R,T,K) when less than T channels are busy, and the mean M_C(A_N,A_R,T,K) and variance V_C(A_N,A_R,T,K) of carried traffic in the K channels.

Assuming that the mean overflow from the aggregate hand-off traffic stream (M_Σ ,V_Σ) is the same as the mean overflow from its constituent streams (V_Σ,V_Σ) & (M_Σ-V_Σ , 0) and since only the Poisson component (V_Σ,V_Σ) and not the constant traffic component (M_Σ-V_Σ,0) produces overflow, the aggregate hand-off blocking p_{hb} experienced by the aggregate stream (M_Σ,V_Σ) is as follows:

$$p_{hb} = \frac{V_\Sigma \; p(A_N, A_R, T, K)}{M_\Sigma} \tag{5.36}$$

The new call blocking p_{nb} is the blocking experienced by the Poisson traffic (A_n,A_n) under channel reservation of level *r*. Therefore,

$$p_{nb} = 1 - g(A_N, A_R, T, K) \tag{5.37}$$

The individual hand-off offered traffic stream $[M_{ho}; V_{ho}]$ is assumed to experience the same blocking p_{hb} as the aggregate hand-off traffic stream $[M_\Sigma; V_\Sigma]$ since they both have the same peakedness (=variance/mean) [14].

We need to evaluate two additional parameters for the cell traffic blocking method, namely, the mean M_{car} and variance V_{car} of carried traffic in cell L. The mean M_{car} and variance V_{car} of carried traffic in a cell L of C servers is the same as the total mean and variance of carried traffic due to the three equivalent processes (A_n, A_n), (V_Σ, V_Σ) & $(M_\Sigma - V_\Sigma, 0)$. The mean $M_C(A_N, A_R, T, K)$ and variance $V_C(A_N, A_R, T, K)$ for the two Poisson streams (A_n, A_n) & (V_Σ, V_Σ) may be obtained from the application of Girard's method in the above cell traffic blocking method. The mean and variance of carried traffic due to the constant traffic stream is $M_\Sigma - V_\Sigma$ and 0 respectively since it is assumed to experience no loss. The constant traffic stream is uncorrelated with the two Poisson streams (A_n, A_n) & (V_Σ, V_Σ). This lack of correlation arises out of the manner in which we have decomposed the offered traffic streams into a constant traffic stream which permanently occupies one set of channels and two Poisson streams which compete for the remaining set of channels. Therefore, the total mean M_{car} and variance V_{car} of carried traffic in a cell L of C servers is:

$$M_{car} = (M_\Sigma - V_\Sigma) + M_C(A_N, A_R, T, K); \; V_{car} = V_C(A_N, A_R, T, K) \tag{5.38}$$

Although, Sanders et al derived their model under the assumption of negative exponential service time distributions, we shall show in the results section that the model is applicable for arbitrary service time distributions. Their model decomposes traffic processes into either constant traffic processes or Poisson traffic processes. The constant traffic process may be generated by a $D/D/\infty$ queue of appropriate mean. The constant traffic process, in a manner of speaking, plays a silent role in determining call congestion and time congestion probabilities. It is the Poisson traffic processes that play a dominant role in calculating these parameters. The reason for the Sanders et al's model's suitability, when moment matching various traffic streams under any type of service distributions, may be attributed to the fact that the Poisson arrival rates are insensitive to service distributions beyond the mean [59]. In the next section we present extensive results to validate our proposed cell traffic blocking models.

5.4 Results

5.4.1 Blocking probabilities for two cell scenario

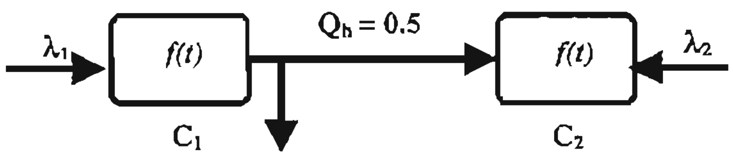


Fig 5-3: Two-cell scenario

We evaluated the accuracy of our proposed cell traffic blocking models using the simple scenario illustrated above. We assume that Poisson new calls with mean arrival rate λ_1 are offered to cell L_1 of C_1=6 channels. These calls are served in this cell for time intervals that are arbitrarily distributed with pdf $f(t)$ and with a normalised mean of 1 unit time. Thereafter, carried calls in the first cell are handed off to the second cell L_2 of C_2=8 channels with hand-off probability Q_h=1/2. For simplicity, these hand-offs are assumed to be simple Bernoulli trials. New calls are also assumed to be offered to cell L_2 with Poisson arrival rate λ_2. All calls in the second cell are served for time intervals that have the same pdf $f(t)$ and mean of 1unit time. To see the effect of different load conditions we varied the mean arrival rate λ_1 from 5 to 7 calls per unit time and set $\lambda_2 = \lambda_1 - 4$. We evaluated the first two moments of the fresh and hand-off traffic offered to the second cell using our cell traffic characterising model from the previous chapter. We then determined the blocking experienced by both the fresh and hand-off traffic streams in the second cell using our extensions to both Delbrouck's model and Sanders et al's model. We consider a blocked calls lost policy whereby all calls that arrive to find no free server available are assumed to be lost. For comparison purposes we obtained the same congestion parameters using an exact simulation and a simple single moment blocking model where all arrivals are assumed to be Poisson. In the single moment blocking model we make use of Akinpelu's model [44] to determine the various blocking values. In the next sections, we examine the different blocking values experienced for the case where $f(t)$ is negative exponential, det-neg and gamma distributed. We also consider the cases where channel reservation of level $r=0$ and $r=1$

are applied for the benefit of hand-off calls at the expense of new calls in the second cell. There is no channel reservation in the first cell as only fresh traffic is offered to it. The results obtained for the various scenarios are presented in graphical form. In these graphs, the horizontal axis represents the various arrival rates and the vertical axis represents the blocking probabilities.

5.4.2 Negative exponential channel holding times and channel reservation of level *r=0*.

We considered the scenario in figure 5-3 when *f(t)* is negative exponential and no channel reservation is applied to the second cell. Figure 5-4 presents the blocking experienced by the Poisson arrival process λ_2 in the second cell.

Figure 5-4: Call congestion experienced by the Poisson process

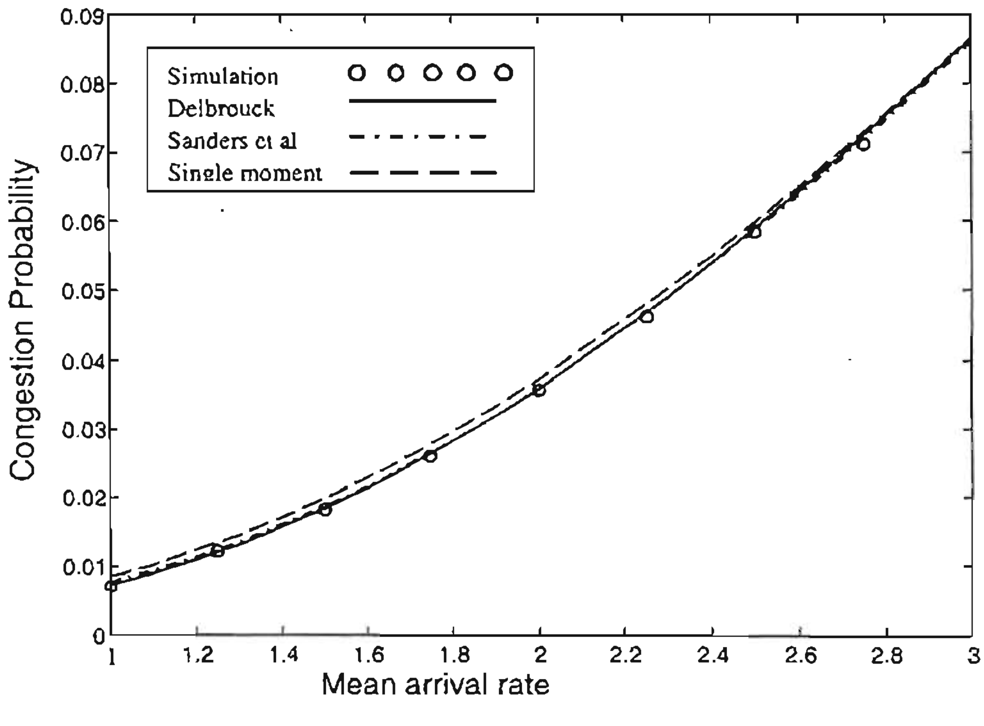
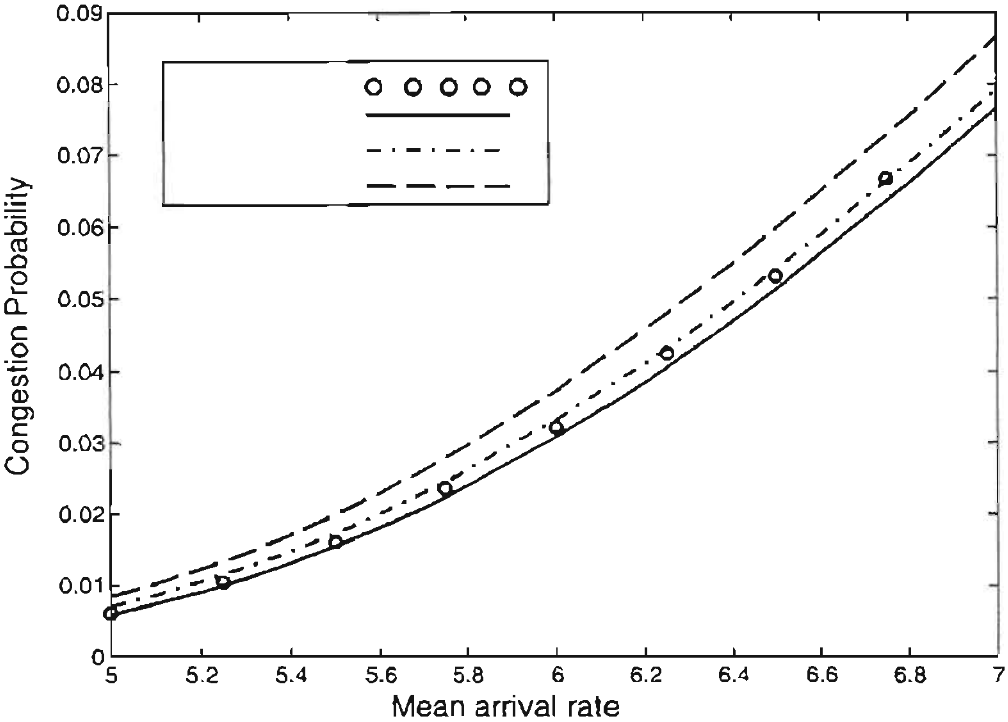


Figure 5-5 presents the blocking experienced by hand-off arrival process in the second cell. The horizontal axis in figure 5-5 is the original arrival rate λ_1 that gave rise to the hand-off process.

Figure 5-5: Call congestion experienced by the Hand-off process



As can be seen, for the congestion experienced by the Poisson stream, simulation and all three analytic methods tied up extremely well. For the congestion experienced by the hand-off arrival process, Sanders et al's model and Delbrouck's model performed better than the single moment model when compared to simulation results. Comparison of simulation values between figures 5-4 and 5-5 shows that hand-off blocking is less than new call blocking for a particular pair of arrival rates [λ_1 ; λ_2]. The reason for this may be attributed to the fact that hand-off traffic is smooth under negative exponential channel holding times and therefore experiences less blocking than the Poisson fresh traffic. Both the two moment blocking models have captured this difference. However, for the single moment model, there is no difference in the blocking values.

5.4.3 Negative exponential channel holding times and channel reservation of level *r=1*.

We considered the scenario in figure 5-3 when *f(t)* is negative exponential and channel reservation of level *r=1* is applied to the second cell. Figure 5-6 presents the blocking experienced by the Poisson arrival process λ_2 in the second cell. Figure 5-7 presents the blocking experienced by hand-off arrival process in the second cell.

Figure 5-6: Call congestion experienced by the Poisson process

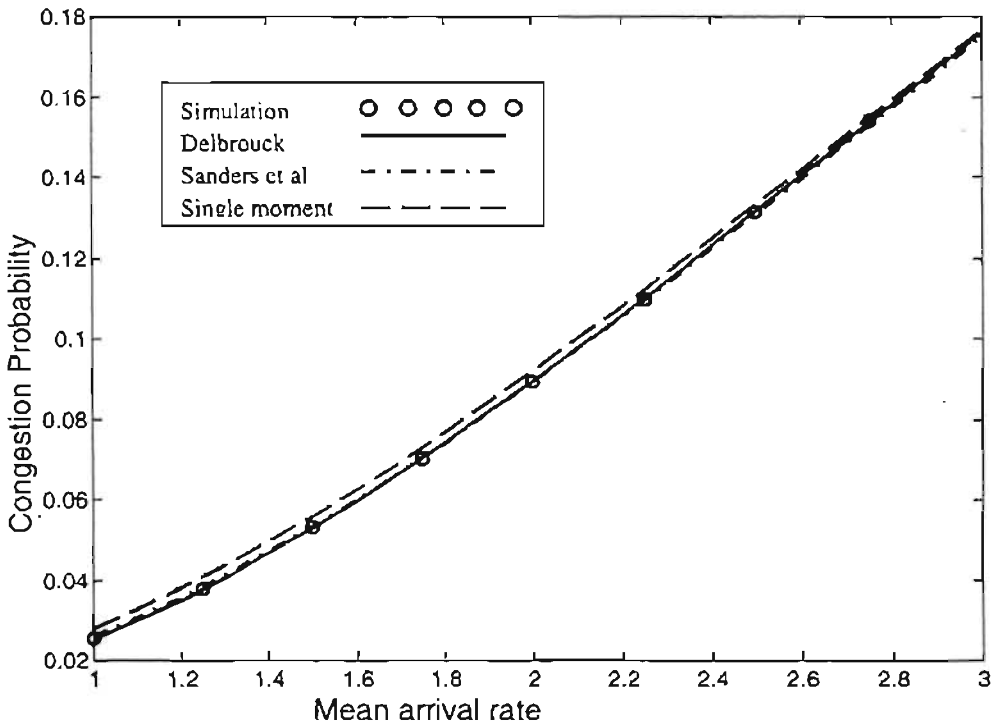
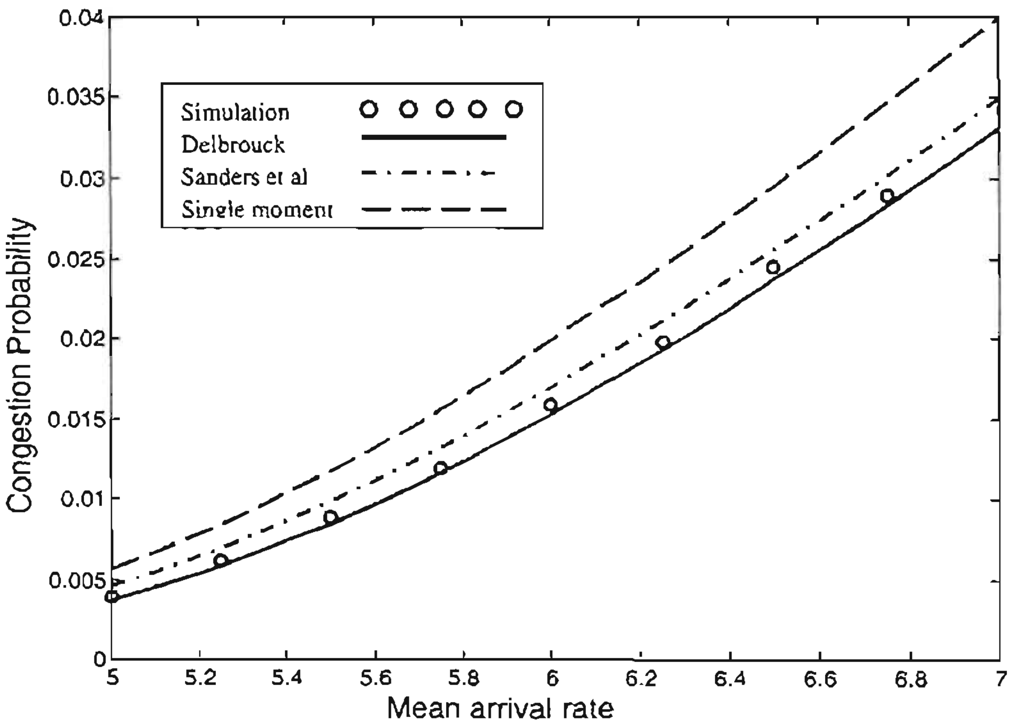


Figure 5-7: Call congestion experienced by the Hand-off process



As can be seen, for the congestion experienced by the Poisson stream, simulation and all three analytic methods tied up extremely well. For the congestion experienced by the hand-off arrival process, Sanders et al's model and Delbrouck's models performed better than the single moment model when compared to simulation results – the single moment method had a worst case error of 37% for underload conditions. Between the proposed two-moment blocking models, Delbrouck's model performed marginally better than Sanders et al's model for under load and nominal load conditions and Sanders et al's method performed marginally better for overload conditions. Sanders et al's method had a worst case error of 17% for underload conditions and Delbrouck's method had a worst case error of 5% for underload conditions. The effect of channel reservation is easy to see by comparing figures 5-4,5 with figures 5-6,7: channel reservation provides reduced hand-off blocking at the expense of increased new call blocking.

5.4.4 Det-neg channel holding times and channel reservation of level $r=0$.

We considered the scenario in figure 5-3 when $f(t)$ is det-neg with probability parameter $p = 0.6666$. No channel reservation is applied to the second cell. Figure 5-8 presents the blocking experienced by the Poisson arrival process λ_2 in the second cell.

Figure 5-8: Call congestion experienced by the Poisson process

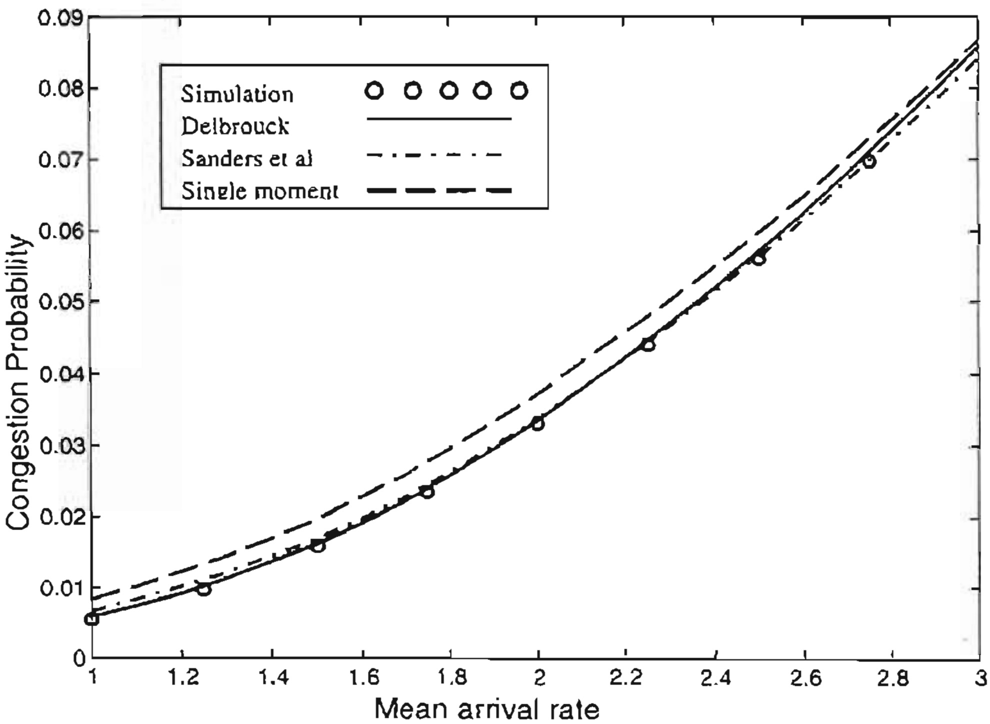
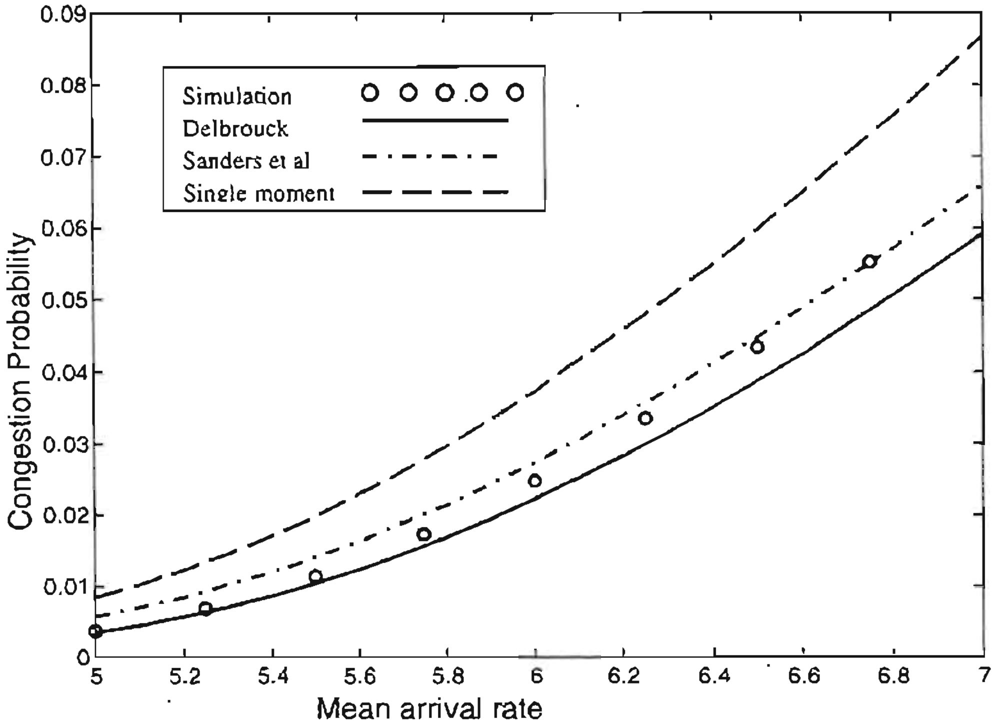


Figure 5-9 presents the blocking experienced by hand-off arrival process in the second cell.

Figure 5-9: Call congestion experienced by the Hand-off process



As can be seen, for the congestion experienced by the Poisson stream, both the proposed two moment models performed better than the single moment method - the worst case error was 51% for the single moment method. For the congestion experienced by the hand-off arrival process, Sanders et al's model and Delbrouck's models performed better than the single moment model when compared to simulation results – the worst case error was 121% for the single moment method. Between the two proposed models, Delbrouck's model performed better for underload and nominal load conditions and Sanders et al's model performed better for overload conditions. Sanders et al's method had a worst case error of 52% for underload conditions and Delbrouck's method had a worst case error of 4% for underload conditions. Comparison of simulation values between figures 5-8 and 5-9 shows that hand-off blocking is less than new call blocking for a particular pair of arrival rates [λ_1 ; λ_2]. The reason for this may be attributed to the fact that hand-off traffic is smooth under the given det-neg channel holding times and therefore experiences less blocking than the Poisson fresh traffic. Both the two moment blocking models have captured this difference. However, for the single moment model, there is no difference in the blocking values.

5.4.5 Det-neg channel holding times and channel reservation of level *r=1*.

We considered the scenario in figure 5-3 when *f(t)* is det-neg with probability parameter p=0.6666 and channel reservation of level *r=1* is applied to the second cell. Figure 5-10 presents the blocking experienced by the Poisson arrival process λ_2 in the second cell. Figure 5-11 presents the blocking experienced by hand-off arrival process in the second cell.

Figure 5-10: Call congestion experienced by the Poisson process

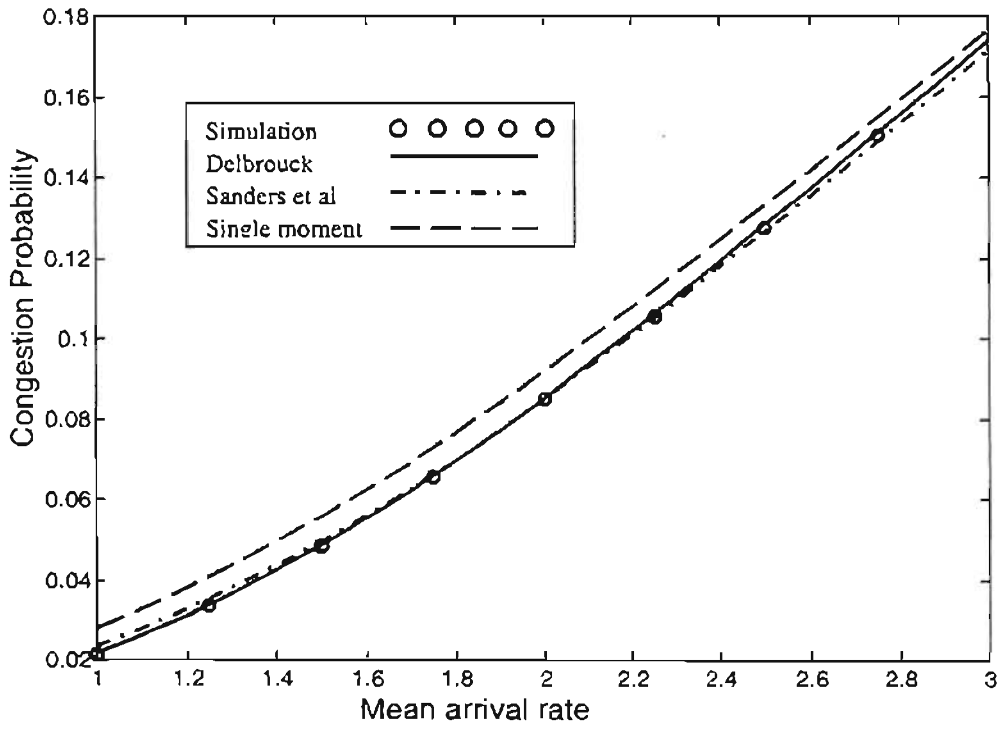
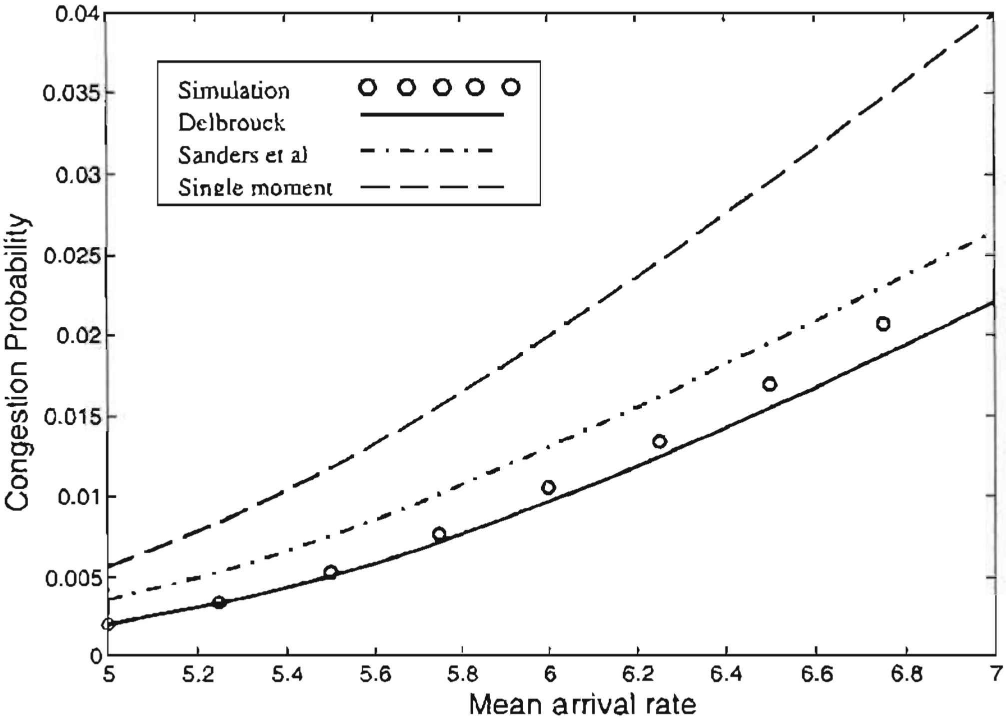


Figure 5-11: Call congestion experienced by the Hand-off process



As can be seen, for the congestion experienced by the Poisson stream, simulation and all three analytic methods tied up reasonably well –the single moment method had the worst case error (30%). For the congestion experienced by the hand-off arrival process, Sanders et al's model and Delbrouck's models performed significantly better than the single moment model when compared to simulation results – the single moment method had a worst case error of 180% for underload conditions. Between the proposed two moment blocking models, Delbrouck's model performed marginally better than Sanders et al's model for under load and nominal load conditions and Sanders et al's method performed marginally better for overload conditions. Sanders et al's method had a worst case error of 77% for underload conditions and Delbrouck's method had a worst case error of 6% for underload conditions. The effect of channel reservation is easy to see by comparing figures 5-8 & 5-9 with figures 5-10 & 5-11: channel reservation provides reduced hand-off blocking at the expense of increased new call blocking.

5.4.6 Gamma channel holding times and channel reservation of level $r=0$.

We considered the scenario in figure 5-3 when $f(t)$ is gamma distributed with gamma

Figure 5-12: Call congestion experienced by the Poisson process

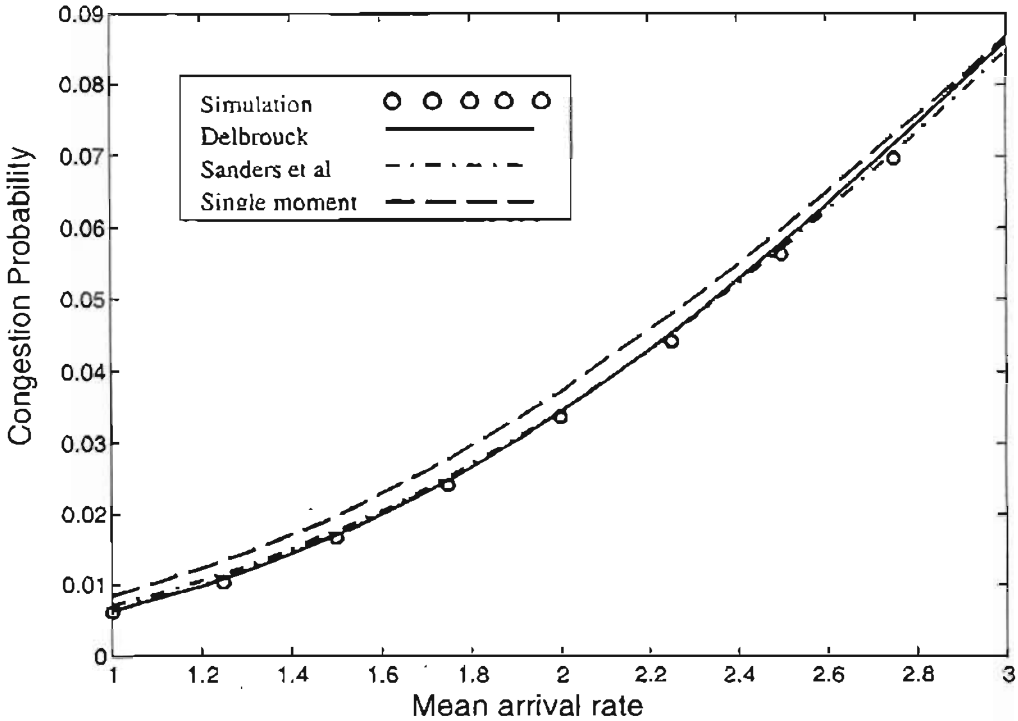
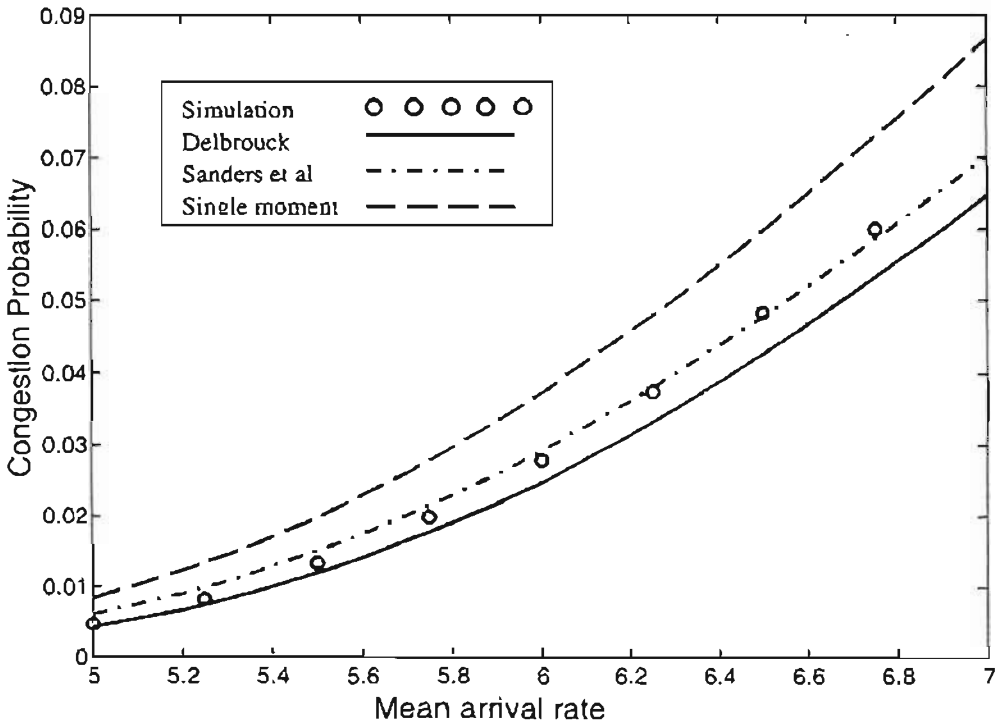


Figure 5-13: Call congestion experienced by the Hand-off process



shape parameter $c = 3$. No channel reservation is applied to the second cell. Figure 5-12 presents the blocking experienced by the Poisson arrival process λ_2 in the second cell. Figure 5-13 presents the blocking experienced by hand-off arrival process in the second cell.

As can be seen, for the congestion experienced by the Poisson stream, both the proposed two moment models performed better than the single moment method - the worst case error was 40% for the single moment method. For the congestion experienced by the hand-off arrival process, Sanders et al's model and Delbrouck's models performed better than the single moment model when compared to simulation results – the worst case error was 79% for the single moment method. Between the two proposed models, Delbrouck's model performed better for underload and nominal load conditions and Sanders et al's model performed better for overload conditions. Sanders et al's method had a worst case error of 30% for underload conditions and Delbrouck's method had a worst case error of 9% for underload conditions. Comparison of simulation values between figures 5-12 and 5-13 shows that hand-off blocking is less than new call blocking for a particular pair of arrival rates [λ_1 ; λ_2]. The reason for this may be attributed to the fact that hand-off traffic is smooth under the given gamma channel holding times and therefore experiences less blocking than the Poisson fresh traffic. Both the two moment blocking models have captured this difference. However, for the single moment model, there is no difference in the blocking values.

5.4.7 Gamma channel holding times and channel reservation of level $r=1$.

We considered the scenario in figure 5-3 when $f(t)$ is gamma with shape parameter $c=3$ and channel reservation of level $r=1$ is applied to the second cell. Figure 5-14 presents the blocking experienced by the Poisson arrival process λ_2 in the second cell. Figure 5-15 presents the blocking experienced by hand-off arrival process in the second cell.

As can be seen, for the congestion experienced by the Poisson stream, simulation and all three analytic methods tied up reasonably well –the single moment method had the worst case error (25%). For the congestion experienced by the hand-off arrival process, Sanders et al's model and Delbrouck's models performed significantly better than the single moment model when compared to simulation results – the single moment method

Figure 5-14: Call congestion experienced by the Poisson process

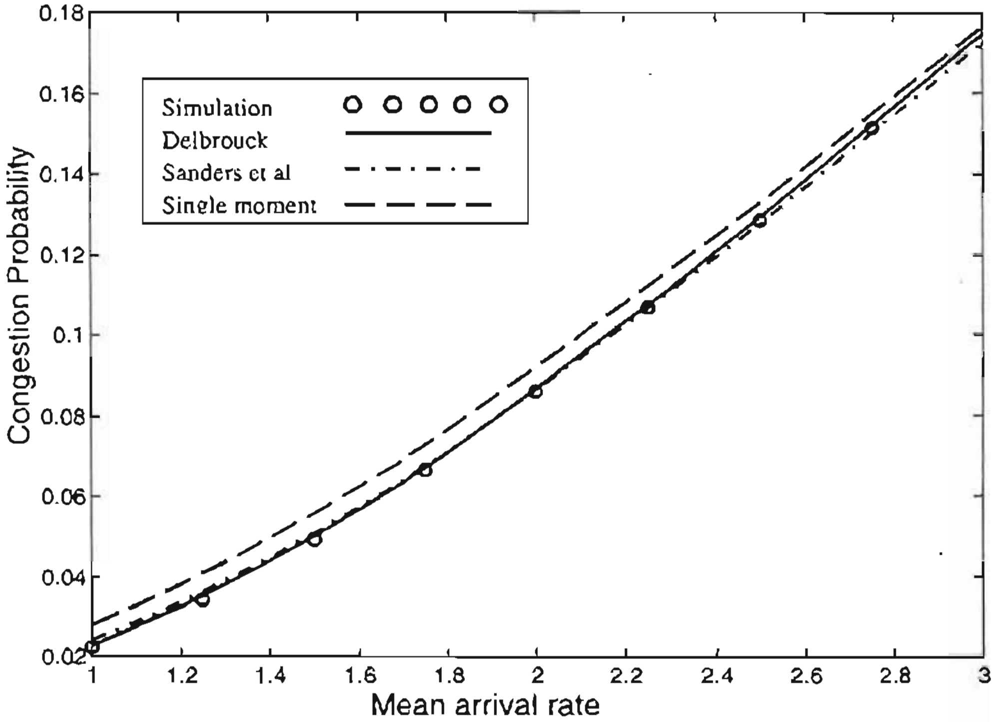
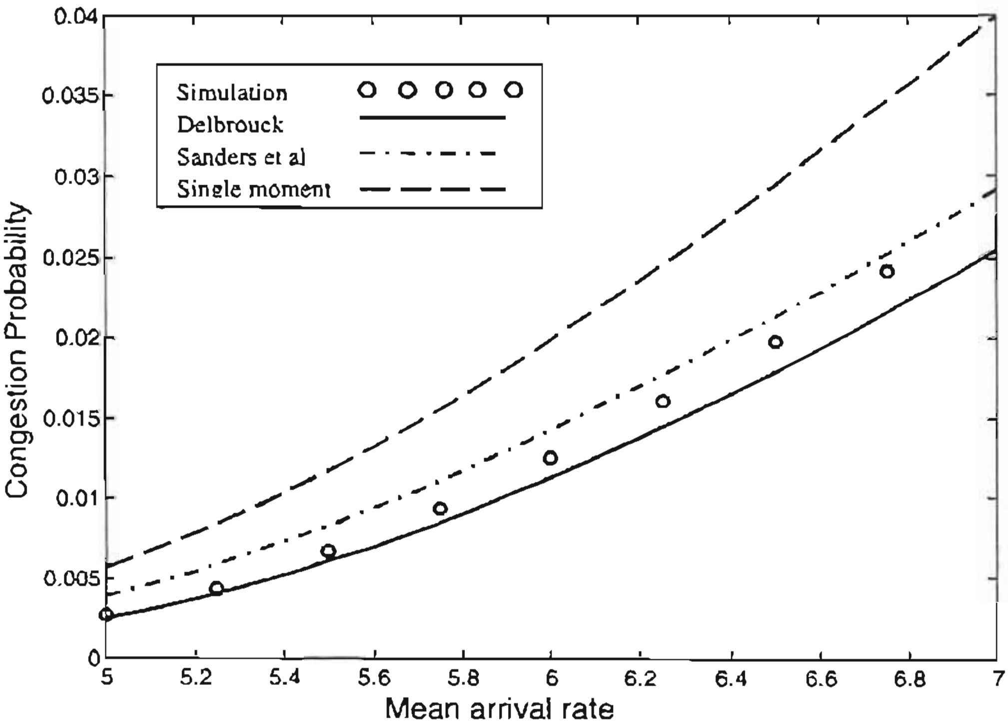


Figure 5-15: Call congestion experienced by the Hand-off process



had a worst case error of 112% for underload conditions. Between the proposed two moment blocking models, Delbrouck's model performed better than Sanders et al's model for under load and nominal load conditions and Sanders et al's method performed better for overload conditions. Sanders et al's method had a worst case error of 46% for underload conditions and Delbrouck's method had a worst case error of 6% for underload conditions. The effect of channel reservation is easy to see by comparing figures 5-12 & 5-13 with figures 5-14 & 5-15: channel reservation provides reduced hand-off blocking at the expense of increased new call blocking.

5.5 Summary

In this chapter we presented our proposed two-moment cell traffic blocking models for use in the performance analysis of cellular networks. The inputs to these cell traffic blocking models are the mean and variance of the various offered traffic streams. The outputs are the blocking probabilities experienced by the various traffic streams and the moments of the respective carried traffic streams.

We first examined two traffic-blocking models from the fixed network arena. They are

- Sander's et al's two moment blocking model
- Delbrouck's two moment blocking model

We extended the above models to apply for the cellular network scenario where two distinct traffic streams, namely, fresh and hand-off traffic, are offered to a cell and where channel reservation may be applied to protect hand-off traffic at the expense of fresh traffic.

We compared the performance of our extended models for a simple two-cell scenario and compared the results with those obtained from a single moment method and from simulation. Both our extended models outperformed the single moment method when compared to simulation results. Between the two proposed models, Delbroucks' model was found to be better for nominal and underload conditions whereas Sanders et al's model was found to be better for overload conditions. However, it must be mentioned that our extension to Delbrouck's model employs a two-dimensional Birth-Death model whereas our extension to Sanders et al's model employs a single dimensional Birth-

Death model, and this makes Delbrouck's model about 10-20 times slower than Sanders et al's model during computation.

CHAPTER 6

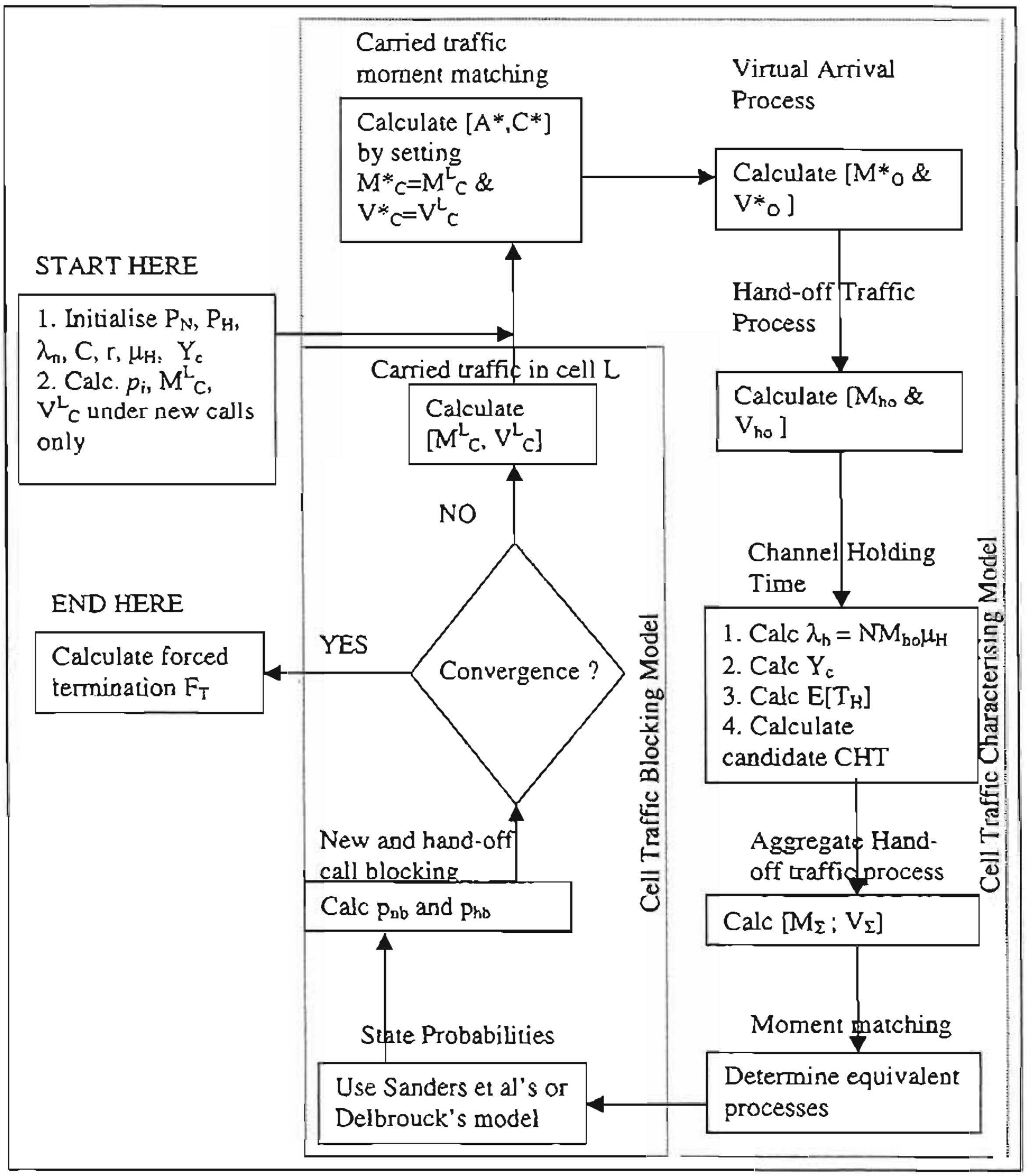
6.1 Introduction

In the previous two chapters, we discussed our proposed cell traffic characterising model and our proposed cell traffic-blocking model. In our cell traffic characterising model, we determined the mean and variance of the hand-off traffic by offered by a cell to its neighbour under various channel holding time distributions for a simple two-cell scenario. In our cell traffic-blocking model, we showed how blocking probabilities may be obtained when heterogeneous traffic streams are offered to a finite server. Both models, when used separately, are of little use in determining customer-orientated grade of service parameters in cellular networks. However, they may be modified and moulded together to determine customer-orientated grade of service parameters. It is this melding of the two models that we consider in detail in this chapter.

6.2 Quality of Service Evaluation

We show how the cell traffic characterising and cell traffic blocking models may be applied in a fixed point iteration to evaluate quality of service parameters such as fresh call blocking, hand-off call blocking and forced termination probability. Under the network homogeneous assumption it is sufficient to model a single cell. For ease of reference later, we label this cell as cell L. Figure 6-1 illustrates the various steps in this quality of service evaluation algorithm. The proposed algorithm is essentially an iterative procedure consisting of repetitive applications of the cell traffic characterising model and cell traffic blocking model.

Figure 6-1: Proposed Quality of Service Algorithm



6.2.1 Initialisation

At the initialisation of the algorithm, it is assumed that a mobility model for the subscribers of the cellular network is known. In chapter 2, we discussed mobility models from the existing literature and in chapter 7 we propose some of our own. For now we simply assume that a mobility model exists for the users being served by our network. We assume that our network satisfies the assumptions mentioned in section 2 of chapter 4. For brevity, we do not repeat these assumptions here. We assume that the mean new call arrival rate λ_n in cell L is known and that the unencumbered call holding time T_M is negative exponentially distributed with mean $1/\mu_M$. We also assume that the cell parameters such as channel size C and channel reservation parameter *r* are known.

Once the mobility model is derived, we assume that new call cell dwell time pdf $f_n(t)$ and hand-off call cell dwell time pdf $f_h(t)$ are known. Thereafter, applying equation (4.1), the new call hand-off probability P_N and hand-off call hand-off probability P_H may be determined.

The iteration illustrated in figure 6-1 is normally started by determining time congestion probabilities in a cell under the assumption that only fresh traffic is offered to it. The mean channel holding time may be obtained under the assumption that only new calls are offered to cell L. Therefore, equation (4.4) for the mean channel holding time $E[T_H]$ reduces to:

$$E[T_H] = \frac{1}{\mu_M}(1 - P_N) \qquad \textbf{(6.1)}$$

The mean offered traffic $A = \lambda_n \cdot E[T_H]$. Since new call arrivals are assumed to be Poisson, the Erlang-B model applies and equation (4.12) gives the blocking experienced by the new call arrivals. Equations (4.13) and (4.14) give the mean $M^L{}_C$ and the variance $V^L{}_C$ of carried traffic in cell L of C channels. We now examine the iterative portions of the quality of service algorithm. It is easier to start our explanation with the application of the cell traffic blocking model.

6.2.2 Application Of The Cell Traffic Blocking Model

We assume that cell L is offered new call arrivals with mean A_n where $A_n=\lambda_n/\mu_H$ and N independent hand-off traffic streams of mean M_{ho} and variance V_{ho} from its immediate neighbours. We assume that the mean and variance of the hand-off traffic streams are obtained from the previous iteration. Following the usual practice in fixed network analysis [14], we represent the N hand-off traffic streams by a single aggregate traffic stream of mean M_Σ and variance V_Σ where

$$M_\Sigma = N M_{ho} \; ; \; V_\Sigma = N V_{ho} \tag{6.2}$$

We may then use our extensions to Sanders et al.'s method or Delbrouck's BPP model as shown in chapter 5 to determine new call blocking p_{nb} and hand-off call blocking p_{hb} under channel reservation of level r. The mean carried traffic M^L_C and the variance of carried traffic V^L_C in cell L of C channels may also be obtained using these cell traffic blocking models. For brevity we will not examine these models in any great detail here since they have been considered in great detail in chapter 5. Suffice it to say they are of the following form:

$$p_{nb} = m(A_n, M_{ho}, V_{ho}, r) \tag{6.3}$$

$$p_{hb} = n(A_n, M_{ho}, V_{ho}, r) \tag{6.4}$$

$$M^L_C = p(A_n, M_{ho}, V_{ho}, r) \tag{6.5}$$

$$V^L_C = q(A_n, M_{ho}, V_{ho}, r) \tag{6.6}$$

where the probabilities p_{nb} and p_{hb} as well as the moments of carried traffic M^L_C and V^L_C are functions, $m(....)$, $n(....)$, $p(....)$,$q(....)$ of the new call and hand-off call offered traffic as well as the channel reservation parameter r.

6.2.3 Application Of The Cell Traffic Characterising Model

It is arguable that the exact Markovian analysis of a multi-cellular network scenario is very cumbersome. One would require a large multi-dimensional state space to properly describe the scenario, shown in figure 6-2, where a cell has new calls initiated within it, receives N incoming hand-off traffic streams and generates N outgoing hand-off traffic streams. For this reason, we suggest an approximate analysis of the multi-cellular scenario that makes use of the results of the simple two-cell scenario presented in chapter 4.

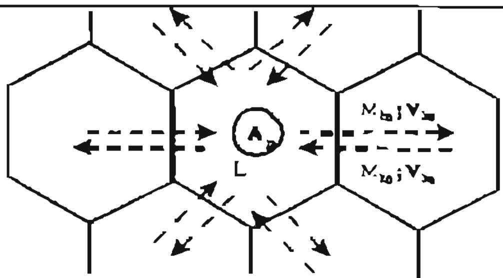


Figure 6-2: Cell L having many neighbours with in-coming and outgoing hand-off traffic

Before we discuss this approximate analysis, it is important to consider the following intrinsic relationship between the carried traffic within a cell and the hand-off offered traffic generated by the same cell. Clearly, a call has to be accepted or carried within a cell before it may be handed-off to another cell. Therefore, a call forms part of the carried traffic distribution before it forms part of the hand-off offered traffic distribution. With this argument in mind, we may employ carried traffic moment-matching to replace the cumbersome scenario where cell L is offered fresh-traffic and N incoming hand-off traffic with an equivalent scenario where only fresh traffic is offered to an equivalent cell L*. See figure 6-3.

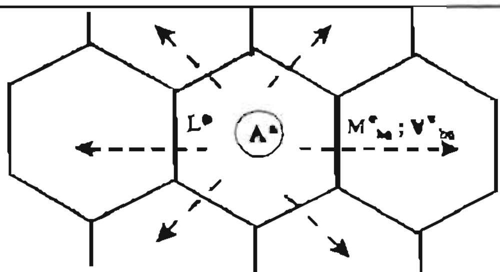


Figure 6-3: Cell L having many neighbours with only outgoing and no incoming hand-off traffic

We replace cell L of C channels being offered fresh and hand-off traffic with an equivalent cell L* of C* channels and being offered fresh traffic of mean A* *such that* the mean M^*_C and variance V^*_C of carried traffic in the equivalent cell L* is the same as the actual mean M^L_C and variance V^L_C of carried traffic in the original cell L. We choose the parameters [A*;C*] of the equivalent cell such that the mean and variance of carried traffic are the same for the equivalent cell and the original cell L. We may then determine the mean and variance of hand-off traffic offered by the equivalent cell to a neighbouring cell *j* and take them to be the mean and variance of hand-off traffic offered by the original cell L to the same neighbour *j*. Using Little's result [59], it is not difficult to show that the mean hand-off traffic generated by the equivalent cell L* to a neighbour *j* is exactly the same as that which the original cell L would have offered to neighbour *j* since the mean carried traffic are the same in both cells L and L*. It is only the variance of hand-off traffic offered by the equivalent cell that is necessarily an approximation for the variance of hand-off traffic offered by the original cell L. However, such an approximation is not far-fetched considering that we have implemented carried traffic moment-matching between the two systems and that carried traffic within a cell is that which gives rise to out-going hand-off traffic from the same cell. The above argument of replacing the original cell L by an equivalent cell L* seems more logical for the case when there is no channel reservation in the original cell L. However, as we shall show in the results section, the above approximation works well even when channel reservation is employed in the original cell L. It is important to understand that there is no question of channel reservation in the equivalent cell L* since it has only fresh traffic A* offered to it. Note that we assume that the various holding times in the equivalent cell are identical to those in the original cell L. Note also that the equivalent cell L* from this chapter and the virtual cell L_V of chapter 4 are two distinct mathematical tools with two distinct purposes.

Since the equivalent cell L* has only fresh (Poisson) traffic offered to it, the Erlang-B model is used to analyse it. One may determine the equivalent cell parameters A* and C* by substituting M^*_C and V^*_C into equation (4.13-4.14) and solving for the Erlang parameters A* and C*. Then using the results of the cell traffic characterising model in chapter 4, the mean M^*_{ho} and variance V^*_{ho} of hand-off traffic offered by the equivalent cell to a neighbouring cell *j* may be determined by substituting A by A*, λ by λ^* (NB. $\lambda^* = A^*.\ \mu_H$), C by C*. Consequently the mean M_{ho} and variance V_{ho} of hand-off traffic offered by the original cell L to neighbour *j* is:

$$M_{ho} = M^*_{ho} = Q^*_{h} \cdot M^*_{O} \tag{6.7}$$

$$V_{ho} = V^*_{ho} = M^*_{ho} \left(1 + Q^*_{h} \cdot \left(\frac{V^*_{O}}{M^*_{O}} - 1 \right) \right) \tag{6.8}$$

where the mean M^*_O and variance V^*_O of the virtual arrival traffic generated by the equivalent cell can be obtained from chapter 4. Probability Q^*_h is defined in terms of the various parameters of cell L as follows:

$$Q^*_h = R_j \left(\frac{p_N}{1 + \Upsilon_c} + \frac{p_H \Upsilon_c}{1 + \Upsilon_c} \right) \quad (6.9)$$

The term in the bracket is the weighted probability with which a call, be it a new call or a previously handed-off call, would require a further hand-off from cell L, and R_j is the probability that cell j receives that hand-off traffic. Note that given the total mean hand-off traffic offered to a cell is $N.M_{ho}$, the total mean hand-off arrival rate into a cell, λ_h, may be calculated by the relationship $\lambda_h = N.M_{ho}.\mu_H'$ where $1/\mu_H'$ is the mean channel holding time calculated from the previous iteration.

6.2.4 Channel Holding Time Distributions

In order to calculate the mean and variance of the virtual arrival traffic offered by the equivalent cell, it is necessary that the channel holding time distribution be modelled accurately. The analysis of our two cell scenario shown in chapter 4 under arbitrary channel holding time distributions is very difficult. Consequently we proposed two candidate distributions, the det-neg distribution and the gamma distribution to model the actual channel holding time distribution and it is for these candidate distributions that we analysed the two-cell scenario.

The explicit derivation of the cell dwell time distributions and thereafter the actual channel holding time distribution, for a given mobility model, is possible using the methodology presented in the appendix of [15]. For the sake of conciseness we will not elaborate on this methodology nor present the results here since we use this methodology to derive a suitable mobility model for highway networks in the next chapter. The resulting channel holding time cdf $F_{CH}(t)$ is generally a rather complicated expression and one that is not easy to employ it in its exact form during the tele-traffic performance analysis of a cellular network. In classical analysis, the resulting channel holding time distribution $F_{CH}(t)$ would most likely be represented by a suitable member of the negative exponential distribution family. In our analysis, we prefer to represent the actual channel holding time distribution with a suitable member from our candidate distributions, the det-neg and the gamma distributions.

Firstly, we set the mean of our candidate channel holding time distribution to be the same as the mean of the actual channel holding time distribution as derived from equation (4.4). Since $E[T_H]$ is a function of Y_c, which in turn is a function of λ_h, and since λ_h changes at each iteration, it is necessary that Y_c and consequently the mean channel holding time $E[T_H]=1/\mu_H$ be recalculated at each iteration. Thereafter, for the det-neg distribution we choose the appropriate det-neg probability parameter *p* and for the gamma distribution we choose the appropriate shape parameter *c* such that the following expression is minimised:

$$G = \frac{\int_0^\infty \left| F_{CH}(t) - F_{CAND}(t) \right| dt}{2 \int_0^\infty F_{CH}(t)\, dt} \qquad \textbf{(6.10)}$$

where $F_{CAND}(t)$ is the cdf of the candidate (det-neg or gamma) distribution.

Once the mean and variance of the hand-off traffic streams, and the candidate channel holding time distributions are determined, these values are plugged into the cell traffic blocking method to obtain a new set of blocking probabilities. Repetitive use is made of the cell traffic blocking and traffic characterising methods until convergence is obtained. Once convergence is obtained, the forced termination probability, F_T, which is the probability with which a call is dropped by an unsuccessful handover after being admitted into the network, may be calculated as:

$$F_T = \sum_{i=0}^{\infty} Prob\{call\ dropping\ on\ (i+1)^{th}\ handover\} \cdot Prob\{i\ successful\ handovers\} \qquad \textbf{(6.11)}$$

$$F_T = \sum_{i=0}^{\infty} P_N\, p_{hb}\, \{ P_H (1 - p_{hb}) \}^i = \frac{P_N\, p_{hb}}{1 - P_H (1 - p_{hb})} \qquad \textbf{(6.12)}$$

Finally, note that we have not analytically investigated the question of unique solutions for the equivalent parameters A^* and C^* when inverting equations (4.13-4.14) given the mean M^*_C and variance V^*_C of carried traffic in the equivalent cell L^*. However, for the numerous cases attempted by a practical iterative algorithm we have always found a unique solution. Furthermore, during the solution of equation (4.13-4.14) for the equivalent parameters A^* and C^* we have generally found non-integral solutions for

C^*. Since the Erlang-B formula has been extended to non-integral values of channel size C^* [14], there is no problem in calculating the mean offered traffic M^*_O. Unfortunately, the same cannot be said for equations (4.33-4.35) for calculating the variance V^*_O under negative exponential channel holding times. In the absence of proper extensions to the non-integral values, we suggest the following alternative for calculating V^*_O : Let FLOOR[C*] and CEILING[C*] be the two closest integers below and above the non-integral number C^*. Let V^*_{FLOOR} and $V^*_{CEILING}$ be the two variance values calculated using offered traffic A^* and channel sizes FLOOR[C*] and CEILING[C*] in equations (4.33-4.38). Then an approximation for V^*_O may be obtained using linear interpolation between V^*_{FLOOR} and $V^*_{CEILING}$:

$$V^*_O = V^*_{FLOOR} + (C^* - FLOOR[C^*])(V^*_{CEILING} - V^*_{FLOOR}) \tag{6.13}$$

Finally, a requirement of the cell traffic blocking method by Sanders et al is that the number of channels $C - r > M_\Sigma - V_\Sigma$. We found this condition to be met in the numerous networks that we analysed. Furthermore Sanders et al [67, equation (19)] provide a simple extension that is applicable when the above condition is not met. The more elegant method by Delbrouck does not have such limitations.

6.3 Results

In the results section of chapter 4 we considered the performance of our proposed cell traffic characterising model. In the results section of chapter 5 we considered the performance of our proposed cell traffic blocking models. In this section we consider the performance of the entire performance analysis algorithm for cellular networks. We consider the performance of simple ring networks under simple mobility models. We leave the analysis of more realistic mobility models and more realistic cellular network layouts for the following chapter where we consider mobility modelling in greater detail. The results section is composed of three parts. In the first part we examine the performance analysis of cellular networks under negative exponential channel holding times. In the next section we examine the performance analysis of cellular networks under det-neg and gamma channel holding times. In the last section we consider the nature of the convergence of our performance analysis algorithm. Finally it must be mentioned that although we have a choice of two possible cell traffic blocking models, Sanders et al's model and Delbrouck's model, for use in our performance analysis algorithm, we opt for the latter in the analysis to follow. The main reason for such a choice is to maintain consistency. Furthermore, as shown in chapter 5, Delbrouck's model performs marginally better.

6.3.1 Performance analysis under negative exponential channel holding times

Using sample network scenarios, we compare the performance of our two moment analysis with existing single moment analysis and simulation results. In our analysis we use the simple negative exponential distribution as the candidate channel holding time distribution. In our simulator, we used some of the mobility models proposed in the literature. Existing mobility models may be separated into two groups based on whether or not they make an explicit distinction between the cell dwell times of new calls and previously handed-off calls.

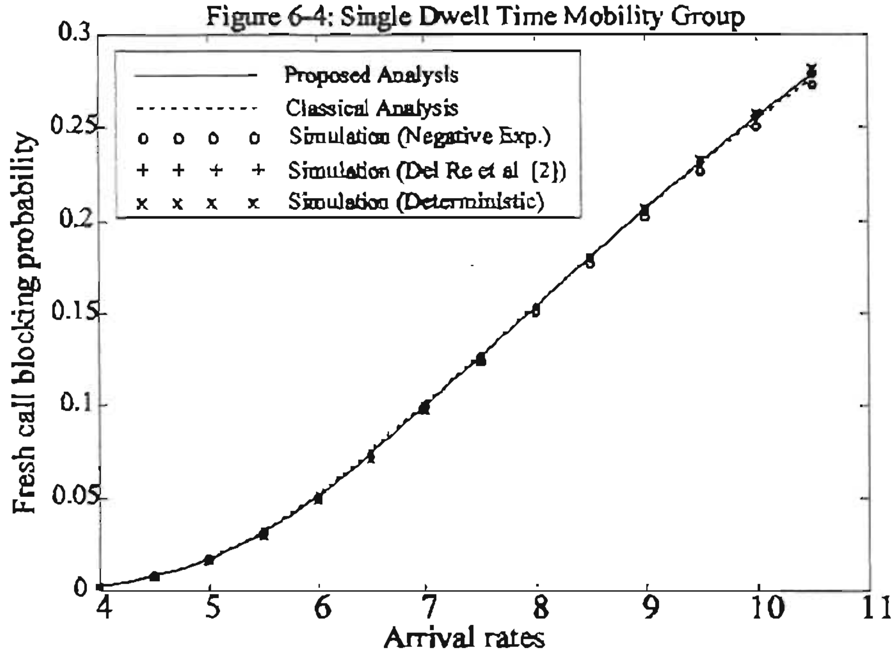
6.3.1.1 Single Dwell Time Mobility Group

In this section we compare results between the proposed analysis, existing analysis and simulation results employing mobility models that do not make an explicit distinction between the cell dwell times of new calls and previously handed-off calls. Similar to Sidi and Starobinski [21], we employ a homogeneous cellular network that consists of a ring of 12 concatenated cells. A ring network is employed to avoid edge effects. In each cell we set C= 20 and r=3 channels. We assume $1/\mu_M$=2minutes. Since the mobility models from the Single Dwell Time group make no distinction between new calls and previously handed-off calls, we set the probability of hand-off for both types of calls to be 2/3. Therefore, the mean channel holding time $1/\mu_H = E[T_H]$= 40 seconds (equation 4.4) for all calls. For simplicity sake, we assume that all the mobile users are moving in the same direction. Therefore, calls are only handed-off in one direction around the ring of cells. We increased the new call arrival rate per cell from 4 to 11 calls/minute to see the effect of different offered load conditions. We determined the various Quality-Of-Service parameters using the proposed method, existing single moment method and simulation. In our simulator we used the following three mobility models. To be able to compare results among the three mobility models listed below, we adjusted the parameters in each mobility model such that network parameters such as mean channel holding times, hand-off probability etc. are identical in all the mobility models:

a)Mobility model of Del Re et al [2]: In this mobility model, users travel a distance that is uniformly distributed between 0 and 2D where D is a cell side. They have a constant velocity during their travel within a cell and their velocity is uniformly distributed between 0 and V_{max}. Del Re et al show that the probability of hand-off is a function of parameter α where $\alpha = 2\, D\, \mu_M / V_{max}$. We chose the parameter $\alpha \approx 0.30$ such that the probability of hand-off (equation (2.18)) is 2/3.

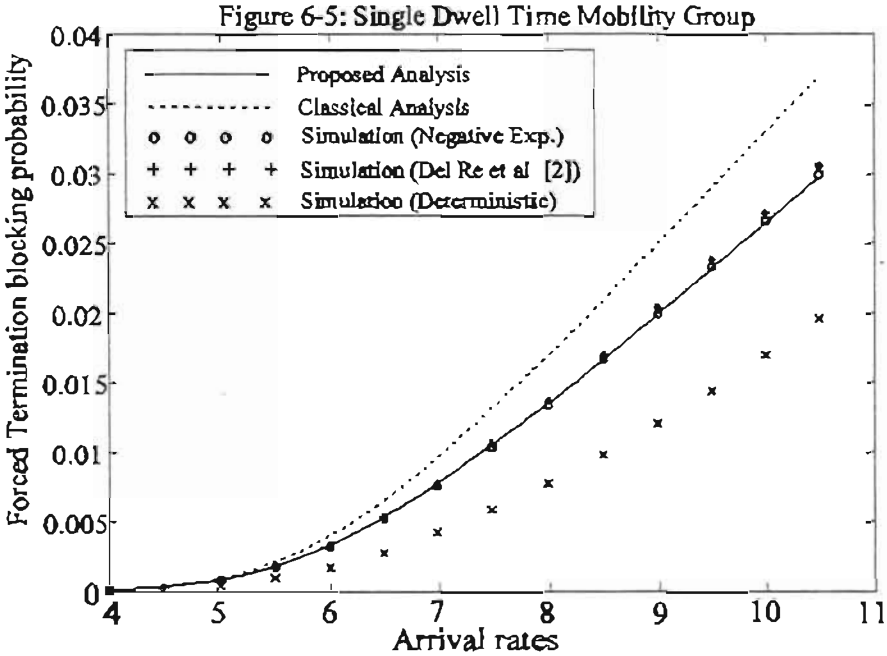
b) Deterministic mobility model: We also consider a mobility model where the cell dwell times are deterministic. Assuming a negative exponentially distributed unencumbered call holding time, it is simple to show that the probability of hand-off for a deterministic cell dwell time $P_H = P_N = e^{-\mu_M . d}$ where d is the deterministic cell dwell time. We choose d=0.810930 minutes such that $P_H = P_N = 2/3$. One may argue that deterministic cell dwell times are unrealistic, but they are useful for examining a worst case scenario. Furthermore, high-way micro/pico cellular networks and LEO satellite systems have dwell times that are better approximated by deterministic cell dwell times than negative exponential dwell times. For example, in a high-way micro-cellular

Figure 6-4: Single Dwell Time Mobility Group



network consisting of concatenated cells, users moving from cell boundary to cell boundary will travel a fixed distance D and their velocity will be distributed within a very narrow range. In such a case their dwell times tend towards the deterministic.

c) Negative exponential mobility model: For further comparison purposes, we also simulated the scenario where dwell times are negative exponentially distributed with mean of 1 minute.



Figures 6-4 and 6-5 list the various results obtained for the above network scenario. As can be seen from figure 6-4, for new-call blocking, the proposed analysis and the existing analysis tie up well with the three simulation results. However, figure 6-5 shows that for forced termination probability there are much larger differences. As can be seen from figure 6-5, the proposed analysis is better than the existing single moment analysis when compared to all three simulation scenarios. However it appears to be much better for Del Re et al's mobility model and the negative exponential mobility model than the worst case scenario where the dwell times are deterministic. The following reasons may be given: The three simulations and the two analytical models tie up well for new call blocking values because in all five cases new call arrivals are modelled as Poisson. Larger differences are seen for forced termination probability values between the proposed method and the existing method. The reason for this is that the existing method assumes that hand-off call arrival is Poisson, whereas the proposed analysis attempts to capture the actual attributes of hand-off traffic using a two-moment representation. In the first part of the results section of chapter 4, we empirically showed that hand-off traffic is smooth under negative exponential channel holding times. This is the reason why the single moment method provided conservative estimates for forced-termination blocking probabilities compared to the proposed two-moment analysis.

The reason why the proposed model was much better for Del Re et al's mobility model as well as the negative exponential mobility model than for the deterministic mobility model may be explained as follows: Given an arbitrary *channel holding time* pdf $f_H(t)$, let us define a parameter Ξ which we call Erlang's number for that pdf. Erlang's number is defined to be the ratio of the square of the mean to the variance of the given pdf. The reason why we call this parameter Erlang's number is that, for integral values, it is the number of stages that would be required by an Erlang-k pdf to achieve the same ratio of mean squared over variance [59]. Table I lists the Erlang's number Ξ for the three different mobility models:

Table I: Erlang's Number for various channel holding time distributions

Mobility Model	Erlang's number
Del Re et al. [2]	0.7
Negative Exponential	1
Deterministic mobility model	7.5

The Negative exponential mobility model has Erlang's number of 1. Since the proposed model assumes negative exponential channel holding times in its analysis, it is not hard to understand why this is the mobility model that has the best agreement with our proposed model. The channel holding time from Del Re et al's mobility model has Erlang's number of 0.7. This implies that the channel holding time is more hyper-exponential in nature and less erlang-k in nature. However since the Erlang's number for Del Re et al's mobility model is not too removed from 1 (which is the Erlang's number for the Negative Exponential distribution), the proposed method works well for Del Re et al's mobility model. Note that although Erlang's number is not on a linear scale, it does give a relative feel as to how close or removed a given channel holding time pdf is to the negative exponential pdf. Also note that the actual range for the Erlang's number ranges from 0 (for a channel holding time that has unbounded variance) to ∞ (for a deterministic channel holding time).

The channel holding times in the deterministic mobility model have Erlang's number of approximately 7.5. This implies that an erlang-7 or erlang-8 distribution would be better suited to model this channel holding time pdf. Since we employ the negative exponential pdf in our analysis, there is much larger discrepancy between the proposed model and this particular simulation scenario. However the proposed model is still a better approximation for this simulation scenario than the existing single moment method. The deterministic dwell time is a good example of a scenario where the negative exponential distribution fails as a candidate channel holding time distribution and one where other candidate distributions such as the gamma and det-neg may be more suitable (as we show later).

6.3.1.2 Double Dwell Time Mobility Group

In this section we compare results between the proposed analysis, existing analysis and simulation results employing mobility models that make an explicit distinction between the cell dwell times of new calls and previously handed-off calls. In our simulator we employed the mobility model of Zonoozi and Dassanayake [18]. We analysed a ring of 12 concatenated cells where C=50 channels and channel reservation of r=3 channels. We also considered the case where C=50 channels and r=2 channels. We assume $1/\mu_M$= 2 minutes. We increased the new call arrival rates per cell from 15 to 30 calls per minute to see the effect of different offered load conditions.

Mobility model of Zonoozi and Dassanayake [18]: Based on their simulations, these authors suggest the use of two generalised gamma distributions (equation (2.33)) as cell dwell times of new and previously handed-off calls. We determined the parameters of the generalised gamma distribution as suggested by them (equation (2.33-2.35)) where we set the radius of our cell to be 0.7 km. Figures 6-6 and 6-7 list the various results obtained for the above network scenario.

Figure 6-6: Double Dwell Time Mobility Group

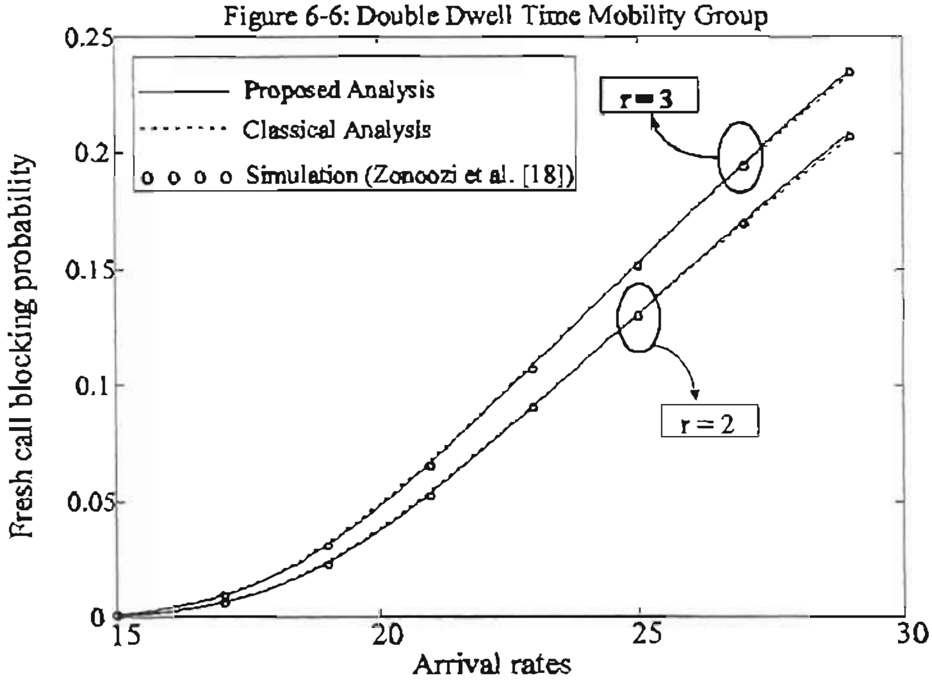
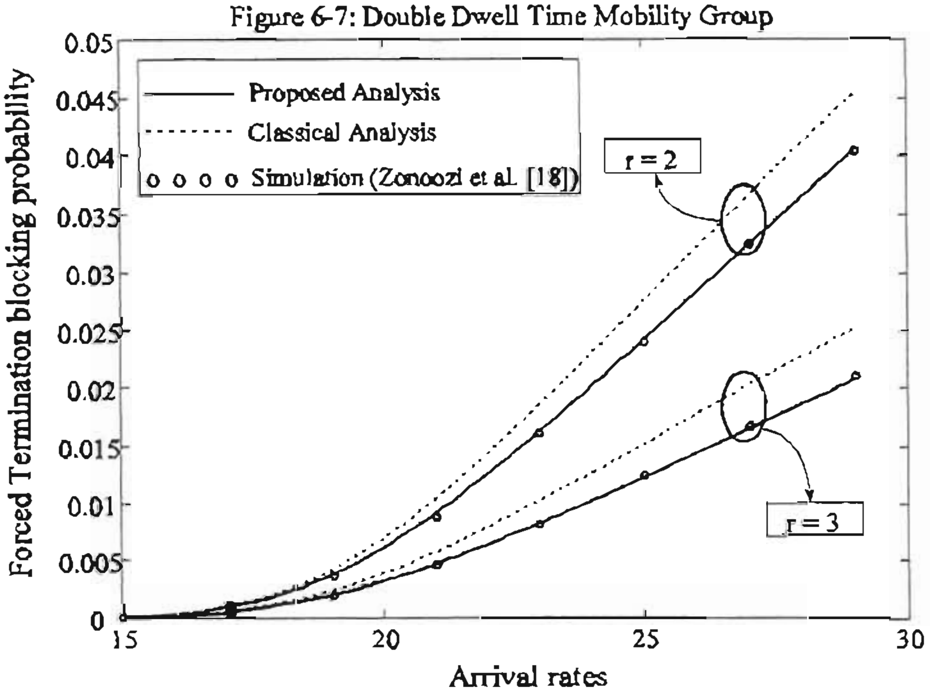


Figure 6-7: Double Dwell Time Mobility Group



As can be seen from figure 6-6, for new-call blocking, the proposed analysis and the existing analysis tie up well with the simulation results for both channel reservation values. However, figure 6-7 shows that for forced termination probability there are much larger differences. As can be seen from figure 6-7, the proposed analysis is better than the existing single moment analysis when compared to the simulation results for both channel reservation scenarios. The existing analytical method's results were conservative compared to the proposed two moment method's results. The reason for all this is that, unlike the existing method, the proposed method captures the true nature of hand-off traffic using a two-moment representation of hand-off traffic. Furthermore, Zonoozi and Dassanayake [18] indicate that the channel holding time as determined by their mobility model is well approximated by the negative exponential distribution. We have found this to be the case for large cell sizes. However as the cell size is reduced, we found the channel holding times to have Erlang's number, Ξ, where Ξ is larger than 1. This implies that, for small cell sizes, the channel holding times of the above mobility model are less negative exponential and more Erlang-k (gamma) in nature. This is to be expected since the dwell times are gamma distributed.

6.3.2 Performance analysis under det-neg and gamma candidate channel holding times

So far we have considered the use of the simple negative exponential distribution as a suitable candidate for modelling the actual channel holding time distributions. However, as shown in the previous results the neg-exp distribution is not always the best choice for modelling realistic channel holding time distributions. It is for this reason that we proposed the det-neg and gamma pdf as alternate candidates for modelling the channel holding time distribution. We now determine the grade-of-service parameters, new call blocking and forced termination probability, for a simple ring network under channel holding time distributions that are exactly det-neg distributed and gamma distributed.

In this section we show that the proposed cell traffic characterising model works well with our proposed cell traffic blocking model in our Quality of Service Evaluation algorithm under controlled conditions where the channel holding time distributions are either exactly det-neg or exactly gamma distributed. In practical networks it is very unlikely that the channel holding times are exactly det-neg or gamma distributed. But the use of these channel holding time distributions give an indication as to how well our entire performance analysis algorithm gels together and performs under controlled conditions. Such an approach can illustrate limitations, if any, in the approximations made in the cell traffic characterising, cell traffic blocking method and the quality of service evaluation method.

We consider two similar network scenarios for the det-neg and gamma channel holding time distributions. Similar to Sidi and Starobinski [21], we employ a homogeneous cellular network that consists of a ring of 12 concatenated cells. A ring network is employed to avoid edge effects. In each cell we set channel size $C= 8$. We set channel reservation $r=1$ channel in each cell. We assume that both new calls and previously handed-off calls have the same channel holding time distribution (single dwell time mobility group). For simplicity sake, we set the probability of hand-off for both new and hand-off calls to be exactly 2/3 (assuming simple Bernoulli trials). We also assume that all the mobile users are moving in the same direction. Therefore, calls are only handed-off in one direction around the ring of cells. We increased the new call arrival rate per cell from 1 to 3 calls per minute to see the effect of different offered load conditions.

In scenario #1, we assumed that the channel holding times are exactly det-neg distributed with mean of 40 seconds but having different det-neg probability parameters p where $p \in \{0, 5/9, 9/10\}$. In scenario #2, we assumed that the channel holding time distributions in each cell are exactly gamma distributed with mean of 40 seconds but having different gamma shape parameters c where $c \in \{1, 3, 15\}$.

We analysed the scenarios #1 and #2 using our proposed performance analysis method under det-neg channel holding times and gamma channel holding times respectively. We also analysed both scenarios using exact simulation runs as well as classical single moment analysis [15,16]. We have plotted these results in figures 6-8 to 6-11.

Figure 6-8: Grade of Service Evaluation under det-neg channel holding times

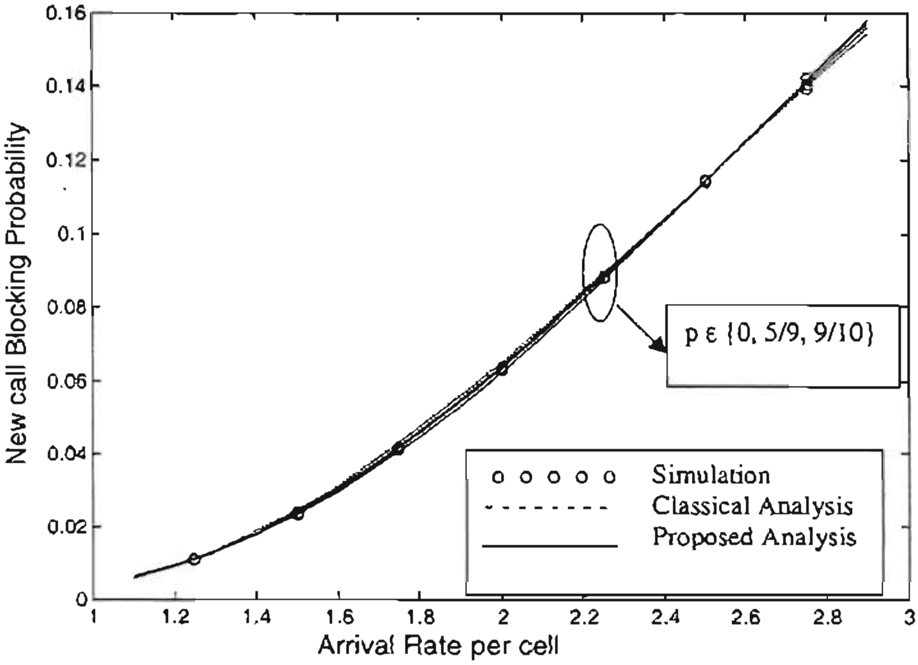


Figure 6-9: Grade of Service Evaluation under det-neg channel holding times

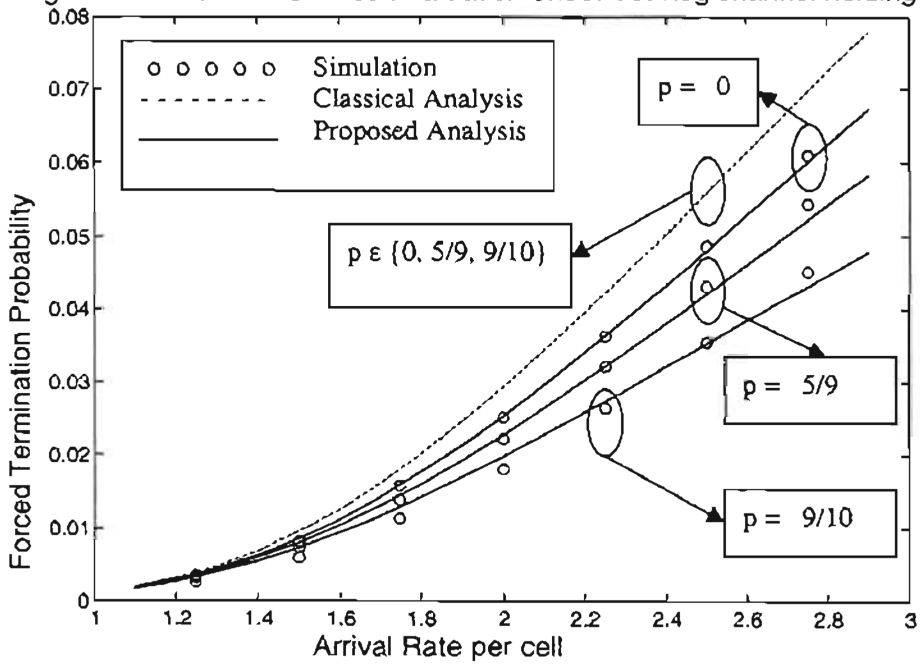


Figure 6-10: Grade of Service Evaluation under gamma channel holding times

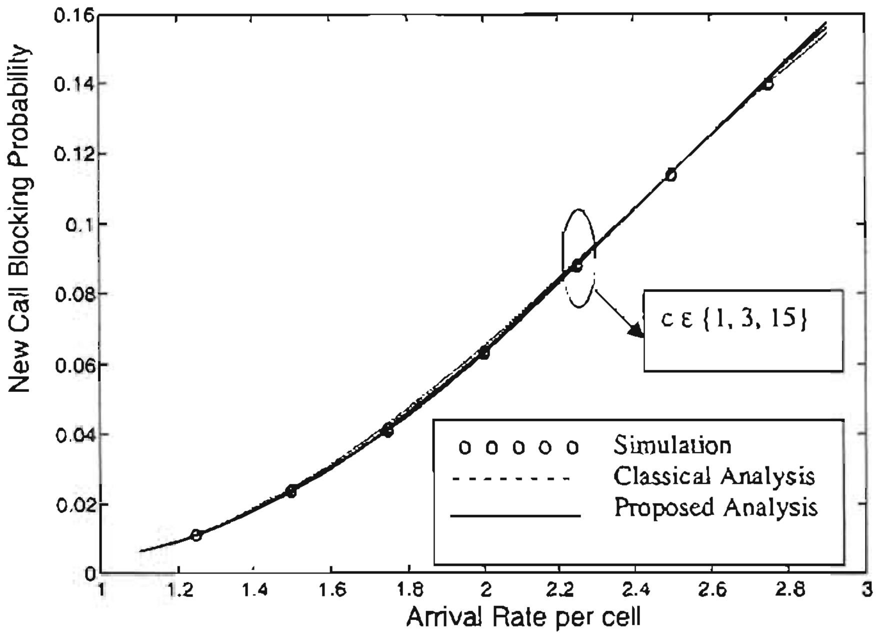
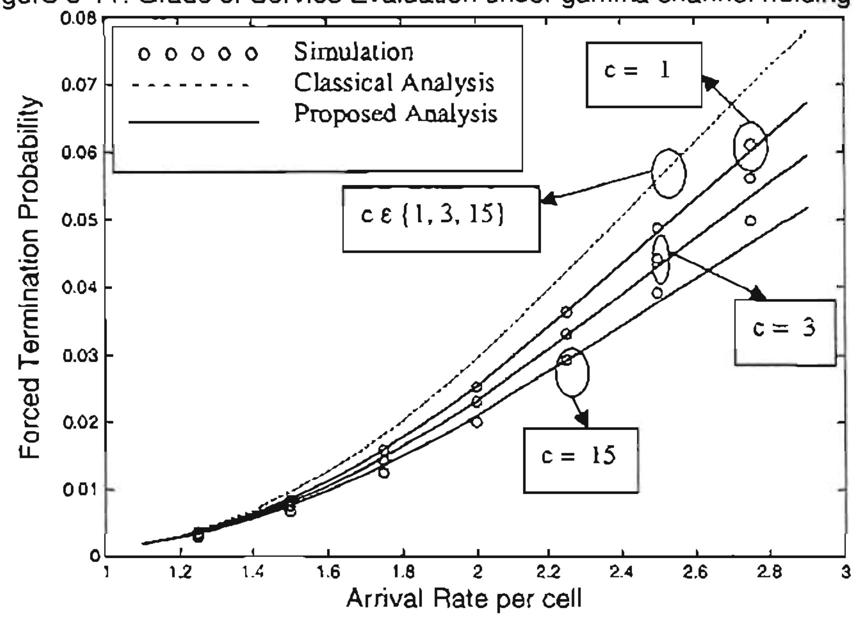


Figure 6-11: Grade of Service Evaluation under gamma channel holding



As can be seen from figure 6-8 and figure 6-10, for new-call blocking, the proposed analysis, the classical analysis and the simulation tie up well. In figure 6-8, there is not much sensitivity in the new call blocking values when the probability parameter, p, of the det-neg distribution takes on different values in the set {0, 5/9, 9/10}. Similarly, in figure 6-10, there is not much sensitivity when the shape parameter, c, of the gamma distribution takes on different values in the set {1, 3, 15}. The following reasons may be given: The simulation, proposed analysis and classical analysis tie up well for new call blocking values because in all of them new call arrivals are modelled as Poisson. The fact that Poisson arrivals are insensitive to service time distributions beyond the mean of the distribution in loss systems [59] may be the reason why different det-neg probability parameters and different gamma shape parameters did not cause any noticeable differences in the new call blocking values in figure 6-8 and 6-10 respectively.

On the other hand, figure 6-9 and figure 6-11, for forced termination probability shows that there are much larger differences when different det-neg probability parameters p and different gamma shape parameters c are used. In figure 6-9, different det-neg probability parameters p produced different forced termination values for the same new call arrival rate for both the simulation results and the proposed analysis. Similarly in figure 6-11, different gamma shape parameters c produced different forced termination values for the same new call arrival rate in both the simulation results and the proposed analysis. However, classical analysis produced the same forced termination values for the different parameters. The "dashed line" representing classical analysis in figure 6-9 is actually three sets of results, one for each det-neg probability parameter p in the set {0, 5/9, 9/10}, superimposed on top of one another. Similarly, the "dashed line" representing classical analysis in figure 6-11 is actually three sets of results, one for each gamma shape parameter c in the set {1, 3, 15}, superimposed on top of one another.

In both figures 6-9 & 6-11, for forced termination probability values, the proposed method approximates simulation results a lot better than classical analysis. The reason for this is that the classical method assumes that hand-off call arrival is Poisson and that channel holding times are negative exponential, whereas the proposed analysis attempts to capture the actual attributes of hand-off traffic using a two moment representation and models the channel holding time distributions exactly as det-neg or gamma distributed as the case may be. A comparison between classical and proposed analysis with respect to simulation is perhaps unfair, since the proposed analysis exactly models the channel holding time distributions found in simulation by employing either det-neg or gamma distributions whereas classical analysis uses only the negative exponential distribution. However such a comparison serves to illustrate the accuracy of the proposed model under controlled conditions and thus validate the assumptions made in it. It also serves to illustrate the limitations in the use of Poisson hand-off arrivals and negative exponential service time distribution in classical performance analysis. Because the Poisson process and the negative exponential distribution are one-parameter models, they are unable to capture any information beyond the mean of the process that they are chosen to model. Thus, for a fixed new call arrival rate and a fixed mean channel holding time, classical analysis produces the same forced termination values when modelling six different channel holding time distributions that have the same mean but entirely different higher moments in figures 6-9 and 6-11.

6.3.3 Convergence Behaviour of our algorithm

Our iteration algorithm converged for every scenario that we tried. This amounts to a very large number of scenarios, considering that the results we have presented in this chapter and the next chapter as well as in our published papers.

The method of iteratively applying a cell traffic characterising model and a cell traffic blocking model in a fixed point iteration is also know as the "relaxation method", or "repeated substitution method". Such relaxation methods are common to fixed and wireless network analysis for both single moment [13,14,15,16] and two-moment analysis [14]. In fact, the reference [14] has an interesting monologue on the convergence of the relaxation method. However the author concludes that although (invariably) relaxation methods work well in the performance analysis of telecommunication networks, the proof for convergence is difficult.

We now consider the nature of the convergence of our algorithm. Since the time for convergence would depend entirely on the software that we implemented the algorithm, we find it more informative to present the number of iterations it required to achieve certain levels of convergence. The following table presents a number of iterations it took for the algorithm to achieve various levels accuracy compared to its final rest position for a typical network scenario.

Table II: The number of iterations required to achieve certain convergence levels

Percentage Difference Compared to final rest position	Number of Iterations it took to achieve particular level of accuracy
1 %	6
0.01 %	10
0.0001 %	14
0.000001 %	19

As can be seen a reasonable level of accuracy is reached with 10-15 iterations. We have found this to be the case in various other networks that we analysed.

6.4 Summary

In this chapter, we presented the Quality of Service evaluation algorithm. This algorithm employs the cell traffic characterising model that we presented in chapter 4 and the cell traffic blocking model that we presented in chapter 5 in an iterative procedure to determine customer orientated grade-of-service parameters such as new call blocking and forced termination probabilities. In the results section, we considered a multitude of scenarios. We considered the performance of our proposed two-moment performance analysis algorithm under the following different conditions:

1. The mobility model of Del Re et al
2. The mobility model of Zonoozi and Dassanayake
3. A mobility model where the cell dwell times are negative exponentially distributed
4. A mobility model where the cell dwell times are deterministically distributed
5. A scenario where the channel holding times are det-neg distributed
6. A scenario where the channel holding times are gamma distributed

For all these different scenarios, we showed that for new call blocking, the proposed analysis, the classical analysis and simulation performed reasonably well. However for forced termination probabilities, we showed that the proposed two-moment analysis easily outperformed the classical single moment analysis compared to simulation. The simulation results, proposed analysis and classical analysis tie up well for new call blocking values because, in all of them, new call arrivals are modelled as Poisson. The fact that Poisson arrivals are insensitive to service time distributions beyond the mean of the distribution is the reason why different channel holding time distributions did not cause any noticeable differences in the new call blocking values. The reason for the superiority of the proposed analysis over classical analysis for determining forced termination probabilities is that hand-off arrival, in practice, is non-Poisson and as such is sensitive to channel holding time distributions. However, classical analysis assumes that hand-off arrival is Poisson and that channel holding times are negative exponentially distributed. These one-parameter models have a serious limitation in that they are unable to capture any information beyond the mean of the process that they are chosen to model. On the other hand, the proposed analysis attempts to capture the actual attributes of hand-off traffic using a two-moment representation and attempts to model the channel holding time distributions more accurately. This is the reason why the proposed analysis performed better. In the next chapter we show that the proposed two-moment performance analysis performs well for more realistic mobility models and network configurations.

CHAPTER 7

7.1 Introduction

In the previous three chapters we have discussed three important aspects of our proposed performance analysis algorithm. Another important and sometimes neglected aspect in the performance analysis of cellular networks is *mobility modelling*. Mobility modelling is used in the performance analysis of cellular networks for deriving the cell dwell time distributions and the channel holding time distributions. However mobility modelling is not limited to the teletraffic performance analysis of cellular networks and may be used in the analysis of:

- Aspects relating to location management (eg. location area planning and paging strategies)
- Aspects relating to handover strategies and mobility related signalling
- Aspects relating to channel assignment schemes

In chapter 2 section 2.5.1 to section 2.5.7 we considered numerous mobility models in the literature. In this chapter we present our contribution to mobility modelling. In the first part we present a mobility model for a highway network serving cellular subscribers in moving vehicles. In the second part we approach the question of mobility modelling from an industrial usage point of view. We show how one may model mobility in an *established* cellular network without actually having to derive the relevant cell dwell time distributions. Our proposed approach is useful in established networks where it is often easier to measure mobility-related traffic parameters from the network than explicitly derive the cell dwell time distributions. For both these approaches to mobility modelling, we provide extensive results to validate our analysis.

7.2 Mobility modelling in highway networks

The user mobility characteristics in a highway cellular network constitute a marked departure from that found in standard cellular networks such as those found in rural and metropolitan areas. The mobility characteristics in a highway environment do not exhibit any of the *memory less* characteristics akin to the negative exponential distribution. On the contrary, the mobility characteristics in a highway have many fixed or *deterministic* features. For example, the distance and direction of travel for vehicles within a highway cell are pretty close to fixed values. The vehicular velocity distribution generally has a small variance around its mean. These mobility attributes can hardly be labelled memory less, and therefore require distributions more generalised than the negative exponential distribution to model.

We consider a highway carrying vehicular traffic in two opposite directions. The cellular communication requirements of occupants in vehicles on the highway are serviced by a *highway cellular network* of concatenated cells along the length of the highway. As proposed in [7], we assume that directional antennas, at street-lamp elevation, generate cell coverage patterns that typically follow the contours of the highway. The highway is divided into segments of length D km, where each segment is a cell with its own base station. We assume that there exists an on-ramp and an off-ramp in each segment of the highway, to bring vehicular traffic to and carry traffic away from the highway.

For the purpose of simplifying the teletraffic analysis, similar to [7], we assume that the highway network does not cater for pedestrians. As shown in figure 7-1, the cellular communication requirements of users outside the highway, namely, pedestrians and

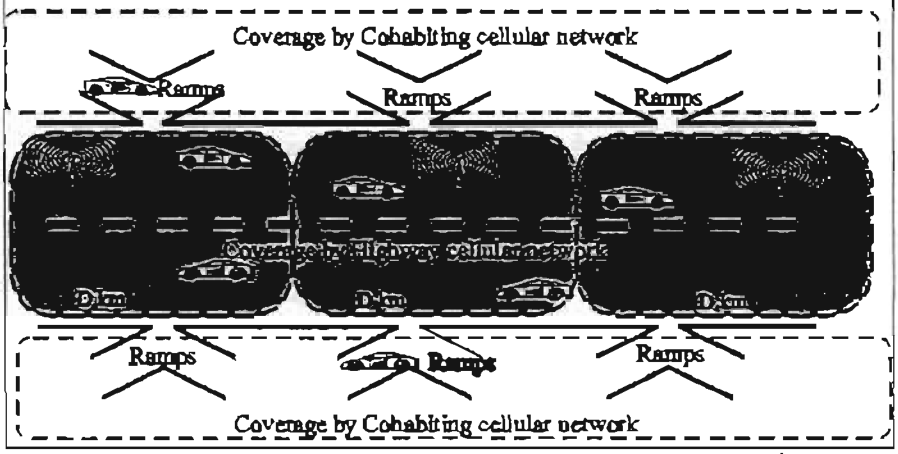


Fig 7-1: Highway Cellular network

vehicles remaining outside the highway as well as vehicles that are waiting to enter or have just exited the highway, are assumed to be serviced by a separate cellular network. This cellular network, called the *cohabiting* network for ease of reference, is assumed to cohabit alongside the highway and cover those areas not covered by the highway micro-cells. To avoid obfuscation of the issues related to the teletraffic performance analysis of the highway network, we artificially assume that the cohabiting network is an infinite capacity or non-blocking network: cellular calls within the cohabiting network are assumed to experience no blocking. Such an assumption allows us to separate the cohabiting network from the highway network during analysis.

We assume that hand-off calls may be offered by the cohabiting network to the highway network when vehicles enter the highway via an on-ramp. Due to the non-blocking environment in the cohabiting network, we assume that the hand-off calls offered by the cohabiting network to the highway network constitute a Poisson process. The Poisson assumption follows the work of Chlebus and Ludwin [19] who showed that the arrival statistics of hand-off calls in a non-blocking environment is the same as that for new calls, namely, Poisson. For simplicity we assume that the Poisson hand-off calls offered by the cohabiting network to the highway network are indistinguishable from the Poisson new calls generated within the highway network from a teletraffic/mobility point of view and that the new call arrival rate λ_n is accordingly inflated to include them.

We also assume that each vehicle in each cell in the highway has a fixed probability, P_x, of leaving the highway via an off-ramp. We assume that cellular calls from such vehicles are seamlessly handed off to the cohabiting network. It is not difficult to accommodate the exit probability, P_x, into our existing framework for the performance analysis of cellular networks, presented in chapters 4, 5 and 6. The exit process (to the cohabiting network) may simply be considered to be a subset of the hand-off process. Each vehicle in each cell is assumed to undergo a simple Bernoulli trial where with a probability of success, P_x, it is assumed to exit the highway network and be handed off to the cohabiting network.

7.2.1 Dwell Time Distribution For New Calls

We assume that the position of the vehicle when an occupant initiates a new call is uniformly distributed along the length D km of a cell. The vehicle is therefore required to travel a further distance Z that is also uniformly distributed between 0 and D km before the cell boundary is reached and hand-off becomes necessary. We consider vehicular velocity V to be constant during travel within each cell and that it is uniformly distributed between a minimum velocity V_{mn} and a maximum velocity V_{mx}.

The pdf of the distance Z is:

$$f_Z(z) = \frac{1}{D} \qquad for\ 0 \le z \le D \tag{7.1}$$

The pdf of the velocity V is:

$$f_V(v) = \frac{1}{V_{mx} - V_{mn}} \qquad for\ V_{mn} \le v \le V_{mx} \tag{7.2}$$

The new call cell dwell time T_n is expressed as:

$$T_n = \frac{Z}{V} \tag{7.3}$$

having the following pdf [15]:

$$f_n(t) = \int_{-\infty}^{\infty} |w| f_Z(tw) f_V(w)\, dw \tag{7.4}$$

$$= \begin{cases} = \dfrac{1}{D}\dfrac{1}{V_{mx} - V_{mn}} \displaystyle\int_{V_{mn}}^{V_{mx}} w\, dw & for\ 0 \le t \le \dfrac{D}{V_{mx}} \\ = \dfrac{1}{D}\dfrac{1}{V_{mx} - V_{mn}} \displaystyle\int_{V_{mn}}^{D/t} w\, dw & for\ \dfrac{D}{V_{mx}} \le t \le \dfrac{D}{V_{mn}} \\ = 0 & elsewhere \end{cases} \tag{7.5}$$

$$= \begin{cases} = \dfrac{V_{mx} + V_{mn}}{2D} & for\ 0 \le t \le \dfrac{D}{V_{mx}} \\ = \dfrac{1}{V_{mx} - V_{mn}} \left(\dfrac{D}{2t^2} - \dfrac{V_{mn}^2}{2D} \right) & for\ \dfrac{D}{V_{mx}} \le t \le \dfrac{D}{V_{mn}} \\ = 0 & elsewhere \end{cases} \quad (7.6)$$

The cdf of T_n is:

$$F_n(t) = \int_0^t f_n(\tau) d\tau \quad (7.7)$$

$$= \begin{cases} = \dfrac{V_{mx} + V_{mn}}{2D} t & for\ 0 \le t \le \dfrac{D}{V_{mx}} \\ = \dfrac{1}{V_{mx} - V_{mn}} \left(V_{mx} - \dfrac{D}{2t} - \dfrac{V_{mn}^2 t}{2D} \right) & for\ \dfrac{D}{V_{mx}} \le t \le \dfrac{D}{V_{mn}} \\ = 1 & for\ \dfrac{D}{V_{mn}} \le t < \infty \end{cases} \quad (7.8)$$

7.2.2 Dwell Time Distribution For Hand-Off Calls

When a call is successfully handed off to a cell from a neighbouring cell, the vehicle in which the mobile user is an occupant, has to travel a further distance, D, in the new cell before hand-off becomes necessary once again. We assume that the vehicular velocity is constant during travel within each cell and that it is uniformly distributed between a minimum velocity V_{mn} and a maximum velocity V_{mx}. The pdf $f_V(v)$ of velocity V is given by equation (7.2).

Let us define an artificial random variable S as follows:

$$S = \frac{V}{D} \quad (7.9)$$

The random variable S is uniformly distributed between $\frac{V_{mn}}{D}$ and $\frac{V_{mx}}{D}$ and has pdf $f_S(s)$:

$$f_S(s) = \frac{1}{\frac{V_{mx}}{D} - \frac{V_{mn}}{D}} \quad for\ \frac{V_{mn}}{D} \le s \le \frac{V_{mx}}{D} \quad (7.10)$$

We are, however, interested in the cell dwell time distribution of hand-off calls. The random variable representing the cell dwell time of hand-off calls, T_h, is related to the artificial random variable S as follows:

$$T_h = \frac{1}{S} \tag{7.11}$$

We define the transformation G: $\mathbb{R} \rightarrow \mathbb{R}$, $S \rightarrow \frac{1}{S}$ where the reciprocal of our artificial variable S is taken. The transformation or 'mapping' G is one-to-one. The inverse transformation G^{-1} exists and is also one where the reciprocal is taken. Using simple calculus, the pdf $f_h(t)$ of the cell dwell time T_h of the hand-off calls is expressed as:

$$f_h(t) = f_S(G^{-1}(t)) \left| \frac{dG^{-1}(t)}{dt} \right| \tag{7.12}$$

$$f_h(t) = \frac{1}{\frac{V_{mx}}{D} - \frac{V_{mn}}{D}} \frac{1}{t^2} \quad for \ \frac{D}{V_{mx}} \le t \le \frac{D}{V_{mn}} \tag{7.13}$$

The cdf $F_h(t)$ is:

$$F_h(t) = \int_{D/V_{mx}}^{t} f_h(\tau)\, d\tau \tag{7.14}$$

$$= \begin{cases} = 0 & for \ 0 \le t \le \frac{D}{V_{mx}} \\ = \frac{D}{V_{mx} - V_{mn}} \left(\frac{V_{mx}}{D} - \frac{1}{t} \right) & for \ \frac{D}{V_{mx}} \le t \le \frac{D}{V_{mn}} \\ = 1 & for \ \frac{D}{V_{mn}} \le t < \infty \end{cases} \tag{7.15}$$

The complementary distribution function [15] (or survivor function) of the channel holding time distribution is $F_H^C(t) = 1 - F_H(t)$:

$$F_H^C(t) = e^{-\mu_M t} - \frac{e^{-\mu_M t}}{1 + \Upsilon_C} (F_n(t) + \Upsilon_C F_h(t)) \tag{7.16}$$

The mean channel holding time $E[T_H]$ and the variance of the channel holding time $V[T_H]$ may be obtained using the following equations by applying simple numerical integration:

$$E[T_H] = \int_0^{\infty} t f_H(t)dt = \int_0^{\infty} F_H^C(t)dt \tag{7.17}$$

$$V[T_H] = \left(\int_0^{\infty} t^2 f_H(t)dt \right) - E^2[T_H] = 2\left(\int_0^{\infty} t F_H^C(t)dt \right) - E^2[T_H] \tag{7.18}$$

It is then possible to use these expressions in the performance analysis of highway cellular networks.

7.3 Results for the highway network

The results section consists of three parts. In all three parts we compare the performance of our proposed analysis with classical analysis [15,16] and simulation results. We vary input parameters such as new call arrival rate λ_n, highway exit probability P_x, channel reservation parameter r, and the cell length D, to gauge the performance of the above-mentioned methods under different conditions. In the following analysis, we employ a homogenous cellular network that consists of a ring of 12 concatenated cells. In each cell we set C=20 channels. There are two opposing directions of vehicular flow, clockwise and counter-clockwise around the ring. We assume that the vehicular velocity is constant during travel within each cell and that it is uniformly distributed between 90-120km/hr (56-75mph). Ring highway networks are not uncommon and probably the most famous one is the M25 motorway encircling the city of London. We analysed the highway network using the performance analysis algorithm presented in chapters 4, 5 and 6. In our analysis we used the det-neg pdf rather than the gamma pdf as our candidate channel holding time distribution since the det-neg pdf better fitted the actual channel holding time pdf than did the gamma pdf.

7.3.1 Performance analysis under varying offered load conditions

We determined the Quality of Service parameters by fixing the highway exit probability $P_x = 0.05$, cell size D = 1.0km, whilst varying λ_n, the new call arrival rate, from 3 to 12 calls per cell per minute. For variety we considered both r=2 and r=3 reservation levels. As can be seen from figures 7-2 & 7-3, the results show that the proposed analysis is better than the classical analysis compared to simulation results for both reservation levels. Classical analysis appears to have underestimated the simulation results for new call blocking and overestimated them for forced termination probabilities. The proposed analysis tracks simulation results more accurately from underload through nominal load to overload conditions. Interestingly, the difference between the proposed method and the classical method is more pronounced for forced termination probability rather than for new call blocking. For convenience of notation we call our model the G/G'/C method since it employs generalised hand-off arrivals and det-neg channel holding time pdf which is more generalised than the simple negative exponential pdf. For simplicity we call the single moment analysis under Poisson hand-off arrivals and neg-exp channel holding times the M/M/C model.

Comparison for new call blocking: The reasoning for the small difference in new call blocking values between classical analysis and proposed analysis may be that new call arrivals are modelled as Poisson arrivals in both methods. The fact that channel holding times are modelled differently, as negative exponential in the classical analysis and as the more generalised det-neg distribution in the proposed analysis, does not appear to have caused large differences in the new call blocking values between the proposed method and classical method. This may be attributed to the insensitivity of the Poisson arrival process to service distributions beyond the mean [59].

Comparison for forced termination probability: Much larger differences are noticeable between the proposed method and classical analysis in the forced termination probability results as shown in figure 7-3. The reasoning is two-fold. Classical analysis models hand-off as Poisson whereas the proposed analysis models hand-off more accurately as a generalised arrival process. Furthermore, unlike the Poisson new call arrivals which were insensitive to service time distributions, it appears that the hand-off call arrival process, being non-Poisson, is sensitive to service time distributions. The accuracy of the proposed analysis may therefore also be attributed to the fact that it models the channel holding time distribution as a more generalised distribution whereas classical analysis models it as the simple negative exponential.

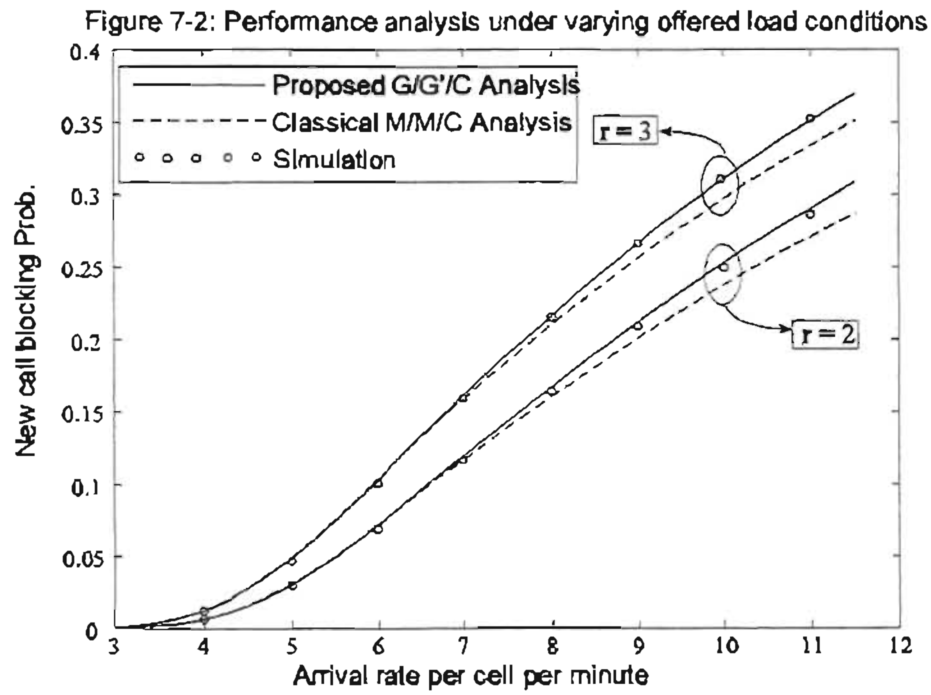


Figure 7-2: Performance analysis under varying offered load conditions
Proposed G/G'/C Analysis
Classical M/M/C Analysis
Simulation
r = 3
r = 2
New call blocking Prob.
Arrival rate per cell per minute
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
3
4
5
6
7
8
9
10
11
12

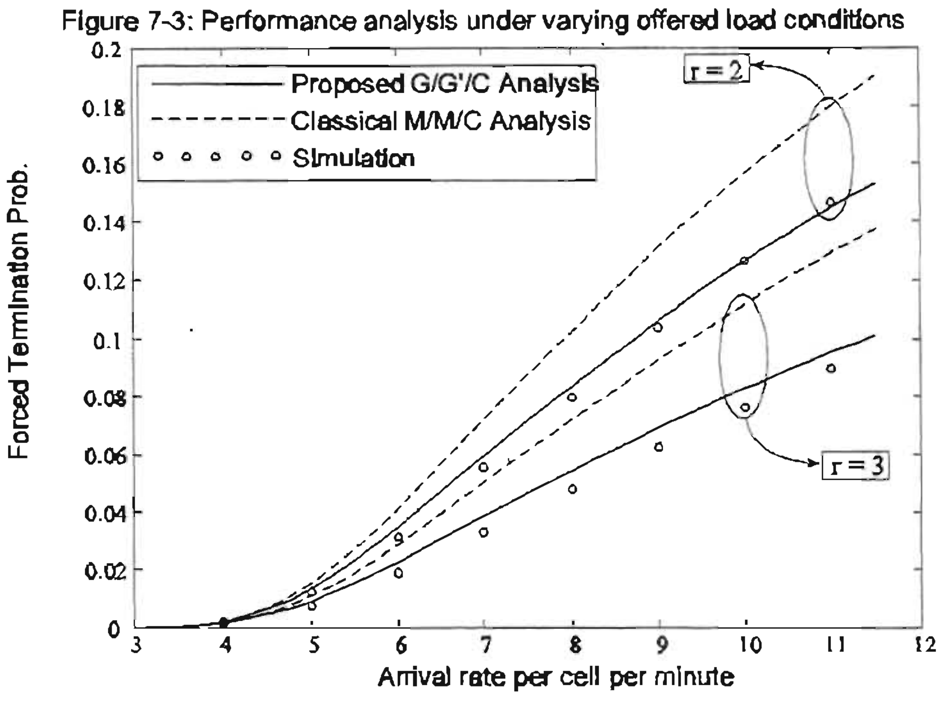


Figure 7-3: Performance analysis under varying offered load conditions
Proposed G/G'/C Analysis
Classical M/M/C Analysis
Simulation
r = 2
r = 3
Forced Termination Prob.
Arrival rate per cell per minute
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0.16
0.18
0.2
3
4
5
6
7
8
9
10
11
12

7.3.2 Performance analysis under varying highway exit probabilities

We determined new call blocking and forced termination probabilities by fixing the new call arrival rate to 11 calls per cell per minute, cell size to D=1.0km, whilst varying the highway exit probability P_x from 0 to 0.3 and setting reservation levels at *r*=2 and *r*=3. Figures 7-4 and 7-5 illustrate these results. Interestingly, new call blocking and forced termination values decrease with increasing probabilities of exit. The reasoning is that with increasing probabilities of exit P_x, more and more traffic is handed out of the cellular network to the cohabiting network instead of being handed to the next cell within the highway ring of cells. The resulting decreasing levels of hand-off traffic translates to more channels being available and therefore less competition between new and hand-off calls for free channels. Decreasing values of blocking are thus observed in association with increasing highway exit probabilities, P_x.

Figure 7-4: Performance analysis under varying highway exit probabilities

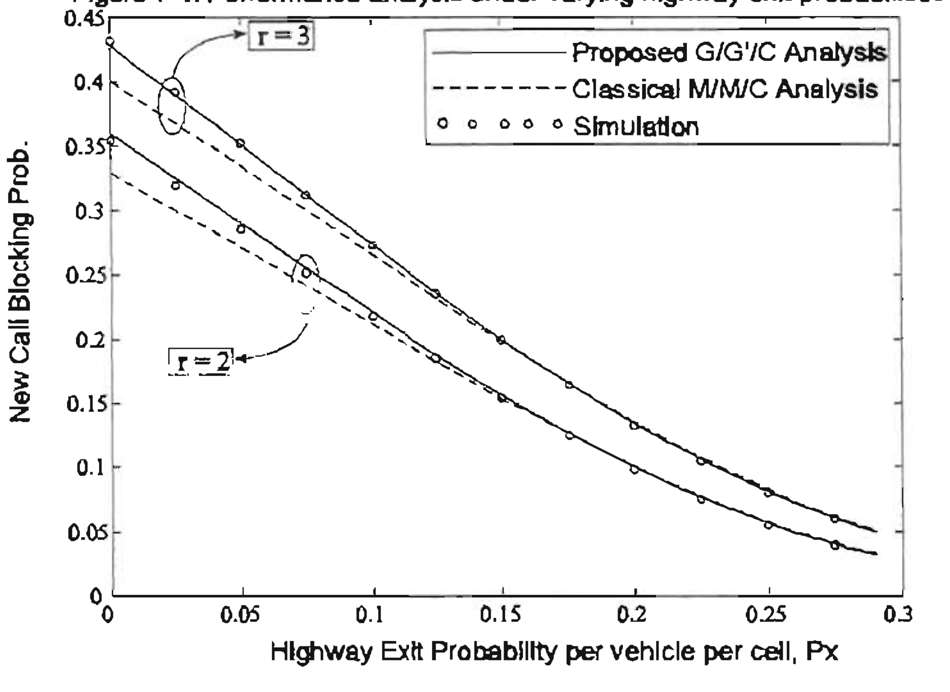
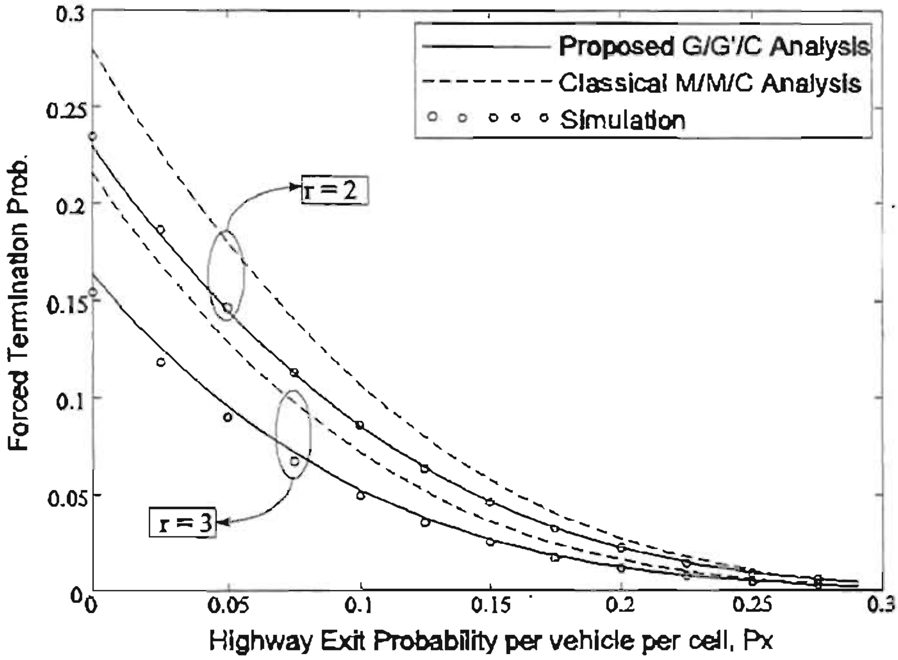


Figure 7-5: Performance analysis under varying highway exit probabilities



As far as accuracy is concerned, once again the proposed analysis is better than classical analysis compared to simulation for both reservation values. As before, the difference between classical analysis and proposed analysis is smaller for new call blocking values and much larger for forced termination values. The reason for smaller differences in the new call blocking values is that in both methods new call arrivals are modelled as Poisson and Poisson arrivals are insensitive to service distributions beyond the mean [59]. The reason for larger differences in the forced termination values is that the classical method models hand-off as Poisson and assumes that the channel holding time distribution is negative exponential, whereas the proposed model attempts to capture the true nature of hand-off traffic using a generalised arrival process and also models the channel holding time distribution as a more generalised distribution.

7.3.3 Performance analysis under varying cell sizes

In this scenario, we determined new call blocking and forced termination probabilities, by setting the new call arrival rate, λ_n, to 6 calls per cell per minute, the highway exit probability P_x to 0.05, channel reservation r to 2, and then varied the cell sizes from 1.5 - 4.5 km (0.94 - 2.8 miles) in length. In figures 7-6 & 7-7, we have compared the proposed G/G'/C analysis with classical analysis and simulation results. For interest sake, we have also included the results as obtained using our earlier G/M/C analysis (using generalised hand-off arrival but neg-exp channel holding times). In table I, we have reported the mean channel holding times as calculated using our proposed G/G'/C model for the different cell lengths. We have also included the Erlang's number of the corresponding channel holding time distributions as calculated by our G/G'/C model. Note that the Erlang's Number of a pdf is the ratio of the square of the mean to the variance of that pdf. The reason why we call this parameter, Erlang's number, is that for integral values, it is the number of stages that would be required by an Erlang-k pdf to achieve the same ratio of mean squared over variance. Erlang's number for the negative exponential distribution is 1.

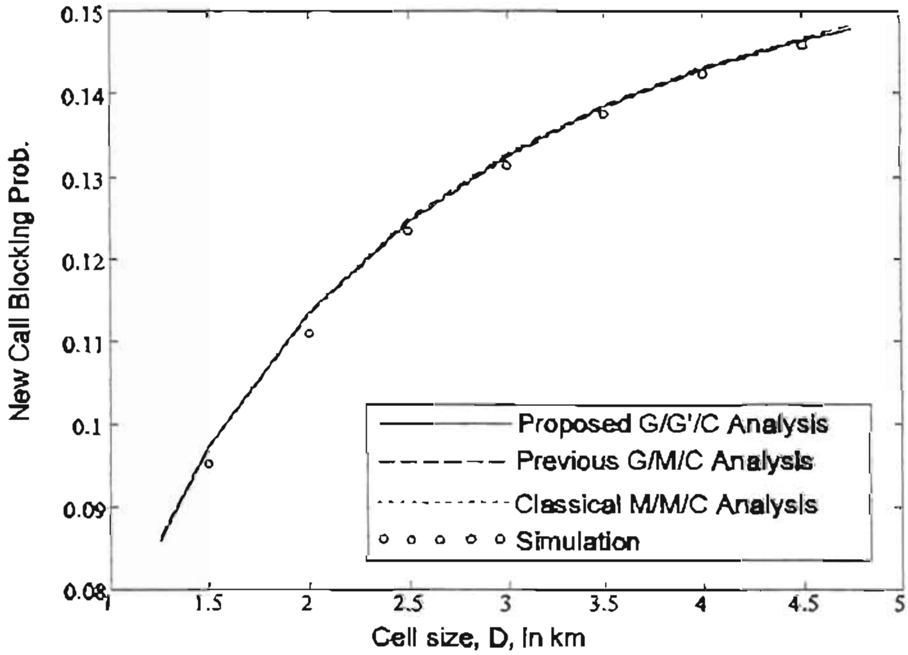
Interestingly, in figure 7-6, new call blocking values increase with increasing cell size, but in figure 7-7, forced termination values decrease with increasing cell size: Although, the new call arrival rate, λ_n, is fixed at 6 calls per cell per minute, the mean channel holding time $E[T_H]$ increases with increasing cell size (see table I). This means that the new call offered traffic, $A_n=\lambda_n.E[T_H]$, increases with increasing cell size and this results in increasing new call blocking values as shown in figure 7-6. The decrease in forced termination values with increasing cell size is a bit more difficult to explain. There are two forces at work. Firstly, as shown in table I, the mean channel holding times increase with increasing cell sizes. However, the mean hand-off arrival rate decreases with increasing cell size. This is because, with increasing cell size, there is a better chance of calls terminating, after completion of conversation, within each cell and therefore there is less need for further hand-offs. This results in decreasing probabilities of hand-off P_N and P_H, which translate to decreasing hand-off arrival rates. Furthermore, equation (6.12) shows that forced termination values decrease in accordance with decreasing probabilities of hand-off, P_N & P_H. We believe that all the above contribute to decreasing forced termination probabilities with increasing cell sizes. Note that the reason why the mean channel holding time increases with increasing cell sizes is simply that the mean channel holding time [equation (7.26)] is a composite of the unencumbered call holding time, which is fixed, and the probabilities of hand-off, P_N and P_H, which decrease with increasing cell size.

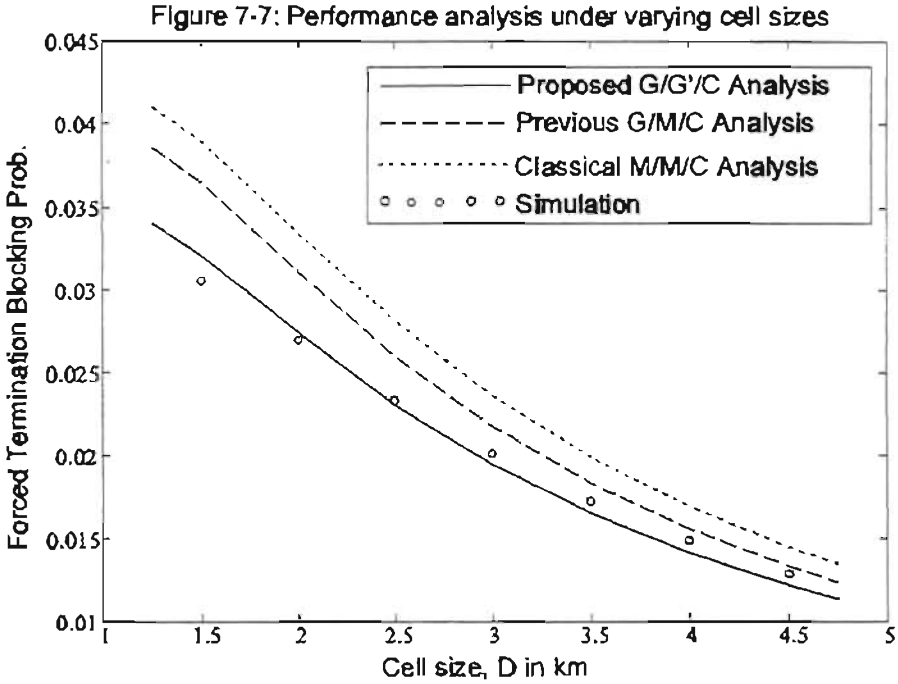
Table I: Mean Channel Holding Times and Erlang's Number for various cell sizes

Cell Size (km)	1.5	2	2.5	3	3.5	4	4.5
Mean Channel Holding Time (s)	39	49	57	64	71	77	82
Erlang's Number	5.1	4.1	3.5	3.0	2.7	2.5	2.3

As far as accuracy is concerned, in figure 7-6, for new call blocking values, all three analytical methods perform equally well. However, larger differences are noticeable in figure 7-7 for forced termination values. As can be seen from figure 7-7, the proposed G/G'/C analysis performed the best compared to simulation. The G/M/C model, although inferior to the G/G'/C model, performed better than the classical M/M/C model. The obvious reasons are that the G/G'/C method, and to a certain extent the G/M/C

Figure 7-6: Performance analysis under varying cell sizes





method, model the teletraffic characteristics of highway cellular networks more accurately and therefore provide better results. However, what is more interesting is that the G/G'/C model outperforms the G/M/C and M/M/C models substantially more for small cell sizes than for large cell sizes. The reason for this phenomenon is evident in table I. The Erlang's Number, for channel holding time distributions in cells that are small in size, is much larger than 1, the Erlang's Number for the negative exponential pdf. This means that in small cell sizes, the corresponding channel holding time distributions are markedly different from the negative exponential distribution. Therefore, larger errors are accrued by the G/M/C and M/M/C methods when they attempt to model these channel holding time distributions as negative exponential.

7.4Mobility Modelling using network measurements

The classical approach to mobility modelling relies heavily on the need to derive explicitly the cell dwell time distributions. In real-life cases, this can be a insurmountable task. In this section we propose a methodology whereby one can circumvent the need to derive explicitly the cell dwell time distributions. We propose to do this by measuring certain mobility related parameters from live networks and thereby obtain sufficient information regarding user mobility to be able to performance analyse these real-life networks. Firstly, we need the following proof regarding mean channel holding times.

7.4.1 Channel holding time distribution under arbitrary cell dwell times

We assume that the unencumbered call holding time, T_M, is negative exponentially distributed with mean $1/\mu_M$ (see equations 2.3–2.4). We define two random variables, T_n and T_h, which describe the cell dwell times of new and previously handed-off calls in a cell. We then assume that the new call cell dwell times and hand-off call cell dwell times in a cell have arbitrary probability density functions (pdfs), $f_n(t)$ and $f_h(t)$ respectively. In chapter 2, we defined P_N and P_H to be the probabilities with which a new call and a previously handed-off call would require a new hand-off and are given by:

$$P_N = Prob\{T_M > T_n\} = \int_0^\infty [1 - F_M(t)]\, f_n(t)\, dt$$
$$P_H = Prob\{T_M > T_h\} = \int_0^\infty [1 - F_M(t)]\, f_h(t)\, dt \qquad (7.19)$$

where $F_M(t)$ is the cumulative distribution function (cdf) of T_M. Hong and Rappaport showed that the cdf, $F_H(t)$, of the channel holding time T_H may be written as:

$$F_H(t) = F_M(t) + \frac{1}{1+\Upsilon_C}[1 - F_M(t)][F_n(t) + \Upsilon_C F_h(t)] \qquad (7.20)$$

where Yc is the ratio of mean hand-off carried rate to mean new call carried rate as given by equation (3.2). We now derive the mean channel holding time $E[T_H]$ from its cumulative distribution function $F_H(t)$ for arbitrary cell dwell time cdfs $F_n(t)$ and $F_h(t)$ given that the unencumbered call holding time T_M is negative exponentially distributed. The pdf $f_H(t)$ of the channel holding time may be found by differentiating (7.20):

$$f_H(t) = \mu_M e^{-\mu_M t} + \frac{e^{-\mu_M t}}{1+\Upsilon_c}\{ f_n(t) + \Upsilon_c f_h(t)\} - \frac{\mu_M e^{-\mu_M t}}{1+\Upsilon_c}\{ F_n(t) + \Upsilon_c F_h(t)\} \qquad (7.21)$$

The mean channel holding time $E[T_H]$ is simply the integral:

$$E[T_H] = \int_0^{\infty} t\, f_H(t)\, dt \qquad (7.22)$$

Let us assume that the pdf $f_n(t)$ of the cell dwell of time of new calls is zero for values of t outside a certain region $[a, b]$ and that the pdf $f_h(t)$ of the cell dwell of time of previously handed off calls is zero for values of t outside a certain region $[c, g]$. This implies that the cdfs $F_n(t<a)=0$, $F_n(t>b)=1$, $F_h(t<c)=0$ and $F_h(t>g)=1$. The actual values for a,b,c & g do not preclude any distribution functions from the following analysis since, if required, a & c can be made arbitrarily close to zero and b & g can be made arbitrarily large. The following three results are useful when integrating equation (7.21): (Note that the results presented in equations (7.23-7.25) are easily obtained using integration by parts):

$$\int_0^{\infty} t\, \mu_M e^{-\mu_M t}\, dt = \frac{1}{\mu_M} \qquad (7.23)$$

$$\begin{aligned}
\int_a^{\infty} t\, [e^{-\mu_M t} f_n(t) - \mu_M e^{-\mu_M t} F_n(t)]\, dt &= t\, e^{-\mu_M t} F_n(t)\Big|_a^{\infty} - \int_a^{\infty} e^{-\mu_M t} F_n(t)\, dt \\
&= t\, e^{-\mu_M t} F_n(t)\Big|_a^{\infty} + \frac{1}{\mu_M} e^{-\mu_M t} F_n(t)\Big|_a^{\infty} - \int_a^{\infty} \frac{1}{\mu_M} e^{-\mu_M t} f_n(t)\, dt \\
&= 0 + 0 - \int_a^{b} \frac{1}{\mu_M} e^{-\mu_M t} f_n(t)\, dt
\end{aligned} \qquad (7.24)$$

$$\begin{aligned}
\int_c^{\infty} t\, [e^{-\mu_M t} f_h(t) - \mu_M e^{-\mu_M t} F_h(t)]\, dt &= t\, e^{-\mu_M t} F_h(t)\Big|_c^{\infty} - \int_c^{\infty} e^{-\mu_M t} F_h(t)\, dt \\
&= t\, e^{-\mu_M t} F_h(t)\Big|_c^{\infty} + \frac{1}{\mu_M} e^{-\mu_M t} F_h(t)\Big|_c^{\infty} - \int_c^{\infty} \frac{1}{\mu_M} e^{-\mu_M t} f_h(t)\, dt \\
&= 0 + 0 - \int_c^{g} \frac{1}{\mu_M} e^{-\mu_M t} f_h(t)\, dt
\end{aligned} \qquad (7.25)$$

Then using the definitions of the probabilities of hand-off, P_N and P_H, as shown in equation (7.19), the mean channel holding time $E[T_H]$ may be shown to be:

$$E[T_H] = \frac{1}{\mu_M}\left(1 - P_N \frac{1}{1+\Upsilon_c} - P_H \frac{\Upsilon_c}{1+\Upsilon_c}\right) \qquad (7.26)$$

This equation for the mean channel holding time is quite illustrative and, by its very nature, suggests the type of measurements that need to be made on live networks for mobility modelling.

7.4.2 Performance analysis of practical cellular networks where cell dwell time distributions are unknown [62,72].

So far we have considered the performance analysis of cellular mobile networks where the mobility of the users is assumed to be known. However, in many practical cases, it is difficult, sometimes impossible, to explicitly derive the cell dwell time distributions or even find reasonable approximations for them. In this section we examine how one may performance analyse established cellular networks in the absence of explicit cell dwell time distributions.

We consider a homogeneous established cellular network with C Channels in each cell. Since our network is already established we assume that the following easy-to-obtain parameters may be measured for a cell L by traffic counters at the base station over a time interval T_{MI}. We measure:

1. The number of new calls carried N_{CC} over time T_{MI}.
2. The number of carried new calls which request for a hand-off $N_{CC\text{-}H}$ over time T_{MI}.
3. The number of hand-off calls carried H_{CC} over time T_{MI}.
4. The number of carried hand-off calls which request for another hand-off $H_{CC\text{-}H}$ over time T_{MI}.
5. The duration T_i out of the total measurement interval T_{MI} for which channel i (for i=1,2...C) is occupied in cell L.

Once the above measurements are made, we make the following common assumptions [15,16]:

(a) New call arrival is Poisson;

(b) Unencumbered call holding time is negative exponentially distributed;

(c) The mobility of the users and therefore their cell dwell time distributions are independent of traffic loading conditions;

The astute reader will recognise that the third assumption is nothing new and is implicit in existing mobility models such as those by Del Re et al [2], Zonoozi and Dassanyake [18], Guerin [17] etc. Using the above assumptions, we exploit the strength of the result stated in equation (7.26) to determine the mean unencumbered call holding time $1/\mu_M$

(NB. The strength of the result in equation (7.26) is that it applies for arbitrary cell dwell time distributions):

$$\frac{1}{\mu_M} = \frac{E[T_H]}{1 - \frac{P_N}{1+\Upsilon_c} - \frac{P_H \; \Upsilon_c}{1+\Upsilon_c}} \tag{7.27}$$

The various parameters of the expression on the right hand side may be determined from the measured parameters of the network as follows:

$$E[T_H] = \frac{\sum_{i=1}^{C} T_i}{N_{CC} + H_{CC}} \tag{7.28}$$

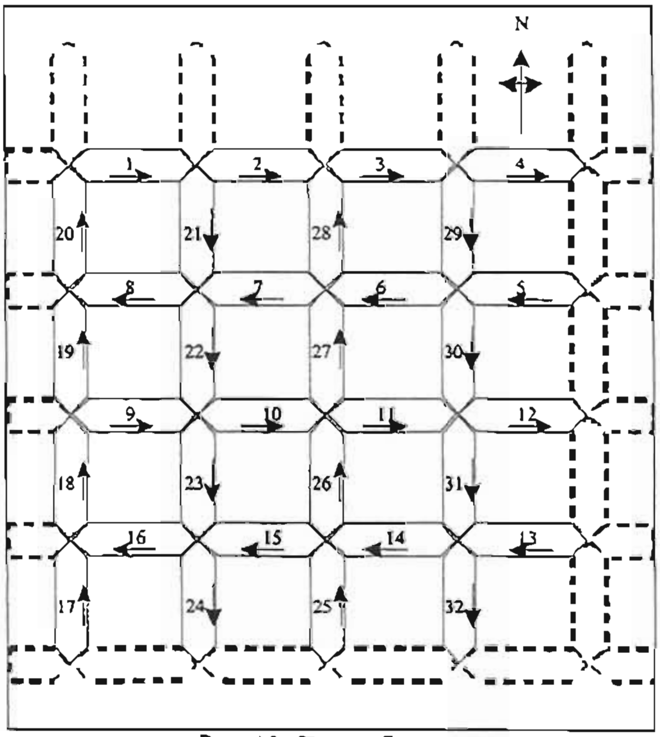
$$P_N = \frac{N_{CC\text{-}H}}{N_{CC}} \; ; \; P_H = \frac{H_{CC\text{-}H}}{H_{CC}} \; ; \; \Upsilon_c = \frac{H_{CC}}{N_{CC}} \tag{7.29}$$

Since the unencumbered call holding time distribution and the cell dwell time distributions are assumed to be independent of traffic loading conditions, the mean unencumbered call holding time $E[T_M] = 1/\mu_M$ as given by equation (7.27) and the probabilities of hand-off P_N and P_H as given by equation (7.19) are also independent of traffic loading conditions. Given $1/\mu_M$, P_N, and P_H, we may performance analyse the network using the methodology suggested in chapters 4, 5 and 6 for arbitrary new call arrival rate λ_n. In this manner we bypass the explicit derivation of the cell dwell time distributions when performance analysing an established cellular network. The obvious short fall of this approach is that it ignores higher moments of the channel holding time distribution. In the absence of any information regarding higher moments, such a mobility model is strictly only useful for scenarios where the channel holding time distribution can be represented by the negative exponential distribution. However, in industry it is common to use the so-called "quick and dirty" methods and the above approach offers such a viable alternative to classical mobility modelling as we shall show in the next section.

7.5 The Performance Analysis of A Realistic Street Micro-Cellular System Where the Cell Dwell Time Distributions Are Unknown

In this section, we simulate a street micro-cellular network proposed for the city of Durban, South Africa. Like most modern cities, the city streets in Durban form a grid of intersecting one-way streets. The traffic flow in the parallel one-way streets are directed in a manner that, one-way streets carrying traffic in a Northerly direction alternate with one-way streets carrying traffic in a Southerly direction. Similarly, one-way streets carrying traffic in an Easterly direction alternate with one-way streets carrying traffic in a Westerly direction. We assume that street micro-cells are set up by exploiting the electromagnetic shielding provided by the various high-rise buildings. We assume that a single micro-cell covers the street segment between two adjacent intersections. In our simulator we consider 32 micro-cells covering the street segments in 16 blocks of city buildings. As shown in figure 7-8, to avoid edge effects, we assume that the network wraps around itself. Each micro-cell is assumed to have 20 channels under a FCA channel assignment method. We consider channel reservation levels of r=2 and r=3. We assume that a vehicle, moving along a street, has a choice of either turning or going straight, at each intersection. Since all the streets are one-way, we assume that a vehicle may turn in only one direction at each intersection but not both. We assume that all vehicles have a probability R_h =0.8 of going straight at an intersection and a probability of R_h = 0.2 of turning in the permissible direction at the same intersection. Table II indicates the direction of hand-off for the various cells.

Figure 7-8: Street Micro-cell Layout



Under the assumption that the network wraps around itself, it can be seen that each of the 32 cells receives two hand-off traffic streams, one from vehicles that went straight at an intersection and the other from vehicles that turned at an intersection to reach the cell.

We assume that each intersection is controlled by traffic lights. We assume that the "stop" (red) and "go" (green) periods of the traffic lights last twenty seconds each. We also assume that the red and green periods of the various traffic lights are timed in a manner that, whilst traffic is flowing in the Northerly/Southerly direction, traffic comes to a standstill along the streets running in the Easterly/Westerly direction and vice versa. Turning into a busy street is not always a smooth process. A vehicle, when turning into a street, requires a bit of time to merge with existing traffic on that street. We assume that vehicles that choose to turn at an intersection require a random time interval ψ to merge with existing traffic on that street. The random variable ψ is assumed to have a truncated negative exponential distribution with mean of 5 seconds. For simplicity we also assume that our micro-cellular network caters for only callers in moving vehicles and not pedestrians. This assumption is made because our proposed model was derived for only one type of users. Extensions are required to the work in this paper to allow for multiple platforms. We assume that the length of a street segment between adjacent intersections is 100m. For simulation purposes each micro-cell is assumed to be rectangular in shape with length 100m. Given that a vehicle's position may be uniformly distributed along the length of the cell when an occupant initiates a new call, the vehicle has to travel a distance that is uniformly distributed between 0 and 100m before hand-off becomes necessary. On the other hand, a vehicle offering hand-off traffic to a cell needs to transit the full distance 100m before hand-off is necessary once again. We assume that vehicular velocity V is a random variable but which remains constant during travel within a cell. We consider velocity V to be uniformly distributed between 0 and 40 km/hr.

TABLE II: DIRECTION OF HAND-OFF FOR VARIOUS CELLS

Cell No.	Hands-off forward to cell number	Hands-off after turning to cell number	**Cell No.**	Hands-off forward to cell number	Hands-off after turning to cell number
1	2	21	17	18	13
2	3	25	18	19	9
3	4	29	19	20	5
4	1	17	20	17	1
5	6	30	21	22	8
6	7	28	22	23	10
7	8	22	23	24	16
8	5	20	24	21	2
9	10	23	25	26	15
10	11	27	26	27	11
11	12	31	27	28	7
12	9	19	28	25	3
13	14	32	29	30	6
14	15	26	30	31	12
15	16	24	31	32	14
16	13	18	32	29	4

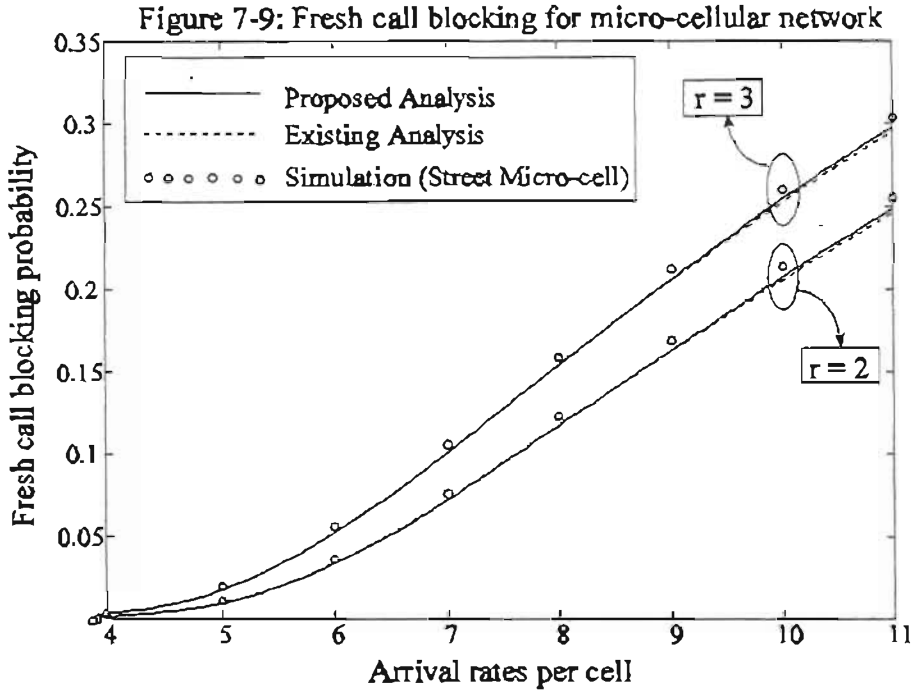
The explicit derivation of cell dwell time distributions for the above mobility model is cumbersome. We circumvent the explicit derivation of the cell dwell time distributions using the methodology suggested in section 7.4. We measured the necessary mobility parameters from the simulator for a particular value of new call arrival rate. Then using the analysis presented in section 7.4, we determined the mean unencumbered call holding time $1/\mu_M$, and the probabilities of hand-off P_N and P_H. Using these values we then determined new call blocking, hand-off call blocking and forced termination probabilities under the proposed two moment analysis and existing single moment analysis (under neg-exp channel holding times) for various new call arrival rates. Note that in assumption (2) of section 4.2, we assumed that a cell has N neighbours and given that a call is to be handed off, it has probability $R_j = 1/N$ of being handed off to neighbour j. However this is not the case in the micro-cellular network of figure 7-8.

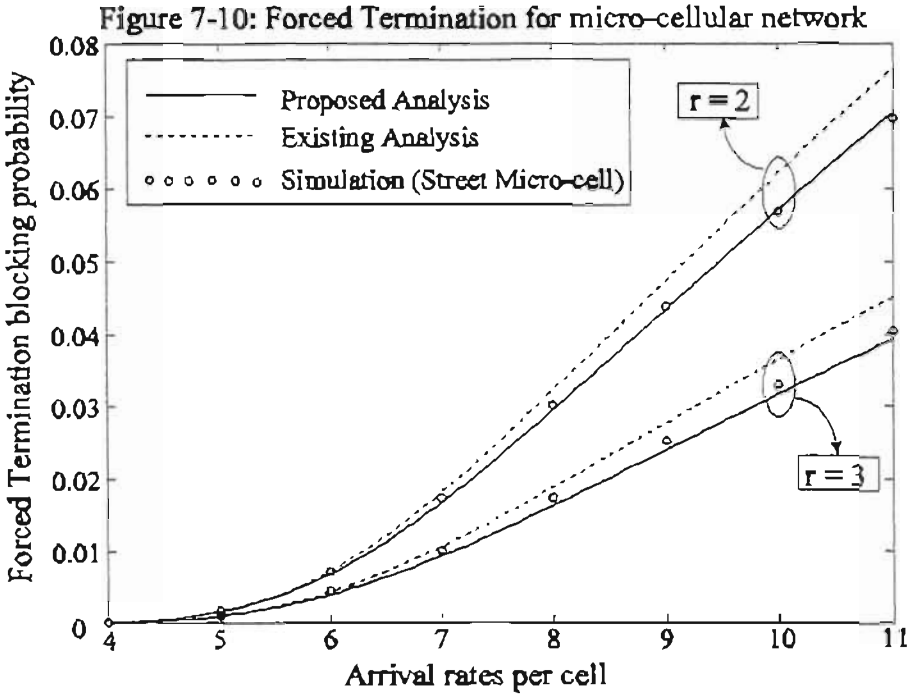
Consider cell #10, it receives hand-off traffic with probability R_h=0.8 from cell #9 and hand-off traffic with probability R_h=0.2 from cell #22. Therefore the mean and variance as determined by equations (6.7-6.9) are not the same for the two different streams. We suggest the following equation in place of equation (6.2) to determine the mean M_Σ and variance V_Σ of the aggregate hand-off offered traffic stream.

$$M_\Sigma = M_{ho}(R_h = 0.8) + M_{ho}(R_h = 0.2) \quad ; \quad V_\Sigma = V_{ho}(R_h = 0.8) + V_{ho}(R_h = 0.2) \tag{7.30}$$

Once the blocking experienced by the aggregate hand-off stream is determined using the proposed cell traffic blocking method, we assume that the two component streams experience the same blocking as well (similar to assumption 8 of section 4.2). Strictly speaking, the two different streams should experience different blocking, but this approximation is used since it simplifies analysis. Furthermore, its effect can be seen from comparison with simulation which is free of such restrictions. Figures 7-9 and 7-10 list the various results obtained for the above network scenario.

Figure 7-9: Fresh call blocking for micro-cellular network





As can be seen from figure 7-9, for new-call blocking, the proposed analysis, existing analysis and simulation provide very similar results. The reason for this is that new call arrival is modelled as Poisson in all three cases. Figure 7-10 shows much larger differences for forced termination probability between the proposed analysis and the existing model when compared to simulation. The reason for this is that the proposed analysis models the hand-off arrival process as a general arrival process whereas the existing method models it as Poisson. The importance of the results derived in this section is that the existing single moment analysis and the proposed two-moment analysis both circumvented the explicit derivation of the cell dwell time distributions. However, such a methodology is only possible for established networks where mobility / traffic measurements are possible.

7.6 Summary

In this chapter, we presented different approaches to mobility modelling. This included an analytic mobility model for a highway vehicular scenario, where exact cell dwell time distributions for vehicular travel were derived under simple mobility assumptions.

The thesis also presented a practical approach to mobility modelling in established cellular networks where appropriate mobility-related measurements from the network may be employed as a simpler alternative to deriving exact cell dwell time distributions.

CHAPTER 8

Conclusions

The implementation cost of a cellular network can easily amount to billions of dollars. A few percentage points reduction in such implementation costs can lead to massive savings. Accurate forecasting, dimensioning and performance-analysis are the keys to achieving these savings. The accurate performance analysis of cellular networks is the main contribution of this thesis.

Performance analysis involves the determination of customer orientated grade-of-service parameters, such as call blocking and dropping probabilities, using the methods of stochastic theory. This stochastic theory analysis is built on certain assumptions regarding the arrival and service process of user-offered calls within a network. In chapter 2, we considered the various arrival and service processes relevant to the performance analysis of cellular networks. These include,

- New call arrivals
- Hand-off call arrivals
- Unencumbered call holding times
- Cell dwell times
- Channel holding times.

In chapter 2, we examined the various models proposed in the literature for the above five processes. We considered analytical models, simulation studies and field measurements results. The analytical models for the above processes mainly consisted of negative exponential distributions or mathematical combinations thereof. As such, the analytical models naturally lent themselves to tractable performance analysis algorithms. In both the simulation studies and the field measurements, suitable distributions for the relevant arrival and service-processes were obtained by best fitting to the results. Although, these distributions are quite useful in their own right, they do not naturally lend themselves to elegant and tractable models during performance analysis.

In chapter 3, we examined the various performance analysis algorithms proposed in the literature. These performance analysis algorithms generally differ from one another in the manner they treat the various arrival and service processes relevant to cellular networks. It is clear that the simple negative exponential distribution and mathematical combinations of negative exponential distributions are the most favoured distributions to model call arrival and call service processes. None of the performance analysis algorithms that we considered in this chapter made use of arrival and service processes obtained directly from simulation studies or field measurement studies. The work by Professor Rappaport's team stands out as a tour de force in the performance analysis of cellular networks. However, these models suffer from limitations as well. The main limitations are that of state space explosion and the assumption of Poisson-type hand-off processes.

This thesis's contribution to cellular network analysis is a *moment-based* performance analysis algorithm that avoids full state space description but ensures that the hand-off arrival process is modelled beyond the first moment. The thesis showed that the performance analysis of a cellular network may be loosely decomposed into three parts, a generic *cell traffic characterising model*, a generic *cell traffic blocking model* and a *quality of service evaluation* model. The cell traffic characterising model is employed to determine the mean and variance of hand-off traffic offered by a cell to its neighbour. The cell traffic-blocking model is used to determine the blocking experienced by the various traffic streams offered to each cell. The quality of service evaluation part is essentially a *fixed-point iteration* of the cell traffic characterising and cell traffic blocking parts to determine customer orientated grade-of-service parameters such as blocking and dropping probabilities. The thesis presented detailed mathematical models for each of the three parts.

In chapter 4 we analysed a simple two-cell scenario and presented results for the mean and variance of traffic offered by a cell to its neighbour in this two-cell scenario. We considered the following four different channel holding time distributions in our two-cell scenario:

- Negative exponential distribution
- Deterministic distribution
- Det-neg distribution
- Gamma distribution

The results that we presented for the neg-exp and deterministic distribution are exact whereas the results that we presented for the det-neg and gamma distribution are approximate, based on asymptotic analysis.

In chapter 5, we presented our proposed two-moment cell traffic blocking models for use in the performance analysis of cellular networks. The inputs to these cell traffic-blocking models are the mean and variance of the various offered traffic streams obtained from the cell traffic characterising model. The outputs are the blocking probabilities experienced by the various traffic streams and the moments of the respective carried traffic streams. We first examined the following two blocking models from the fixed network arena:

- Sander's et al's two moment blocking model
- Delbrouck's two moment blocking model

We extended the above models to apply for the cellular network scenario where two distinct traffic streams, namely, fresh and hand-off traffic, are offered to a cell and where channel reservation may be applied to protect hand-off traffic at the expense of fresh traffic.

In chapter 6, we presented the Quality of Service evaluation algorithm. This algorithm employed the cell traffic characterising model that we presented in chapter 4 and the cell traffic blocking model that we presented in chapter 5 in an iterative procedure to determine customer orientated grade-of-service parameters such as new call blocking and forced termination probabilities. The end product of the work that we presented in chapters 4, 5 & 6 is a performance analysis model that is based on Poisson new call arrivals, generalised hand-off call arrivals and suitable channel holding times such as the negative exponential distribution, the det-neg distribution and the gamma distribution.

In chapter 7, we presented different approaches to mobility modelling. This included an analytic mobility model for a highway vehicular scenario, where exact cell dwell time distributions for vehicular travel were derived under simple mobility assumptions. The thesis also presented a practical approach to mobility modelling in established cellular networks where appropriate mobility-related measurements from the network may be employed as a simpler alternative to deriving exact cell dwell time distributions.

The thesis provided extensive results to validate the proposed analysis. We showed that the proposed two-moment analysis is superior to the classical single moment analysis when compared to simulation results. The single moment method makes use of Poisson hand-off arrivals and negative exponential service time distributions. Because the Poisson process and the negative exponential distribution are one-parameter models, they are unable to capture any information beyond the mean of the process that they are chosen to model. The accuracy of the proposed analysis may be attributed to the fact that it models the channel holding time distribution as a more generalised distribution. Thereafter, the proposed model characterises the resulting hand-off traffic using its mean and its variance. Our generalisation of these arrival and service processes has not been accompanied by the state space explosion, previously found in the literature, when similar generalisations have been attempted. The ability to accommodate more accurate arrival and service processes without any added cost in tractability and computational effort is the true strength of our proposed model.

Since the write-up of this thesis, the work presented in this thesis has been extended in two significant directions:

- In a recently published paper [73], we consider the extension of the proposed two moment performance analysis algorithm from the Poisson (infinite population) new call arrival process to the more suitable Engset (finite population) new call arrival process. The Engset new call arrival process is more realistic for the analysis of micro-cellular and pico-cellular networks. The accommodation of the Engset new call arrival process requires modification to both our cell traffic characterising model and our cell traffic blocking models. Preliminary analysis of our two-cell scenario under Engset new call arrivals can be found in [73].
- With the growth in data traffic and the advent of GPRS packet services [12] on the GSM cellular network, we are extending the analysis in this thesis to apply for the scenario where circuit-switched voice calls and packet switched data calls compete for the same resources on the air-interface. The incorporation of packet-switched traffic requires modifications to our performance analysis algorithm and our initial foray in this matter can be found in [74].

THE END

APPENDIX A

In this section, we analyse the carried traffic in a $E_2/E_2/\infty$ queuing system. We show that for identical arrival and service distributions that are Erlang-2 distributed, the mean carried traffic is 1 Erlang and the variance of carried traffic is 0.75 Erlang.

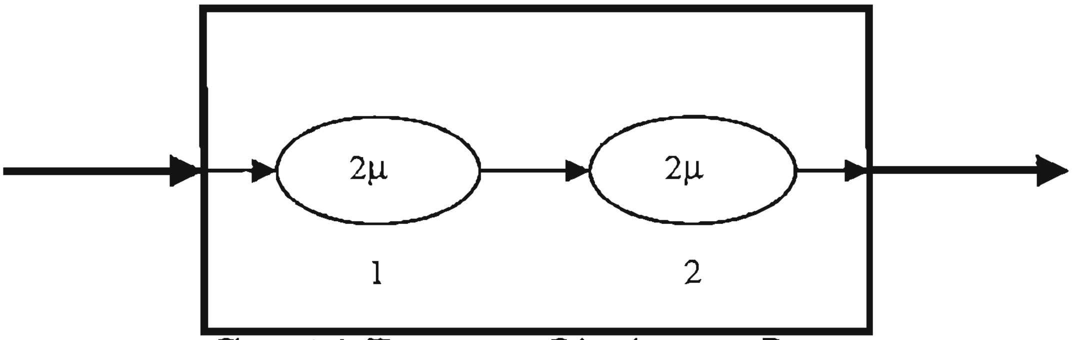


Figure A-1: The two-stage Erlangian server E_2.

Figure A-1 illustrates a two stage Erlangian system. The large rectangular structure could represent either an arrival process or a service facility. For explanation purposes, we consider it to be a service facility. The internal structure of this service facility is a series or tandem connection of two single stage exponential server systems with rate 2μ. A call that undergoes the two stage Erlangian service process, in fact undergoes two tandem single stage exponential processes. The second stage starts as the call finishes its first stage. It is not difficult to show that the mean service time of the single stage exponential server system is $1/2\mu$ and the mean service time of the two-stage Erlangian server E_2 is $1/\mu$.

Let *a(t)* represent the pdf of the inter-arrival time of a two-stage Erlangian arrival process where the mean arrival rate is λ. Let *b(t)* represent the pdf of the service time of a two-stage Erlangian service process where the mean service rate is μ. Hence,

$$a(t) = 2\lambda(2\lambda t)e^{-2\lambda t} \quad \text{(A.1)}$$

$$b(t) = 2\mu(2\mu t)e^{-2\mu t} \quad \text{(A.2)}$$

We now consider a Markov model for the $E_2/E_2/\infty$ queuing system. Let $p_{j,k,m}$ represent the probability of being in state $[j,k,m]$ of the system. The subscript j represents the stage of the call arrival process and therefore can take the values 1 or 2. The subscript k represent the number of calls in stage 1 of their two-stage Erlangian service process and the subscript m represent the number of calls in stage 2 of their two-stage Erlangian service process. The subscripts k and m may take any integral values from 0 to infinity. The state equations for the above system may be derived under the assumptions that state transitional rates to and from a particular state $[j,k,m]$ are equal.

$$\text{for } j = 1 : (2\lambda + 2k\mu + 2m\mu)\, p_{1,k,m} = 2\lambda\, p_{2,k-1,m} + 2(k+1)\mu\, p_{1,k+1,m-1} + 2(m+1)\mu\, p_{1,k,m+1} \qquad \textbf{(A.3)}$$

$$\text{for } j = 2 : (2\lambda + 2k\mu + 2m\mu)\, p_{2,k,m} = 2\lambda\, p_{1,k,m} + 2(k+1)\mu\, p_{2,k+1,m-1} + 2(m+1)\mu\, p_{2,k,m+1} \qquad \textbf{(A.4)}$$

Note that in the above equations we assume that $p_{j,k,m}$ equals zero for all $k<0$ and/or $m<0$. Using the probability generating function $G_j(z,y)$, where,

$$G_j(z,y) = \sum_{k=0}^{\infty}\sum_{m=0}^{\infty} p_{j,k,m}\; z^k y^m \qquad \textbf{(A.5)}$$

we may convert equation (A.3) to the following form:

$$(z-y)\frac{\partial G_1(z,y)}{\partial z} + (y-1)\frac{\partial G_1(z,y)}{\partial y} + A\; G_1(z,y) = A\; z\, G_2(z,y) \qquad \textbf{(A.6)}$$

where we set $A = \lambda/\mu$. Using a change of variable, x & w, where $x = z-1$ and $w = y-1$:

$$(x-w)\frac{\partial G_1(x+1,w+1)}{\partial x} + w\frac{\partial G_1(x+1,w+1)}{\partial w} + A\; G_1(x+1,w+1) = A\,(x+1)\,G_2(x+1,w+1) \qquad \textbf{(A.7)}$$

Since the PGF is related to the Binomial Moment Generating function $F_j(x,w)$ as follows [17, pp. 461]:

$$F_j(x,w) = G_j(x+1,w+1) \qquad \textbf{(A.8)}$$

$$\text{where } F_j(x,w) = \sum_{n=0}^{\infty}\sum_{p=0}^{\infty} \beta_{j,n,p}\; x^n w^p \qquad \textbf{(A.9)}$$

and $\beta_{j,n,p}$ is the n^{th} and p^{th} multinomial moment of the number of calls in the first stage and second stage of their service processes in the $E_2/E_2/\infty$ queuing system when the arrival process is in the jth stage. Equation (A.7) becomes:

$$(x-w)\frac{\partial F_1(x,w)}{\partial x} + w\frac{\partial F_1(x,w)}{\partial w} + A\; F_1(x,w) = A\,(x+1)\,F_2(x,w) \qquad \textbf{(A.10)}$$

Similar manipulations of equations (A.4) gives equation (A.11):

$$(x-w)\frac{\partial F_2(x,w)}{\partial x}+w\frac{\partial F_2(x,w)}{\partial w}+A\,F_2(x,w) = A\,F_1(x,w) \quad \textbf{(A.11)}$$

Substituting (A.9) into equations (A.10-A.11) and equating coefficients of equal powers of x and w give different equations relating the different Binomial moments. Assuming A=1 (identical arrival and service processes) and since the total number of calls in progress in the $E_2/E_2/\infty$ queuing system is simply $k+m$, the mean M and the V of the number of calls in progress in the $E_2/E_2/\infty$ queuing system is:

$$M = \sum_{j=1}^{2}\beta_{j,0,1} + \beta_{j,1,0} = 1 \quad \textbf{(A.12)}$$

$$V = 2\beta_2 - M^2 + M = \sum_{j=1}^{2} 2\beta_{j,2,0} + 2\beta_{j,0,2} + 2\beta_{j,1,1} - 1 + 1 = 0.75 \quad \textbf{(A.13)}$$

REFERENCE

[1] Siegmund M. Redl, Matthias K. Weber, Malcolm W. Oliphant, "An introduction to GSM", *Artech House Mobile Communications Series*, 1995.

[2] E. Del Re, R. Fantacci, and G. Giambene, "Handover and Dynamic channel allocation techniques in mobile cellular networks," *IEEE Trans. on Vehic. Tech.*, vol. 44, No. 2, pp 229-237, May 1995.

[3] D. Everitt, D. Mansfield, "Performance analysis of cellular mobile communications systems with dynamic channel assignment" *IEEE J. Select. Areas Comm.*, vol 7, no.8, pp1172-1180, Oct 1989.

[4] D.C Cox, D.O. Reudnik, "Dynamic channel assignment in high capacity mobile communications systems," *Bell System Technical Journal*, vol. 50, no. 6, pp 1833-1857, July-August 1971.

[5] M. Zhang and P. Tak-Shing Yum, "Comparisons of channel assignment strategies in cellular mobile telephone systems", *IEEE Transactions on Vehicular Technology*, vol. 38, no. 4, pp 211-215, Nov 1989.

[6] S.M. Elnoubi, R.Singh, and S.C. Gupta, "A new frequency channel assignment algorithm in high capacity mobile communications systems, "*IEEE Transactions on Vehicular Technology*, vol VT-31, no. 3, pp 125-131, Aug 1982.

[7] S. A. El-dolil, W-C Wong, R. Steele "Teletraffic performance of highway micro-cells with overlay macro-cell" *IEEE J. on Selected Areas in Comms.*, January 1989, Vol. 7, No. 1, pp. 71-78.

[8] M. Madfors, K. Wallstedt, S. Magnusson, H. Olofsson, P. Blackman, S. Engstrom " High Capacity with Limited Spectrum in Cellular systems" *IEEE Communications Magazine*, Aug 1997, Vol 35, No.8. pp.38-45.

[9] Lon-Rong Hu., and Rappaport, S.: "Personal Communication Systems Using Multiple Hierarchical Cellular Overlays", *IEEE Journal on Selected Areas in Comms.*, February 1995, Vol. 13, No. 2, pp. 406-415.

[10] X. Lagrange "Multitier Cell Design" *IEEE Communications Magazine*, Aug 1997, Vol 35, No.8 pp 60-64

[11] K.L. Yeung and S. Nanda "Channel Management in Microcell/Macrocell Cellular Radio Systems" *IEEE Transaction on Vehicular Technology*, vol 45, no. 4, 1996, pp 601-612.

[12] Gotz Brasche and Bernhard Walke, "Concepts, services and protocols of the new GSM Phase 2+. General Packet Radio Service", *IEEE Communications Magazine*, pp 94-102, Vol. 35, Number 8, August 1997.

[13] P.M. Lin, BJ Leon, CR. Stewart, "Analysis of circuit-switched networks employing originating office control with spill, " *IEEE Transactions on Communications*, pp754-765, 1978.

[14] Girard, A., *Routing and dimensioning in circuit-switched networks*, Addison-Wesley, Quebec, 1992.

[15] Hong, D., and Rappaport, S.: 'Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and non-prioritized hand-off procedures', *IEEE Trans. on Vehicular Technology*, VT-35, pp. 77-92, August 1986.

[16] D.Macmillan, "Traffic Modelling and Analysis for Cellular Mobile Networks" *Proc of 13th ITC*, pp 627-632,1991

[17] R.Guerin, "Channel Occupancy time distribution in a cellular Radio System" *IEEE Transactions on Vehicular Technology*, VT-35, No. 3, pp. 89-99, Aug 1987.

[18] Zonoozi, M. M., and Dassanayake, P., "User mobility modelling and characterization of Mobility Patterns," *Journal on Select. Areas in Comm.*, Volume 15, Number 7, pp. 1239-1252, September 1997.

[19] E. Chlebus, W. Ludwin, "Is Hand-off Traffic Really Poissonian?" *Proc ICUPC 95*, Tokyo, Japan, pp 348-353, Nov 6-10, 1995.

[20] KK Leung, W.A. Massey and W Whitt, "Traffic Models for Wireless Communication Networks" *IEEE Journal on Selected Areas in Communications*, Vol 12, No. 8, pp. 1353-1364, October 1994.

[21] M Sidi, D Starobinski, "New call blocking versus Hand-off Blocking in Cellular Networks" *Proc. IEEE INFOCOM'96*, pp 35-42, 1996.

[22] P. Tran-Gia and M. Mandjes. "Modelling of customer retrial phenomenon in cellular mobile networks" *IEEE Journal on Select. Areas in Communications*, pp 1406-1414, vol 15, No. 8, Oct 1997.

[23] Y. Fang, I. Chlamtac, and Y.-B. Lin, "Call Performance for a PCS network" *IEEE Journal on Selected Areas in Communications*, pp 1568-1581, vol 15, No. 8, Oct 1997.

[24] B. Jabbari and W.F. Fuhrmann, "Tele-traffic modelling and analysis of flexible hierarchical cellular networks with speed sensitive hand-off strategy" *IEEE Journal on Selected Areas in Communications*, pp 1539-1548, vol 15, No. 8, Oct 1997.

[25] P. Fitzpatrick, C.S. Lee, and B. Warfield, "Tele-traffic performance of Mobile Radio Networks with Hierarchical cells and overflow" *IEEE J. Sel. Areas in Comm.*, pp 1549-1557, vol 15, No. 8, Oct 1997.

[26] Nokia, "Nokia Technical Documentation: Software release BSS9", Finland, 1999.

[27] Rappaport S. S., "Blocking, hand-off and traffic performance for cellular communications with mixed platforms," *Proc. of Institute of Electrical Engineers (IEE)*, vol 140, No. 5, pp 389-401, 1993.

[28] P. V. Orlik, and S. S. Rappaport, "A model of Teletraffic Performance and Channel holding time characterisation in Wireless Cellular Communication with General Session and Dwell Time Distributions," *IEEE Journal on Selected Areas in Communications*, Vol 16, No. 5, pp. 788 - 803, June 1998.

[29] Rappaport S.S., "The multiple call hand-off problem in high capacity cellular communication systems," *IEEE Transactions in Vehicular Technology*, vol 40, pp546-557,1991

[30] P.V. Orlik, and S.S. Rappaport, "Traffic performance and mobility modelling of cellular communications with mixed platforms and highly variable mobilities", *Proceedings of the IEEE*, Vol. 86, No.7, July 1998.

[31] C. Jedrzycki and V.C.M. Leung, "Probability distribution of channel holding time in cellular telephony systems" *Proceedings of IEEE Vehicular Technology conference*, Atlanta, Georgia, May 1996, pp 247-251.

[32] J.I. Sanchez, F. Barcelo and J. Jordan, "Inter-arrival time distribution for channel arrivals in cellular telephony", Proceedings of the 5th International Workshop on Mobile Multimedia Communication MoMuc'98, October 12-13, 1998, Berlin.

[33] F. Barcelo and J. Jordan, "Channel holding time distribution in cellular telephony", Electronics Letters, Vol. 34, No. 2, pp 146-147, 1998.

[34] E. Chlebus, "Empirical validation of call holding time distribution in cellular Communications systems", Proceedings of the 15th ITC, Vol 2b, pp 1179-1188, Elsevier,1999.

[35] S. Tekinay, B. Jabbari, "Handover and Channel Assignment in Mobile Cellular Networks", IEEE Communications Magazine, pp. 42-46, Nov. 1991.

[36] S. Tekinay, B. Jabbari, "A Measurement-Based Prioritisation Scheme for Handovers in Mobile Cellular Networks", IEEE Journal on Selected Areas in Communications, Vol. 10, No 8, pp. 1343-1350, Oct.1992.

[37] K.C. Chua, B. Bensaou, W. Zhuang, S.Y. Choo, "Dynamic channel reservation (DCR) scheme for handoffs prioritisation in mobile micro/picocellular networks", Proc of IEEE Int. Conf. on Universal Personal Communications, pp. 383-387, 1998.

[38] N.D. Tripathi, J.H. Reed, H.F. VanLandingman, "Adaptive handoff algorithms for cellular overlay systems using fuzzy logic", Proc of IEEE the Vehicular Technology Conf. (VTC99), 1999.

[39] K. Alagha, D. Zeghlache, "A Communications Scheme for Adaptive Resource Allocation in Wireless Networks", Proc of IEEE Int. Conf. on Universal Personal Communications (ICUPC98), pp. 65-69, 1998.

[40] A. Chandra, D. Bansal, A.R. Shorey, "A new priority based dynamic handoff algorithm minimizing unnecessary handoffs in cellular systems", Proc of IEEE the Vehicular Technology Conf. (VTC99),1999.

[41] Ross S. M., "Stochastic processes", Wiley Series in probability and mathematical statistics, John Wiley and Sons, 1983.

[42] Takacs L., "Introduction to the Theory of Queues", University Texts in the mathematical sciences, Oxford University Press, 1962.

[43] Chlebus, "Grade of service determination in cellular mobile networks based on conservation of traffic flow," Electronic Letters, 30 (21), 1741-1742,1994

[44] Akinpelu, "The overload performance of Engineering networks with Non-Heriarchical and Hierarchical routing, ITC10, Montreal 1982.

[45] L.E.N. Delbrouck, "A unified approximate Evaluation of Congestion Functions for Smooth and Peaky Traffic", *IEEE Transaction on Communications*, Vol. COM-29, No.2, pp 85-91, Feb 1981.

[46] Akimaru H. & Kawashima K., "Teletraffic Theory and Applications", Springer-Verlag, 1993.

[47] Leonard Kleinrock, "Queuing systems, Volume 1: Theory", *John Wiley and Sons*, 1975

[48] F Abrishamkar, Z. Siveski "PCS Global Mobile Satellites" *IEEE Communications Magazine*, Sep 1996, Vol 34, No.9 pp 132-136.

[49] A Ganz, Y. Gong, Bo Li, "Performance study of Low Earth Orbit Satellite Systems" *IEEE Transactions on Communications*, Vol 42, No. 2,3,4, Feb/March/April 1994.

[50] H. Uzunalioglu, W Yen., "Managing Connection Handover in Satellite Networks" *Proc IEEE Globecom'97*, pp 1606-1610.

[51] Kuczura A., "The interrupted Poisson process as an overflow process", Bell System Technical Journal, volume 52, pp 437-448, March 1973.

[52] Yum and Leung "Blocking and handoff performance analysis of directed retry in cellular mobile systems" IEEE Transactions on vehicular technology, volume 44, no 3. Pp 645-655, August 1995.

[53] Y. Fang, I. Chlamtac, and Y.B. Lin, "Modelling PCS networks under general call holding time and cell residence time distributions," *IEEE Transactions on Networking*, Vol 5, No. 6, pp 893-906, 1997.

[54] Y. Fang, I. Chlamtac and Y.B. Lin, "Hand-off Probability in PCS networks", *Proceedings of the IEEE ICUPC'98 conference*, Vol 1, pp 765 - 768, Oct 1998.

[55] Lin P, Lin Y-B, and Jeng J-Y, "Improving GSM call completion by call reestablishment" IEEE Journal on Selected Areas on communications, Volume 17, number 7, pp 1305-1317, July 1999.

[56] Li W. and Alfa A.S., "A PCS network with correlated arrival process and splitted-rating channels" IEEE Journal on Selected Areas on communications, Volume 17, number 7, pp 1318-1325, July 1999.

[57] Neuts M., "Matrix-Geometric Solutions in Stochastic Models", Baltimore-MD: John Hopkins University press, 1981.

[58] Rajaratnam, M., and Takawira, F., "Performance Analysis of Highway Cellular Networks using generalised arrival and generalised service time distributions", *Proceeding of International Teletraffic Congess (ITC-16)*, Vol. 3a: Teletraffic Science and Engineering, pp 11-22, June 1999, Edinburgh, Scotland.

[59] Syski, R., 'Introduction to congestion theory in telephone systems', North-Holland, 1986.

[60] Rajaratnam, M., and Takawira, F., "Hand-off traffic modelling in cellular networks", *Proceedings of IEEE Global Communications Conference (GLOBECOM'97)*, Phoenix, Arizona, pp. 131-137, Nov 1997.

[61] Rajaratnam, M., and Takawira, F., "The two moment performance analysis of cellular mobile networks with and without channel reservation," *Proc IEEE Conference on Universal Personal Communications (ICUPC'98)*, pp. 1157-1161, Florence, Oct 98.

[62] Rajaratnam, M., and Takawira, F., "Non-classical traffic modelling and performance analysis of cellular mobile networks with and without channel reservation," *IEEE Transactions on Vehicular Technology*, Vol. 49, Number 03, pp 817-834, May 2000.

[63] Rajaratnam, M., and Takawira, F., " Hand-off traffic characterisation in cellular networks under non-classical arrivals and service time distributions," *submitted for IEEE Transactions on Vehicular Technology*

[64] Gross D.,Harris CM., *Fundamentals of Queuing Theory*, John Wiles and sons, New York, 1998

[65] Rajaratnam, M., and Takawira, F., "Asymptotic approximation for hand-off traffic characterisation in cellular networks under non-classical arrival and service time distributions", *Proceedings of the ITC mini-seminar*, pp 95-106, Lillehammer, Norway, Mar 2000.

[66] Rajaratnam, M., and Takawira, F., "Hand-Off Traffic Characterisation In Cellular Networks Under Non-Classical Arrivals And Gamma Service Time Distributions", *Proceedings of the IEEE Personal, Indoor, Mobile Communications conference PIMRC 2000 conference*, pp. 1635-153, Volume 2, September 2000, London.

[67] B. Sanders, W.H.Haemers, R.Wilcke, "Simple Approximation techniques for congestion functions for smooth and peaked traffic" *Proc. International Teletraffic Congress (ITC) #10*, pp 4.4.b.1-4.4.b.7, Montreal, June 1983.

[68] Jagermann, D.L. "Some properties of the Erlang-loss function" *Bell System Technical Journal*, vol. 53, pp525-551, 1974.

[69] Krupp, R.S. "Stabilisation of alternate routing networks, *International conference on communications*, pp 31.2.1-31.2.5, 1982.

[70] A. Girard, "Blocking probability of non-integer groups with trunk reservation" IEEE Transactions on communications, Vol. COM-33, No. 2, pp 113-120, February 1985.

[71] Cooper R.B., *Introduction to Queuing Theory*, Macmillan, 1972.

[72] M. Rajaratnam & F.Takawira, "Practical Approach to mobility modelling in cellular mobile networks", SATNAC'98, pp456-465, Capetown, September 1998

[73] Rajaratnam, M., and Takawira, F., "Hand-off traffic characterisation in cellular networks under Engset new call arrivals and generalised service times" *Proceedings of IEEE Wireless & Networking Conference (WCNC'99)*, pp. 423-427, New Orleans, September 1999.

[74] Rajaratnam, M., and Takawira, F., "*Performance analysis of the radio layer in GSM phase 2+ (GPRS) networks under voice and data traffic*" submitted for IEEE Conference on Communications (ICC) 2001, Helsinki, Finland. Available by email request to myuran@ieee.org.