MODELING AND PERFORMANCE EVALUATIONS OF TELETRAFFIC IN CELLULAR NETWORKS

by

Emre Altug Yavuz

B. Sc., Middle East Technical University, 1995
M. Sc., Middle East Technical University, 1998

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Electrical and Computer Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA

February 2007

ABSTRACT

The growing interest for cellular technology has motivated operators to provide a wide variety of services from conventional circuit-switched voice to packet-switched data and multimedia applications. Providing these services anytime and anywhere is challenging due to not only frequent status changes in network connectivity, but also limited resources such as bandwidth. Different priorities are assigned to services to satisfy diverse QoS requirements. Call admission control schemes have been proposed to manage resources by selectively limiting the number of admitted calls to ensure that QoS measures such as call blocking/dropping probabilities stay within acceptable limits. Exact analysis methods based on multidimensional Markov chain models are used to evaluate performance of these schemes, yet they suffer from curse of dimensionality that results in very high computational cost. Large sets of equations are avoided using approximation methods based on one dimensional Markov chain models assuming that channel occupancy times are exponentially distributed with equal mean values and all calls require equal capacities. Existing approximation methods lead to significant discrepancies when average channel occupancy times differ. We propose a novel performance evaluation approximation method, *effective holding time*, with low computational complexity to relax this assumption.

In multi-service networks, voice is accompanied by data and multimedia applications that require distinct capacities. When capacity requirements differ, existing approximation methods based on one dimensional Markov chain models become inaccurate if not obsolete. We propose a computationally efficient approximation method, *state space decomposition*, to relax this assumption. Numerical results show that the proposed method provide highly accurate results that match well with exact solutions.

Traffic statistics are essential to understand the distribution of idle periods of voice channels to overlay packet-switched services on circuit-switched technology and to feed simulations with realistic data. Call holding and channel occupancy times are key elements for computing performance metrics such as call blocking/dropping probabilities. We present an empirical approach to determine the distribution of call holding and channel occupancy times. We show that lognormal distribution is the closest fitted candidate to approximate channel occupancy times and call holding times for stationary/mobile users along with the number of handoffs committed by a user.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

C	Number of channels in a cell.
c	Number of occupied channels in a cell.
K	Threshold for new call bounding scheme.
m	Threshold for the cutoff priority scheme.
λ_n	Arrival rate for new calls.
λ_h	Arrival rate for handoff calls.
λ_{np}	Arrival rate for non-prioritized calls.
λ_p	Arrival rate for prioritized calls.
$1/\mu_n$	Average channel holding time for new calls.
$1/\mu_h$	Average channel holding time for handoff calls.
$1/\mu_{np}$	Average channel holding time for non-prioritized calls.
$1/\mu_p$	Average channel holding time for prioritized calls.
$1/\mu_{eff}$	Average effective channel holding time.
b_{np}	Capacity requirement in bandwidth units for non-prioritized calls.
b_p	Capacity requirement in bandwidth units for prioritized calls.
ρ_n	Traffic intensity for new calls (i.e., λ_n / μ_n).
ρ_h	Traffic intensity for handoff calls (i.e., λ_h / μ_h).
ρ_{np}	Traffic intensity for non-prioritized calls (i.e., λ_{np} / μ_{np}).
ρ_p	Traffic intensity for prioritized calls (i.e., λ_p / μ_p).
n_{np}	Number of non-prioritized calls in the system.
n_p	Number of prioritized calls in the system.
p_{nb}	Blocking probability for new calls.
p_{hd}	Dropping probability for handoff calls.
p_{nb}^{a}	Blocking probability for new calls from the normalized approximation.
p_{hd}^{a}	Dropping probability for handoff calls from the normalized approximation.

p^{t}_{nb} Blocking probability for new calls from the traditional approximation.

p^{t}_{hd} Dropping probability for handoff calls from the traditional approximation.

p^{eff}_{nb} Blocking probability for new calls from the proposed approximation.

p^{eff}_{hd} Dropping probability for handoff calls from the proposed approximation.

β_i User defined admission probability for non-prioritized (new) calls in a call admission control scheme.

$q(c)$ Equilibrium channel occupancy probability when c channels are occupied.

$\hat{q}(c)$ Approximated equilibrium channel occupancy probability when c channels are occupied.

B_{np} Blocking probability for non-prioritized calls.

B_p Blocking probability for prioritized calls.

k_j Admission probability for prioritized calls.

h_r Admission probability for non-prioritized calls.

$q_p(j)$ Equilibrium channel occupancy probability when j prioritized calls exist in the system.

$q_{np}(r)$ Equilibrium channel occupancy probability when r non-prioritized calls exist in the system.

$\hat{q}(n_{np}, n_p)$ Estimated equilibrium channel occupancy probability for n_{np} non-prioritized and n_p prioritized calls.

ACKNOWLEDGEMENTS

Yes ! A fierce and savage wind tore at us.

We were on top of Annapurna ! 8,075 meters ...

Our hearts overflowed with an unspeakable happiness.

"If only the others could know ... "

If only everyone could know !

Maurice Herzog on summiting Annapurna

To me, writing a Ph.D. thesis is like climbing a high altitude mountain for the first time. Often you do not have any spectators and acknowledgement is not acquired unless you make it to the top. The graduate level courses guide you while you are on your way to the base camp and every paper you manage to publish afterwards means reaching higher camps, taking you closer to the summit for the final push yet. Neither everybody climbs the same mountain nor under the same conditions. That is why climbing with the right guide is important not only to pick the right route on the way to the top but also to know where your high altitude camps shall be set up. Sometimes the conditions and the difficulty of the route force you to back off only to return later knowing that what does not kill makes stronger. Yet, from an outsider's point of view what matters most is whether you climb it or not regardless of what happened on the way.

This is a self-guided journey where a passionate climber never gives up for the sake of learning about one's self and limits. However, choosing a competent guide is not only vital to avoid storms, climbing to dead ends and being left to destiny when in trouble but also to lift the spirits up when needed. I am thankful to my perseverance and diligence for not only making it to the top through storms and turmoil but also getting myself back on the right track despite surrounded by a thick fog at most times. With a little twist on what Hillary said when he climbed Mount Everest, "It wasn't the title I was after, it was to conquer myself."

I dedicate this work to my parents for their unconditional love and support. I am thankful for their unlimited patience and confidence in me during this arduous climb. I would not be able to complete my graduate studies without their dedicated sacrifice, understanding, and encouragement. I am grateful to my friends; Aliye, Görkem, Kaan, Rübab and Verda, who helped me in many ways that kept me climbing in good times and bad. Last but not the least I would like to express my gratitude to my supervisor Dr. Victor C. M. Leung for his guidance, comments and financial support.

To conclude, I would once more like to quote Maurice Herzog as he says,

"There are other Annapurnas in men's life."

To my mom, Semra, and my dad, Hüseyin,

for their unconditional love and absolute support …

Co-Authorship Statement

I am the first author and principle contributor of all manuscript chapters. All manuscript chapters are co-authored with V.C.M. Leung, who co-supervised the thesis research.

CHAPTER 1 INTRODUCTION

Cellular network technology is one of the fastest growing ways of mobile communications today. The bounds of an existing communication network infrastructure have been extended by cellular technology via connecting mobile units to public network operated by the local exchange or long distance carriers to make special features and functions specific to both cellular and public networks available to all users. Global standards have been developed to provide voice and data services anytime and anywhere regardless of user mobility while satisfying their diverse Quality of Service (QoS) requirements. Radio resources such as bandwidth, transmit power, channel codes and base stations are generally limited in cellular networks due to physical and regulatory restrictions as well as the interference-limited nature of the cellular structure. Efficient management of these resources is not only crucial to the efficiency of system operation and congestion prevention, but it is also very important to satisfy the QoS requirements of user applications. Services with specific data rate, bandwidth, power and latency requirements need specific amount of system resources to be allocated at the time of call admission to ensure that these requirements are attained and sustained during communication. Significant challenges confronted by cellular networks due to frequent status changes in connectivity and highly variable noisy communication channels can be overcome by QoS provisioning which has become one of the most demanding problems.

Resource management requires more sophisticated techniques in a mobile environment than those used in fixed systems since a blind spot may be reached where QoS is severely limited due to a weak signal or a loss of communication may occur during a handoff when the new point of attachment may not be able to provide resources similar to the old one. Nevertheless, the problem of maintaining service continuity for users' applications during handoffs has been intensified with the increasing number of microcells and picocells in cellular networks. Call admission control (CAC) schemes have been developed to manage scarce radio resources to maximize network utilization by selectively limiting the number of admitted calls. Probabilities of call blocking and dropping are two important QoS measures used when evaluating performance of call admission control schemes. Call blocking occurs when a cellular network is unable to assign network resources to enable a call initiated in a cell, whereas call dropping occurs when a cellular network is unable to assign network resources to enable a call handed off from a neighboring cell. Sufficient network resources shall be provided to ensure that call blocking and dropping probabilities stay within acceptable limits for user applications. A higher priority is normally assigned to handoff calls over the new ones to minimize call dropping probability since dropping an on-going call is generally more objectionable to a user than blocking a new call request. However reducing blocking probabilities of calls with higher priorities increases the probability of blocking for calls with relatively lower priorities resulting in a trade off between both types of calls. Therefore the goal is to sustain a balance between calls of different priorities while satisfying the respective QoS requirements.

Call admission control has been intensively studied in the past [1][2] and many priority-based CAC schemes have been proposed [3]-[18]. One dimensional Markov chain models are commonly used to evaluate these schemes (e.g., [6] [7] [15]) assuming that call requests that originate from different types of users are independently Poisson distributed, channel occupancy times for each call are exponentially distributed with equal mean values and each call requires an equal channel capacity. Yet these assumptions may not be appropriate since calls with different priorities, such as new and handoff, may have different average channel occupancy times if not different distributions as shown in [19], [20] and references therein. Existing performance evaluation approximation methods based on one dimensional Markov chain modeling lead to significant discrepancies when average channel occupancy times for different call types are not equal [21]. Gersht and Lee proposed an iterative algorithm in [22] by modifying the approximation suggested by Roberts in [23] to improve its accuracy when service rates differ. However the algorithm is only accurate when appropriate initial values are chosen and therefore may not be competent [21]. Li and Chao obtained a product form solution in [24] by modeling a multicell network as a network of queues employing a hybrid guard channel/queuing priority scheme with transfer of unsuccessful requests to neighboring cells. Their solution is restrictive to the protocol considered and therefore may not be appropriate to be used for the performance evaluation of call admission control schemes in general. Thus, exact analysis methods based on multidimensional Markov chain models appeared to be the only means to obtain accurate solutions for evaluating call admission control schemes. Rappaport obtained call blocking probabilities for calls with various priorities by using a multidimensional model of a cellular network in [25]. Rappaport and Monte developed an analytical model for traffic performance analysis using a multidimensional birth death process in [26] by considering the effects of various platform types distinguished by different mobility characteristics on performance. Even though multidimensional Markov chain models are capable of providing the exact solutions, these methods suffer from the curse of dimensionality, which results in very high computational cost for large systems. An easy to implement analytical approximation method with highly accurate solutions and low computational cost is needed to compute new/handoff call blocking/dropping probabilities of calls for several widely known call admission control schemes under more general assumptions.

The fast evolution of cellular networks has been accompanied not only by basic voice services but also by development, growth and use of a wide variety of network applications that range from text-based utilities such as SMS messaging, file transfer, remote login and electronic mailing to multimedia utilities such as video conferencing and streaming, web surfing and electronic commerce. This has motivated cellular network operators to move away from just providing conventional voice services to embracing a wide variety of traffic from data to multimedia applications due to increasing demand coming from users that these services shall also be available on the move. Different techniques have been proposed to allocate limited and varying network resources efficiently to a variety of services with different characteristics and QoS requirements at different network layers. [27]. Modulation and power control schemes are designed to be QoS aware at the physical layer [28] as medium access control is adjusted to support reservations and QoS guarantees at the data link layer. At the network layer, techniques of mobility management and seamless connectivity, including the extension of routing mechanisms to be QoS aware and able to handle mobility, are applied [29]. Multimedia coding systems such as H.263L video codecs are introduced for the application layer [30].

Call admission control schemes are analyzed using Markov chain models based on circuit-switched network architectures, however conventional circuit-switched services such as voice are gradually being replaced by packet-switched data and multimedia applications.

Conventional call admission control schemes will continue to be useful when applied with suitable scheduling techniques to guarantee QoS at the packet level since most data and multimedia applications are inherently connection oriented and packet-switched connections can be provisioned to their effective bandwidths [31]-[33]. Effective bandwidth represents the physically dedicated bandwidth of a packet-switched connection to match its overall traffic demand. Markov chain modeling will still be useful since a cellular network can have a similar form to a circuit-switched network operating with fixed routing [32]. However calculating channel occupancy distribution of such a multi-service cellular network using an exact solution method based on multidimensional Markov chain modeling involves numerically solving the balance equations, which is demanding for all but smallest channel capacities in the absence of a product form solution. Existing approximation methods based on one dimensional Markov chain modeling, on the other hand, become obsolete when packet-switched connections such as data and multimedia applications have distinct capacity requirements. In [34], Borst and Mitra developed computational algorithms for a multi-service cellular network by coupling the computation of joint channel occupancy probabilities with that of used capacity assuming that channels are occupied independently. The authors solved the balance equations through numerical iteration but the results can only be comparative when the number of existing call arrival types are high due to authors' channel occupancy independence assumption. A novel performance evaluation approximation method with highly accurate solutions and low computational cost is needed to compute call blocking probabilities of circuit and packet-switched connection type of calls that have distinct capacity requirements for several widely known call admission control schemes under more general assumptions.

The advantage of having packet-switched connection type of services overlaid on circuit-switched technology over the same air interface is the utilization of excess network capacity available in each cell. When large numbers of sources with bursty characteristics are multiplexed, it is unlikely that all of them transmit at their peak rates at the same time. The network can then allocate each user less resource than the corresponding requested peak capacity while meeting the statistical performance requirements. Data packets can be transmitted over radio interface using statistical multiplexing to provide a QoS level comparable to that of circuit-switched services. Statistical multiplexing gain arises from the talk spurt to silence ratio found in speech which makes it possible to multiplex more than one service on to the same radio channel. However accurate voice traffic statistics are needed to understand the length and frequency distributions of idle periods of cellular channels assigned to voice in order to exploit the statistical multiplexing gain in cellular networks.

Traffic statistics are important for network management and optimization along with traffic modeling, billing and allocation of safety buffers. These statistics are also used to evaluate performance during network simulation or analysis using mathematical models. Two of these traffic statistics, call holding and channel occupancy times, are key elements to compute performance metrics such as call blocking and dropping probabilities. In cellular network analysis call holding times are generally assumed to be exponentially distributed due to studies on wireline traffic statistics. In [3], Hong and Rappaport proposed a traffic model for cellular mobile radio telephone systems and approximated channel occupancy time distribution by exponential distribution when call holding times are assumed to be exponentially distributed. Ramjee *et al.* [6], Fang and Zhang [7], Naghshineh and Schwartz [15], Gersht and Lee [22], Borst and Mitra [34] and Yavuz and Leung [21] studied the performance of various call admission control schemes using one dimensional Markov chain models assuming that channel occupancy times are exponentially distributed based on Hong and Rappaport's study due to its tractability. In [25], Rappaport developed multidimensional models under the same assumption and with Monte the author obtained call blocking probabilities using this model [26]. However simulation studies and field data have shown that these assumptions are not perpetually valid. In [35], Guerin used a simulation model to show that channel occupancy time distribution displays a rather poor agreement with the exponential fitting for mobile users with low change rate of movement direction. Jedrzycki and Leung showed in [36] that exponential distribution assumption for channel occupancy times is not correct and a lognormal model approximation fits much better using real cellular data. Fang *et al.* demonstrated in [37] and [38] that channel occupancy times in a cellular network depend not only on call holding times but also on users' mobility which can be characterized by cell residence time distribution. In [20], the authors showed that channel occupancy time is exponentially distributed only if cell residence time is exponentially distributed. Yet, it is also observed in the same study that channel occupancy time distribution have a good approximation by exponential distribution in general when the mobility is low.

Barcelo and Jordan analyzed a cellular network based on a fully empirical approach in [39] and observed that channel occupancy is less spread out than if exponential distribution was assumed.

Markov chain models are developed to evaluate performance analytically in cellular networks. Calls arriving to a particular cell are grouped into QoS classes or call types, such as new and handoff, based on their first appearance in the corresponding cell. Channel occupancy times for each group are measured from call starting time till the occupied channel in the respective cell is discarded due to call termination or handoff. However, call holding times are measured from call starting time till call termination regardless of occupying a channel in the same cell or not. The characteristics of various types of channel occupancy times are needed to be analyzed to provide sufficiently representative channel occupancy time statistics not only when developing analytical models but also for feeding simulations with realistic traffic statistics to obtain network performance metrics.

The rest of this chapter is organized as follows: Section 1.1 discusses the motivations and objectives of our work. Section 1.2 presents an overview of our contributions. Section 1.3 describes the organization of this dissertation.

1.1 Motivations and Objectives

Many guard channel based call admission control schemes have been proposed to provide the desired quality of service to new and handoff calls in cellular networks. One dimensional Markov chain modeling is generally used under specific assumptions to compute blocking and dropping probabilities of these calls approximately to avoid solving large sets of flow equations that makes exact analysis of these schemes using multidimensional Markov chain models infeasible. The "traditional" approximation method provides accurate results only when channel occupancy times for new and handoff calls have equal mean values while the "normalized" approach relaxes this assumption only for the new call bounding call admission control scheme. Yet, these assumptions may not be appropriate since these two types of calls may have different average channel occupancy times if not different distributions [19] [20]. This motivates us to develop an accurate yet easy to implement method to compute new and handoff call blocking and dropping probabilities for several widely known call admission control schemes when channel occupancy times for new and handoff calls have separate mean values.

A wide variety of network applications that range from text-based to multimedia utilities are provided by cellular networks along with basic voice services. These utilities are grouped under various quality of service classes and a higher priority is normally assigned to calls with higher bandwidth and lower latency requirements based on the application's importance for cellular network operator. Call admission control schemes are also used to optimize call blocking and dropping probabilities of these applications for quality of service provisioning in cellular networks. Performances of these call admission control schemes are evaluated using either one dimensional or multidimensional Markov chain models with the former preferred over the latter to avoid solving large sets of flow equations. A computationally efficient solution is also important for quality of service provisioning when dynamic call admission control schemes are used since efficient adaptive reservation depends on reliable and up to date system status feedback simultaneously provided to the call admission control mechanism. However, relaxing the assumptions to have separate mean values for channel occupancy times of different classes of calls is not sufficient to evaluate multi-service cellular networks using approximation methods based on one dimensional Markov chain modeling. When assumptions are relaxed further to have calls with separate capacity requirements, previously developed one dimensional Markov chain models become obsolete due to the multidimensionality introduced by calls with unequal bandwidth requests. Borst and Mitra proposed an approximation method [34] with a closed form and therefore a fast solution, but it only approximates sufficiently accurately when the number of existing call arrival types are high. The need for an accurate and computationally efficient performance evaluation approximation method for call admission control schemes motivates us to develop an easy to implement method to compute call blocking and dropping probabilities of different classes of calls in multi-service cellular networks.

Packet-switched services are overlaid on circuit-switched technology over the same air interface to use the access capacity in cellular networks. Statistical multiplexing is used to transmit data packets over radio interface to provide a quality of service level comparable to that of circuit-switched services. Accurate voice traffic statistics are needed to understand the distribution of idle periods of voice channels to multiplex more than one service on to the same radio channel. In classical voice traffic modeling call holding times are approximated by exponential distribution and this assumption is widely used due to its tractability to obtain analytical results for evaluating cellular networks [6] [7] [15] [21] [22] [34]. However it has been shown that a lognormal distribution approximation fits much closer [36] [39]. In a cellular network when a call admission control scheme is modeled for each cell, arriving calls to a particular cell are grouped into quality of service classes or call types, such as new and handoff, based on their first appearance in the corresponding cell. Channel occupancy time distribution for each group includes respective channel occupancy times counted only until the corresponding calls discard the occupied channels in the cell due to call termination or handoff. Call holding time distribution, on the other hand, includes the amount of times that the channels are occupied by a call until it terminates either in its originating cell or another. The above discussion motives us to provide sufficiently representative channel occupancy time statistics to develop analytical models since call holding time statistics are not sufficient alone. This is also very useful for feeding simulations with realistic traffic statistics to obtain network performance metrics.

1.2 Main Contributions

The main contributions of this dissertation are as follows:

- **Develop a computationally efficient approximation method to evaluate performance of call admission control schemes in single service cellular networks**: We propose an easy to implement approximation method to evaluate call admission control schemes when average channel occupancy times for new and handoff calls are not necessarily equal. Our proposed approximation method yields more accurate results compared with the previously proposed "traditional" and "normalized" methods while keeping the computational complexity low.
- **Develop computationally efficient approximation methods to evaluate performance of call admission control schemes in multi-service cellular networks**: We classify call admission control schemes into two categories called symmetric and asymmetric. We present the product form solution formula to evaluate symmetric call admission control schemes and propose a novel performance evaluation approximation method to evaluate asymmetric call admission control

schemes when average channel occupancy times for different classes of calls are not necessarily equal and all arriving calls may have distinct capacity requirements. The proposed method performs better in accuracy compared to the previously proposed method by Borst and Mitra while keeping the computational complexity low.

- **Statistical modeling of channel occupancy times for voice service in cellular networks**: We present an empirical approach to determine the probability distribution functions that fit various types of channel occupancy times in cellular networks. We show that these channel occupancy times can be approximated by lognormal distribution.
- **Statistical modeling of call holding times for stationary and mobile users along with the number of handoffs committed by a mobile user in cellular networks**: We show that the closest fit candidate to approximate stationary and mobile users' call holding times is lognormal distribution along with the distribution of the number of handoffs committed by a mobile user.

1.3 Organization of the Dissertation

This dissertation is organized as follows. In chapter 2, we propose a computationally efficient approximation method to evaluate performance of call admission control schemes in single service cellular networks. We reevaluate the analytical methods for computing new/handoff call blocking/dropping probabilities for widely known call admission control schemes and show that the proposed approach gives more accurate results under relaxed assumptions when compared with the existing methods. In chapter 3, we propose computationally efficient approximation methods to evaluate performance of call admission control schemes in multi-service cellular networks assuming that average values for channel occupancy times of different classes of calls are not equal and all arriving calls have different capacity requirements. We present the numerical results that show the proposed methods provide results that match well with the exact solutions while keeping the computational complexity low. In chapter 4, we determine the probability distribution functions that fit various types of channel occupancy times in cellular networks and show that these channel occupancy times can be approximated by lognormal distribution. We also show that the closest fit candidate to approximate stationary and mobile users' call holding times is lognormal distribution along with the distribution of the number of handoffs committed by a mobile user. Chapter 5 concludes the thesis with a summary of the presented work, and describes the future works.

1.4 Bibliography

[1] D. E. Everitt, “Traffic engineering of the radio interface for cellular mobile networks,” *Proc. IEEE*, vol. 82, no. 9, pp.1371-1382, 1994.

[2] H. Chen, L. Huang, S. Kumar, and C. C. J. Kuo, *Radio Resource Management for Multimedia QoS Support in Wireless Networks*. Boston: Kluwer Academic Publishers, 2004, chapter 2.

[3] D. Hong and S. S. Rappaport, “Traffic model and performance analysis for cellular mobile radiotelephone systems with prioritized and non-prioritized handoff procedures,” *IEEE Transactions on Vehicular Technology*, vol. 35, pp. 77-92, Aug. 1986.

[4] B. Li, C. Lin, and S. T. Chanson, “Analysis of a hybrid cutoff priority scheme for multiple classes of traffic in multimedia wireless networks,” *Wireless Networks*, vol. 4, no. 4, pp. 279-290, July 1998.

[5] Y.B. Lin, S. Mohan, and A. Noerpel, “Queuing priority channel assignment strategies for handoff and initial access for a PCS network,” *IEEE Transactions on Vehicular Technology*, vol. 43, no. 3, pp. 704-712, Aug. 1994.

[6] R. Ramjee, R. Nagarajan, and D. Towsley, “On optimal call admission control in cellular networks,” *Wireless Networks*, vol. 3, no. 1, pp. 29-41, March 1997.

[7] Y. Fang, and Y. Zhang, “Call admission control schemes and performance analysis in wireless mobile networks,” *IEEE Transactions on Vehicular Technology*, vol. 51, no.2, pp. 371-382, March 2002.

[8] M. D. Kulavaratharasah, and A. H. Aghvami, “Teletraffic performance evaluation of microcell personal communication networks (PCNs) with prioritized handoff procedures,” *IEEE Transactions on Vehicular Technology*, vol. 48, no. 1, pp. 137-152, Jan. 1999.

[9] R. A. Guerin, “Queuing-blocking system with two arrival streams and guard channels,” *IEEE Transactions on Communications*, vol. 36, no. 2, pp. 153-163, Feb. 1988.

[10] E. D. Re, R. Fantacci, and G. Giambene, “Handover queuing strategies with dynamic and fixed channel allocation techniques in low earth orbit mobile satellite systems,” *IEEE Transactions on Communications*, vol. 47, no. 1, pp. 89-102, Jan. 1999.

[11] C. H. Yoon, and C. K. Un, “Performance of personal portable radio telephone systems with or without guard channels,” *IEEE Journal on Selected Areas in Communications*, vol. 11, no. 6, pp. 911-917, Aug. 1993.

[12] C. Chang, C. J. Chang, and K. R. Lo, "Analysis of a hierarchical cellular system with reneging and dropping for waiting new calls and handoff calls," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 4, pp. 1080-1091, July 1999.

[13] V. K. N. Lau, and S. V. Maric, "Mobility of queued call requests of a new call queuing technique for cellular systems," *IEEE Transactions on Vehicular Technology*, vol. 47, no. 2, pp. 480-488, May 1998.

[14] A. S. Acampora and M. Naghshineh, "Control and quality of service provisioning in high-speed micro-cellular networks," *IEEE Personal Communications*, vol. 1, no. 2, pp. 36-43, 1996.

[15] M. Naghshineh and S. Schwartz, "Distributed call admission control in mobile/wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 14, no. 4, pp. 711-717, May 1996.

[16] D. Levine, I. Akyildiz, and M. Naghshineh, "A resource estimation and call admission algorithm for wireless multimedia networks using the shadow cluster concept," *IEEE/ACM Transactions on Networking*, vol. 5, no. 1, pp. 1-12, Feb. 1997.

[17] C. Oliveira, J. B. Kim, and T. Suda, "An adaptive bandwidth reservation scheme for high-speed multimedia wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 858-874, Aug. 1998.

[18] P. Ramanathan, K. M. Sivalingam, P. Agrawal, and S. Kishore, "Dynamic resource allocation schemes during handoff for mobile multimedia wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 7, pp. 1270-1283, July 1999.

[19] Y. Fang and I. Chlamtac, "Teletraffic analysis and mobility modeling for PCS networks," *IEEE Transactions on Communications*, vol. 47, pp. 1062-1072, July 1999.

[20] Y. Fang, I. Chlamtac, and Y. B. Lin, "Channel occupancy times and handoff rate for mobile computing and PCS networks," *IEEE Transactions on Computers*, vol. 47, pp. 679-692, June 1998.

[21] E. A. Yavuz and V. C. M. Leung, "Computationally efficient method to evaluate the performance of guard-channel-based call admission control in cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 4, pp. 1412-1424, July 2006.

[22] A. Gersht and K. J. Lee, "A bandwidth management strategy in ATM networks," Technical report, GTE Laboratories, 1990.

[23] J. W. Roberts, "Teletraffic models for the Telecom 1 integrated services network," *Proceedings of the 10th International Teletraffic Conference*, Montreal, 1983.

[24] W. Li and X. Chao, "Modeling and performance evaluation of a cellular mobile network," *IEEE/ACM Transactions on Networking*, vol. 12, no. 1, Feb. 2004.

[25] S. S. Rappaport, "The multiple call handoff problem in personal communications networks," *IEEE 40th Vehicular Technology Conference*, pp. 287–294, May 1990.

[26] S. S. Rappaport and G. Monte, "Blocking, hand-off and traffic performance for cellular communication systems with mixed platforms," *IEEE 42nd Vehicular Technology Conference*, vol. 2, pp. 1018–1021, May 1992.

[27] L. Huang, S. Kumar and C. C. J. Kuo, "Adaptive resource allocation for multimedia services in wireless communication networks," *Distributed Computing Systems Workshop, 2001 International Conference on*, 2001, pp. 307 - 312.

[28] L. Qiu, P. Xia and J. Zhu, "Study on wideband CDMA modulation, power control and wireless access for CDMA multimedia systems," *Proceedings of IEEE 50th Vehicular Technology Conference*, vol. 5, pp. 2944 - 2948, 1999.

[29] S. Chen and K. Nahrstedt, "Distributed quality of service routing in ad-hoc networks," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 8, August, 1999.

[30] ITU-T SG16/Q.15 video coding experts group, ITU-T recommendation H.263: Video coding for low bit rate communication, October 1995.

[31] R. Guerin, "Equivalent capacity and its application to bandwidth allocation in high-speed networks," *IEEE Journal on Selected Areas in Communications*, vol. 9, no. 7, pp. 968-981, Sept. 1991.

[32] J. S. Evans and D. Everitt, "Effective bandwidth-based admission control for multi-service CDMA cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 1, Jan. 1999.

[33] Q. Ren and G. Ramamurthy, "A real-time dynamic connection admission controller based on traffic modeling, measurement, and fuzzy logic control," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 2, Feb. 2000.

[34] S. C. Borst and D. Mitra, "Virtual partitioning for robust resource sharing: computational techniques for heterogeneous traffic," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 5, pp. 668-678, June 1998.

[35] R. A. Guerin, "Channel occupancy time distribution in a cellular radio system," *IEEE Transactions Vehicular Technology*, vol. 35, no. 3, pp. 89-99, 1987.

[36] C. Jedrzycki, and V. C. M. Leung, "Probability distribution of channel holding time in cellular telephony systems," *IEEE Vehicular Technology Conference (VTC'96)*, vol. 1, pp. 247 - 251, Apr. 1996.

[37] Y. Fang, I. Chlamtac, and Y.B. Lin, "Call performance for a PCS network," *IEEE Journal on Selected Areas in Communications* vol. 15, no. 8, pp. 1568–1581, Oct. 1997.

[38] Y. Fang, I. Chlamtac, and Y.B. Lin, "Modeling PCS networks under general call holding times and cell residence time distributions," *IEEE Transactions on Networking* vol. 5, no. 6, pp. 893 – 906, Dec. 1997.

[39] F. Barcelo, and J. Jordan, "Channel holding time distribution in public telephony systems (PAMR and PCS)," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 5, pp. 1615-1625, Sep. 2000.

CHAPTER 2 Computationally Efficient Method To Evaluate The Performance Of Guard-Channel-Based Call Admission Control In Cellular Networks[1]

2.1 Introduction

The emerging global standards for wireless communication networks, such as the third generation (3G) cellular networks, promise the efficiency and flexibility of multiplexing a wide variety of traffic from conventional circuit-switched voice service to packet-switched voice, data, and multimedia services while providing the quality of service (QoS) expected by service subscribers and their applications. Services with specific bandwidth and latency requirements need specific amount of system resources to be allocated at the time of call admission to ensure that the required QoS can be maintained during the call. The performance of call admission control (CAC) in a cellular network is specified by the blocking probability of new calls in a cell and dropping probabilities of handoff calls entering a cell. Call dropping may occur before a call is terminated by the communicating parties if the cellular network is unable to assign network resources to enable the call to be handed off to a new cell that the mobile user has moved into. Network planners need to provision sufficient network resources such as channel bandwidth to ensure that call blocking/dropping probabilities stay within acceptable limits for various call types or applications. Since dropping an on-going call is generally more objectionable to a mobile user than blocking a new call request, a higher priority is normally assigned in CAC for handoff calls over the new ones in order to minimize the call dropping probability.

Call admission control for wireless networks has been intensively studied in the past [1] [2] and many priority-based CAC schemes have been proposed [3]-[18]. In the context of a given number of channels being available in each cell for assignment to admitted calls,

[1] **A version of this chapter has been published. E. A. Yavuz and V. C. M. Leung, "Computationally Efficient Method to Evaluate the Performance of Guard-Channel-Based Call Admission Control in Cellular Networks,"** ***IEEE Transactions on Vehicular Technology*****, vol. 55, no. 4, pp.1412-1424, July 2006.**

these CAC schemes can be classified into two broad categories.

1) *Guard Channel (GC) Schemes:* A number of guard channels are reserved for handoff calls. There are four different schemes.
    a) The *cutoff priority scheme* blocks a new call if the number of free channels is less than the number of guard channels reserved for handoff calls [3]-[5].
    b) The *fractional guard channel scheme* admits a new call with certain probability that depends on the number of busy channels in the cell [6].
    c) The *new call bounding scheme* limits the number of new calls admitted to the cell to some number less than the total number of available channels [7].
    d) The *rigid division-based scheme* divides all channels available in a cell into two groups: one for common use and the other only for handoff calls [8].
2) *Queuing Priority (QP) Schemes*: Calls are accepted whenever there are free channels; otherwise either new calls are queued while handoff calls are dropped [9], new calls are blocked while handoff calls are queued [10][11], or all arriving calls are queued with certain rearrangements in the queue [12][13].

In addition to those given above, many dynamic GC schemes [14]-[18] have also been discussed in the literature to improve system efficiency. These dynamic schemes manage to accept more lower-priority calls as compared to the fixed schemes by adaptively reserving the amount of resources needed for high-priority calls.

In the literature CAC schemes are analyzed using Markov chain models based on a circuit-switched network architecture with the assumption that independent Poisson distributed call requests are originated from different classes of service or call types, cell residence time for each call is exponentially distributed, and each call requires a predetermined channel bandwidth. Conventional circuit-switched services including telephony are gradually being replaced by packet-switched ones as today's communication networks evolve. Conventional CAC schemes, on the other hand, will continue to be useful when applied with suitable scheduling techniques to guarantee QoS at the packet level since most applications such as interactive multimedia are inherently connection oriented and packet-switched connections can be provisioned according to their effective bandwidths [19]-[33].

To avoid the computational complexity of solving multidimensional Markov chain models, one dimensional Markov chain models are commonly used (e.g., [11][15]) to obtain the blocking/dropping probabilities for new/handoff calls under the assumption that the channel occupancy times for both types of calls are identically distributed with the same average values. Yet this assumption may not be appropriate as new and handoff calls may have different average channel occupancy times if not different distributions, as shown in [19][20] and references therein. Even though analysis in [7] accounts for the more general case of different average channel occupancy times for new and handoff calls, the approximation employed still leads to significant discrepancies with the exact solutions. In [22], Li and Chao modeled a multi-cell wireless network employing a hybrid *GC/QP* scheme with transfer of unsuccessful requests to neighboring cells as a network of queues and obtained a product form solution. However, the solution is restrictive to the protocol considered and may not be applicable for performance evaluations of *GC* schemes in general.

Thus, using multidimensional Markov chain models appears to be the only means to obtain accurate solutions for the analysis of *GC* schemes. This approach is used in [22] to evaluate a cellular network's performance by determining the new call blocking and handoff call dropping probabilities, and is extended in [26] to analyze the traffic performance taking into consideration the effects of various types of mobile platforms distinguished by different mobility characteristics. Even though the multidimensional Markov chain model is capable of providing the exact results, the method suffers from the curse of dimensionality, which results in very high computational cost for large systems. Therefore it is still desirable to come up with approximate solutions that have high accuracy and is computationally efficient.

In this chapter, we propose a novel method called the "*effective holding time*" approach to compute the above-mentioned CAC performance metrics using an approximation based on one dimensional Markov chain modeling under the condition that the channel occupancy times for new and handoff calls are independent and exponentially distributed with different average values. The proposed method is easy to implement and has low computational cost.

This chapter is organized as follows. In the next section we examine three of the widely known CAC schemes by evaluating their performances using the *traditional* and the *normalized* analytical methods proposed in the literature under the assumption that the new and handoff calls have different average channel occupancy times. In Section 2.3, we present the new analytical method based on *effective holding time*, which yields more accurate approximations than the traditional or normalized methods. In addition to the numerical results obtained using the traditional, normalized, proposed and the direct methods, the accuracy of the results are also presented and compared along with the runtime computational costs in Section 2.4. We will conclude the paper in Section 2.5.

2.2 Existing Methods to Analyze Performance of CAC Schemes

In this section we examine three of the widely known CAC schemes: the *new call bounding priority*, the *cutoff priority*, and the *fractional guard channel* schemes by evaluating their performance using the *traditional* and the *normalized* analytical methods proposed in the literature under the assumption that the new and handoff calls have different average channel occupancy times.

Let λ_n and λ_h denote the arrival rates and $1/\mu_n$ and $1/\mu_h$ denote the average channel occupancy times for new and handoff calls, respectively. Let C denote the total number of channels in a cell. We assume that the arrival processes for new and handoff calls are Poisson, and the channel occupancy times for new calls and handoff calls are exponentially distributed.

Assuming that both new and handoff calls have the same channel occupancy time distributions and average values, the system model is approximated by a one dimensional Markov chain with a fixed average channel occupancy time for the total cell traffic. We refer to this method of deriving blocking/dropping probabilities analytically as the *traditional* approach [3][6][7][10][11]. To improve the inaccurate results obtained when the above assumption on equal average channel occupancy time is no longer valid, Fang and Zhang [7] proposed normalizing the average service times for new and handoff traffic to unity so that they become identical for both streams. Although this approximation, which we call the *normalized* approach, seems to give better accuracy than the traditional approach, it is still inaccurate especially for CAC schemes like cutoff priority and fractional guard channel.

2.2.1 New Call Bounding Scheme

Fig. 2.1 shows the state transition diagram for the new call bounding scheme modeled by a two-dimensional Markov chain, where λ_n, λ_h, μ_n, μ_h and C are as defined before and K is the threshold between 0 and C such that a new call request is admitted only when there are less than K channels occupied by new calls. Let $p(n_1,n_2)$ denote the steady state probability that there are n_1 new calls and n_2 handoff calls in the cell, which can be found by solving the global balance equations obtained from the state transition. Yet as mentioned above, solving these balance equations becomes computationally very intensive when the state dimension increases. Analytical results to compute the blocking probabilities of new calls p_{nb} and the dropping probabilities of handoff calls p_{hd} have been derived for this scheme in [7], and an approximation is developed by normalizing the average service time for new and handoff calls. This *normalized* approach allows the arriving traffic for each type of call to be scaled appropriately and the blocking/dropping probabilities to be related to the traffic intensities.

Here is how the *normalized* approach works. Let $\rho_n = \lambda_n / \mu_n$ and $\rho_h = \lambda_h / \mu_h$, then we can consider an equivalent Poisson new call arrival stream with arrival rate ρ_n and service rate equal to 1, and an equivalent Poisson handoff call arrival stream with arrival rate ρ_h and service rate equal to 1. Let $p^a(n_1,n_2)$ denote the steady state probability that there are n_1 new calls and n_2 handoff calls in the cell for the normalized approximation model. Thus, the following stationary distribution is obtained for this approximate model from the balance equations:

$$p^a(n_1,n_2) = \frac{\rho_n{}^{n_1}}{n_1!} \cdot \frac{\rho_h{}^{n_2}}{n_2!} \cdot p^a(0,0), \quad 0 \le n_1 \le K,\ n_1 + n_2 \le C,\ n_2 \ge 0,$$

where

$$p^a(0,0) = \left[\sum_{0 \le n_1 \le K, n_1+n_2 \le C} \frac{\rho_n{}^{n_1}}{n_1!} \cdot \frac{\rho_h{}^{n_2}}{n_2!} \right]^{-1}$$

$$= \left[\sum_{n_1=0}^{K} \frac{\rho_n^{n_1}}{n_1!} \cdot \sum_{n_2=0}^{C-n_1} \frac{\rho_h^{n_2}}{n_2!} \right]^{-1}$$

The formulas for new call blocking and handoff call dropping probabilities are as follows:

$$p_{nb}^{a}=\frac{\sum_{n_2=0}^{C-K}\frac{\rho_n^K}{K!}\cdot\frac{\rho_h^{n_2}}{n_2!}+\sum_{n_1=0}^{K-1}\frac{\rho_n^{n_1}}{n_1!}\cdot\frac{\rho_h^{C-n_1}}{(C-n_1)!}}{\sum_{n_1=0}^{K}\frac{\rho_n^{n_1}}{n_1!}\cdot\sum_{n_2=0}^{C-n_1}\frac{\rho_h^{n_2}}{n_2!}} \tag{1}$$

$$p_{hd}^{a}=\frac{\sum_{n_1=0}^{K}\frac{\rho_n^{n_1}}{n_1!}\cdot\frac{\rho_h^{C-n_1}}{(C-n_1)!}}{\sum_{n_1=0}^{K}\frac{\rho_n^{n_1}}{n_1!}\cdot\sum_{n_2=0}^{C-n_1}\frac{\rho_h^{n_2}}{n_2!}} \tag{2}$$

On the other hand, the *traditional* approach uses the average channel occupancy time μ_{av} for total cell traffic given by:

$$\frac{1}{\mu_{av}}=\frac{\lambda_n}{\lambda_n+\lambda_h}\cdot\frac{1}{\mu_n}+\frac{\lambda_h}{\lambda_n+\lambda_h}\cdot\frac{1}{\mu_h}=\frac{\rho_n+\rho_h}{\lambda_n+\lambda_h} \tag{3}$$

to replace μ_n and μ_h in (1) and (2) and to obtain new call blocking and handoff call dropping probabilities in the resulting one dimensional Markov chain model. In that case, the traffic intensities for new and handoff calls are given by:

$$\hat{\rho}_n=\frac{\lambda_n}{\mu_{av}}=\frac{\lambda_n}{\lambda_n+\lambda_h}\cdot(\rho_n+\rho_h)$$

$$\hat{\rho}_h=\frac{\lambda_h}{\mu_{av}}=\frac{\lambda_h}{\lambda_n+\lambda_h}\cdot(\rho_n+\rho_h)$$

When channel occupancy times for both new and handoff calls are assumed to be identically distributed with the same parameters, the cell average channel occupancy time also becomes the same as can be easily deduced from (3). However, the *traditional* approach can yield significantly inaccurate results when the channel occupancy times for new and handoff calls have different average values except when the non-prioritized scheme (K = C) is used.

The *normalized* approach overcomes this inaccuracy in the new call bounding scheme by exploiting the symmetric nature and the product form of the detailed balance equations that characterizes this CAC scheme. We will show in the following sections that both approximations mentioned above are not good enough to obtain new call blocking and handoff call dropping probabilities with an acceptable accuracy for other CAC schemes.

2.2.2 Cutoff Priority Scheme

Let $m < C$ denote the channel occupancy threshold for acceptance of new calls. If the total number of busy channels is less than m when a new call arrives, the call is accepted; otherwise the new call is blocked. A handoff call is always accepted if a free channel is available. This scheme has been extensively studied using one dimensional Markov chain modeling under the assumption that the average channel occupancy times of new and handoff calls are equal [3]. However, this approach provides inaccurate results when the equal mean channel occupancy time assumption does not hold true.

Let λ_n, λ_h, μ_n, μ_h, and C be defined as before. The system can be modeled by the two dimensional Markov chain shown in Fig. 2.2, where (n_1, n_2) denotes a feasible state with n_1 and n_2 representing the numbers of new and handoff calls in the cell, respectively. In contrast with the new call bounding scheme, the state flows for the cutoff priority scheme no longer have the symmetric nature since the flows for some of the states are unidirectional, spoiling the symmetry as shown within circles drawn in Fig. 2.1 and Fig. 2.2. Hence, no product form solution exists for this scheme when $\mu_n \neq \mu_h$.

The following stationary distribution is obtained for the cutoff priority scheme using the *normalized* approach, where p_j^a denotes the probability that there are j ($j = 0, 1, \ldots, C$) busy channels in the steady state:

$$p_j^a = \begin{cases} \dfrac{(\rho_n + \rho_h)^j}{j!} \cdot p_0^a, & j \le m \\ \dfrac{(\rho_n + \rho_h)^m \cdot \rho_h^{j-m}}{j!} \cdot p_0^a, & m+1 \le j \le C \end{cases}$$

where

$$p_0^a = \left[\sum_{j=0}^{m} \frac{(\rho_n + \rho_h)^j}{j!} + \sum_{j=m+1}^{C} \frac{(\rho_n + \rho_h)^m \cdot \rho_h^{j-m}}{j!} \right]^{-1}$$

Using the stationary distribution given above, the blocking probabilities for new calls, p_{nb}^{a} and the dropping probabilities for handoff calls, p_{hd}^{a} are obtained as follows:

$$p_{nb}^{a} = \frac{\sum_{j=m}^{C} \frac{(\rho_n + \rho_h)^m \cdot \rho_h^{j-m}}{j!}}{\sum_{j=0}^{m} \frac{(\rho_n + \rho_h)^j}{j!} + \sum_{j=m+1}^{C} \frac{(\rho_n + \rho_h)^m \cdot \rho_h^{j-m}}{j!}} \tag{4}$$

$$p_{hd}^{a} = \frac{\frac{(\rho_n + \rho_h)^m \cdot \rho_h^{C-m}}{C!}}{\sum_{j=0}^{m} \frac{(\rho_n + \rho_h)^j}{j!} + \sum_{j=m+1}^{C} \frac{(\rho_n + \rho_h)^m \cdot \rho_h^{j-m}}{j!}} \tag{5}$$

On the other hand, the corresponding results for the *traditional* approach in which the new and handoff call channel occupancy times are replaced with the average channel occupancy time, given by (3), are as follows:

$$p_{nb}^{t} = \frac{\sum_{j=m}^{C} \frac{1}{j!} \cdot (\frac{\lambda_n + \lambda_h}{\mu_{av}})^m \cdot (\frac{\lambda_h}{\mu_{av}})^{j-m}}{\sum_{j=0}^{m} \frac{1}{j!} \cdot (\frac{\lambda_n + \lambda_h}{\mu_{av}})^j + \sum_{j=m+1}^{C} \frac{1}{j!} \cdot (\frac{\lambda_n + \lambda_h}{\mu_{av}})^m \cdot (\frac{\lambda_h}{\mu_{av}})^{j-m}}$$

$$p_{hd}^{t} = \frac{(\frac{\lambda_n + \lambda_h}{\mu_{av}})^m \cdot (\frac{\lambda_h}{\mu_{av}})^{C-m} \cdot \frac{1}{C!}}{\sum_{j=0}^{m} \frac{1}{j!} \cdot (\frac{\lambda_n + \lambda_h}{\mu_{av}})^j + \sum_{j=m+1}^{C} \frac{1}{j!} \cdot (\frac{\lambda_n + \lambda_h}{\mu_{av}})^m \cdot (\frac{\lambda_h}{\mu_{av}})^{j-m}}$$

2.2.3 Fractional Guard Channel Scheme

The fractional guard channel scheme admits new calls with certain probabilities that depend on the number of busy channels, i. When the number of busy channels is i, an arriving new call will be admitted with probability β_i, where $0 \leq \beta_i \leq 1$, $i = 0, 1, \ldots, C-1$. The new call stream is smoothly throttled by decreasing β_i as the network traffic is building up. An arriving handoff call will always be admitted unless there is no free channel available, in which case all calls will be blocked. Obviously, this scheme becomes the cutoff priority scheme when $\beta_0 = \ldots = \beta_{m-1} = 1$ and $\beta_m = \ldots = \beta_C = 0$. When $\beta_0 \geq \beta_1 \geq \beta_2 \geq \ldots \geq \beta_C$, it can also be observed that the new calls stream grows at a decreasing rate as the number of busy channels increases. Due to this flexible choice of new call admission probabilities, the fractional guard channel scheme can be made very general.

Let λ_n, λ_h, μ_n, μ_h, and C be defined as before and let $q(c)$ denote the equilibrium channel occupancy probability when exactly c channels are occupied in a cell. It has been shown [27] that $q(c)$ satisfies the following recursive equation.

$$(\frac{\lambda_n \cdot \beta_{c-1}}{\mu_n} + \frac{\lambda_h}{\mu_h}) \cdot q(c-1) = c \cdot q(c), \quad c = 1,\ldots,C \tag{6}$$

Replacing the new and handoff call arrival rates, λ_n and λ_h, with the new and handoff traffic intensities, ρ_n and ρ_h, respectively, and setting the corresponding channel occupancy times to 1 as suggested by the *normalized* approach transform (6) to

$$(\rho_n \cdot \beta_{c-1} + \rho_h) \cdot q(c-1) = c \cdot q(c), \quad c = 1,\ldots,C \tag{7}$$

which gives us a general expression for all *GC* schemes.

Solving for $q(0)$ in the equation $\sum_{j=0}^{C} q(j) = 1$, we obtain

$$q(j) = \frac{\prod_{k=0}^{j-1} (\rho_n \cdot \beta_k + \rho_h)}{j!} \cdot q(0), \quad 1 \leq j \leq C \tag{8}$$

where

$$q(0) = \left[1 + \sum_{j=1}^{C} \frac{\prod_{k=0}^{j-1} (\rho_n \cdot \beta_k + \rho_h)}{j!} \right]^{-1} \tag{9}$$

From (8) and (9), the blocking and dropping probabilities for new and handoff calls for the normalized approach are respectively given by:

$$p^{a}_{nb} = \sum_{j=0}^{C} (1 - \beta_j) \cdot q(j), \quad \beta_C = 0 \qquad (10)$$

$$p^{a}_{hd} = q(C) \qquad (11)$$

On the other hand, the corresponding results for the *traditional* approach can be obtained by replacing the average channel occupancy times of both types of calls in (6) with the average channel occupancy time of the total cell traffic given by (3).

2.3 The Proposed Performance Evaluation Method

Even though the *normalized* approach [7] provides a better approximation than the *traditional* one, it is still inaccurate for many *GC* schemes. Therefore, in order to provide more accurate results while keeping the computational complexity low, we present the following novel performance evaluation method for *GC* schemes, referred as the *effective holding time* approach.

Since it is crucial to find an approximation that overcomes the curse of dimensionality when the state dimension is large, it is inevitable to attempt reducing the two dimensional Markov chain model to a one dimensional one. As shown previously, this enables a product form solution of the detailed balance equations to be obtained using (6).

However, we proceed to simplify (6) by replacing the average channel occupancy times for both new and handoff calls with an average channel occupancy time for the total cell traffic which we refer as the *average effective channel occupancy time* and denote by $1/\mu_{eff}$ instead of $1/\mu_{av}$ used by the *traditional* approach. The average channel occupancy time, $1/\mu_{av}$, given by (3) can not approximate the value of the average channel occupancy time for the total cell traffic accurately, since new and handoff calls are not blocked equally. To obtain $1/\mu_{eff}$, we apply an idea proposed by Gersht and Lee [28] when developing an iterative algorithm to calculate approximate occupancy distributions of objects being placed into a knapsack to maximize the total reward that is accrued each time an object is placed into the knapsack. Inspired by the well known Little's theorem, $\mu = \lambda/N$, they defined μ_{eff} as the ratio of expected number of both types of call arrivals to the expected number of occupied channels,

$$\mu_{eff} = \frac{\sum_{j=0}^{C-1}(\lambda_n \cdot \beta_j \cdot q(j)) + \sum_{j=0}^{C-1}(\lambda_h \cdot q(j))}{\sum_{j=0}^{C} j \cdot q(j)} \tag{12}$$

We now simply approximate the occupancy probabilities by setting $q(c) = \hat{q}(c)$, $c = 0,\ldots,C$ in (6) and using the updated recursive formula given below.

$$(\lambda_n \cdot \beta_{c-1} + \lambda_h) \cdot \hat{q}(c-1) = c \cdot \mu_{eff} \cdot \hat{q}(c), \quad c = 1,\ldots,C \tag{13}$$

Solving for $\hat{q}(0)$ with $\sum_{j=0}^{C} \hat{q}(j) = 1$, we obtain

$$\hat{q}(j) = \frac{\prod_{k=0}^{j-1}(\lambda_n \cdot \beta_k + \lambda_h)}{\mu_{eff}^{j} \cdot j!} \cdot \hat{q}(0), \quad 1 \le j \le C \tag{14}$$

where

$$\hat{q}(0) = \left[1 + \sum_{j=1}^{C} \frac{\prod_{k=0}^{j-1}(\lambda_n \cdot \beta_k + \lambda_h)}{\mu_{eff}^{j} \cdot j!} \right]^{-1} \tag{15}$$

In their knapsack problem approach, Gersht and Lee [28] suggested obtaining μ_{eff} using (12) by replacing $q(c)$ with $\hat{q}(c)$ and updating the approximate equilibrium occupancy probabilities iteratively, using (14) and (15) until each $\hat{q}(c)$ changes by no more than ε for all $c = 0,\ldots,C$, where ε is a small number.

Although their approach did not emphasize starting with an appropriate initial value for each approximate equilibrium occupancy probability, we realize that it becomes a problem since more than one set of equilibrium occupancy probabilities can satisfy (14) and (15) at the same time. What makes a set to be the unique solution depends on the values of the arrival rates and average channel occupancy times of both types of calls. Therefore, we consider it important that these values are included directly in the computation to obtain better approximate results. The call arrival rates for both types of calls are embraced in (12), (14) and (15); however we observe that the average channel occupancy time of each type of call is not considered directly in these equations when computing the approximate equilibrium occupancy probabilities since they are replaced by the average effective channel occupancy time, $1/\mu_{eff}$.

Hence, we propose to set the approximate equilibrium occupancy probabilities initially with the values obtained by the *normalized* approach proposed by Fang and Zhang [7] in order to make the approximate equilibrium occupancy probabilities closer to the unique solution that we look for, then apply (14) and (15) to obtain the new and handoff call dropping probabilities.

To summarize, the algorithmic description of our proposed *effective holding time* approach is as follows:

```
1. Initialize equilibrium occupancy probabilities $\hat{q}(c)$
   ($c = 0,\ldots,C$) by setting them to the corresponding values
   obtained from the normalized approach.
2. Calculate $\mu_{eff}$ using(12)by replacing $q(c)$'s with $\hat{q}(c)$'s.
3. Calculate $\hat{q}(c)$ for $c = 0,\ldots,C$ using(14)and(15).
4. Obtain new call blocking and handoff call dropping
   probabilities using $\hat{q}(c)$.
```

Although Gersht and Lee suggested an iterative approach for the solution of the knapsack problem, we present our approximation method in closed form since once we compute the effective channel occupancy time, μ_{eff}, from (12) using the initial conditions obtained from the *normalized* approach and the corresponding values of the estimated equilibrium occupancy probabilities, $\hat{q}(c)$, from (14) and (15) followed by that, the value of the recomputed effective channel occupancy time using the estimated equilibrium occupancy probabilities will remain the same which will also stabilize the values of the estimated equilibrium occupancy probabilities. After the estimated equilibrium occupancy probabilities are obtained, the new call blocking and handoff call dropping probabilities are calculated as follows:

$$p_{nb}^{eff} = 1 - \frac{\sum_{j=0}^{C-1} \lambda_n \cdot \beta_j \cdot \hat{q}(j)}{\sum_{j=0}^{C} \lambda_n \cdot \hat{q}(j)} \tag{16}$$

$$p_{hd}^{eff} = 1 - \frac{\sum_{j=0}^{C-1} \lambda_h \cdot \hat{q}(j)}{\sum_{j=0}^{C} \lambda_h \cdot \hat{q}(j)} \tag{17}$$

2.4 Numerical Results

2.4.1 Performance Evaluation of Existing and Proposed Methods

In this section we present numerical results of performance evaluations using our novel *effective holding time* approach presented in Section 2.3 and compare them with those obtained using the existing *traditional* and the *normalized* approaches based on one dimensional Markov chain models. We also obtain accurate results using a multidimensional Markov chain model for comparison purposes. This is accomplished using a numerical method called *direct* (also known as *LU decomposition*) to calculate the exact values of the performance metrics and the corresponding results are labeled as “direct method”.

The numerical results obtained for this study not only show that the results obtained from the *normalized* and the *traditional* approaches can deviate significantly from the accurate results obtained from the *direct* approach, but also show that our new approach, *effective holding time*, can achieve results very close to the accurate *direct* ones.

We do not give the results here for the *new call bounding scheme* since the *normalized* approach can overcome the inaccuracy of the *traditional* approach which overestimates the new call blocking probability while it underestimates the handoff call dropping probability when the handoff call traffic is higher than the new call traffic load (i.e.,

$\rho_h > \rho_n$) or vice versa by exploiting the symmetric nature of the scheme's state transitions and thus leaves little room for improvement. However, we focus on the other schemes considered above for which this property does not apply.

Since *fractional guard channel* is a generalization of the *cutoff priority* scheme, we evaluate only the *cutoff priority scheme* here due to space constraints as same property applies for both schemes. Before examining this scheme to evaluate its performance, we should determine the range of values that system parameters such as λ_n, λ_h, $1/\mu_n$ and $1/\mu_h$ take in order to reflect practical situations.

It is generally accepted in the literature [3][7][11] to set the arrival rate, λ_n, and the average channel occupancy time, $1/\mu_n$, of new calls in proportion with the arrival rate, λ_h, and the average channel occupancy time, $1/\mu_h$, of handoff calls, respectively. Therefore following the scenarios that have been considered in the literature, we apply the values which are grouped under four different cases and presented in Table 2.1 as the arrival rates and average channel occupancy times for new and handoff calls in order to evaluate the performances of the selected schemes. To put it shortly, we assume both ratios, λ_n/λ_h, μ_h/μ_n, to have values changing within the range of 4 and 0.5 in order to cover the scenarios commonly considered in the literature. We set the total number of channels, C, to 30 and the channel occupancy threshold, m, to 25 as in [7].

Reducing handoff call dropping by assigning higher priorities or other means increases the probability of blocking for new calls and thus results in a tradeoff between these two QoS measures [6]. Nevertheless, the goal of a network service provider is to maximize the revenue by improving network resource utilization, which is usually associated with minimizing the new call blocking probability while keeping the handoff dropping probability below a certain threshold. Hence we evaluate the approximate evaluation methods mentioned above by grouping the possible scenarios into four different cases with parameter values chosen as shown in Table 2.1 in order to obtain handoff call dropping probabilities within the range of 0.01 and 0.1 since this is the interval of interest when providing QoS guarantees in a cellular network.

In the literature the commonly accepted method to evaluate a scheme is to simulate a system modeled with that scheme with call arrival rates being varied to change the traffic load while the average channel occupancy times of different types of calls are kept constant.

However, considering that the objective of this study is to relax the assumption made by previously suggested approximation methods for different types of calls to have different average channel occupancy times, we simulate the system by varying the average channel occupancy times instead as in [7], since having different average channel occupancy times for new and handoff calls is what makes the existing methods inaccurate. One can also notice that blocking/dropping probability values obtained by (4) and (5) or (8) and (9), which were derived for *cutoff priority* and *fractional guard channel* schemes, respectively, by using the *normalized* approach, remain the same as long as the call arrival rates and average channel occupancy times change with same proportions, thus keeping the traffic loads for both types of calls, ρ_h and ρ_n, constant. Yet, the direct method is expected to give different results in such a case as it considers the call arrival rates and average channel occupancy times separately. Considering that average channel occupancy times play a stronger role compared to call arrival rates when obtaining call blocking probabilities accurately using given approximation methods, varying average channel occupancy times of both types of calls in proportion with each other can be justified as we expect that it would be more challenging to obtain accurate results when differences in average channel occupancy times exist.

In Figs. 2.3-2.10, we show the new call blocking and handoff call dropping probabilities computed by the mentioned approaches for each case given in Table 2.1. In Figs. 2.3 and 2.4, for simulation scenario 1, it is observed that as the new call traffic load gets higher, the traditional approach overestimates both the new call blocking and handoff call dropping probabilities by the largest margin while the normalized approach underestimates the handoff call dropping probability. However, the results obtained by our proposed *effective holding time* approximation method match the curve obtained by using the direct method very well. A similar conclusion can be drawn for the comparisons of new call blocking and handoff call dropping probabilities from Figs. 2.5 and 2.6 for simulation scenario 2, Figs. 2.7 and 2.8 for scenario 3 and Figs. 2.9 and 2.10 for scenario 4, respectively. As expected and mentioned above, we observe from Figs. 2.2, 2.4, 2.6 and 2.8 that the results obtained by the proposed approximation method diverge from the exact ones slowly when the ratio of average channel occupancy times for new to handoff calls increases. In short, when the new and handoff calls have different average channel occupancy times, the *traditional* and the *normalized* approaches result in significant discrepancies compared to the *direct* method especially with respect to handoff call dropping probability, which is overestimated by the former approach while it is underestimated by the latter one. However, the results obtained by the proposed approach can overcome such inaccuracy.

2.4.2 Accuracy and Runtime Computational Costs

In this section we examine the accuracy of *traditional*, *normalized*, and the proposed *effective holding time* approximation methods by using the "mean average error (MAE)" and the "root mean square error (RMSE)" which are calculated as given below.

$$MAE = \frac{1}{N} \cdot \sum_{i=1}^{N} \frac{abs(p_{i,estimated} - p_{i,real})}{p_{i,real}} \tag{18}$$

$$RMSE = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} \left(\frac{p_{i,estimated} - p_{i,real}}{p_{i,real}} \right)^2} \tag{19}$$

The results are given in Table 2.2 for handoff call dropping and Table 2.3 for new call blocking probabilities. Further to the results in Figs. 2.3-2.10, the most significant results on accuracy are shown in Table 2.2 for handoff call dropping probabilities, where our proposed *effective holding time* approximation method reduces estimation errors substantially compared to the existing approximation methods. Even though our proposed method estimates the new call blocking probabilities with very small errors as shown in Table 2.3, it does not significantly reduce them with respect to the estimation errors given by the existing methods due to the relatively small error margins.

The percentage gains in accuracy of the results on handoff call dropping probability obtained by the proposed approximation method relative to those obtained by the *normalized* and the *traditional* methods are given in Table 2.4. The percentage gains in accuracy of the results on new call blocking probability obtained by the proposed approximation method are not given here since they are very small compared with the results on handoff call dropping shown in Table 2.4. In Table 2.4, it is observed that the percentage gains in estimation accuracy provided by our proposed approximation method relative to the existing methods decrease as the ratio of new to handoff call holding times decreases even though the gains obtained still remain significant.

The reason why an acceptable approximation method is needed to evaluate the performance of a call admission control scheme when an exact solution with a numerical method based on multidimensional Markov chain modeling exists, is to avoid solving large sets of flow equations and therefore the curse of dimensionality. To give the reader a better idea regarding the "*CPU time*" and the amount of "*memory*" used for evaluating the performance of any of the policies mentioned above, we implement one direct and two widely used iterative methods, which are *direct (LU decomposition)*, *method of Jacobi* (iterative), and *method of Gauss-Seidel* (iterative), in order to compare their runtime computational costs with that of the proposed method. The results are given in Table 2.5 for three different scenarios of total and shared number of channels, where "*CPU time*" represents the total processing time (in seconds) the CPU spent from the time that the computation was started for each method, "*number of iterations*" represents the number of times that each algorithm (except the direct one) needs to iterate before it converges with respect to the chosen tolerance value, and "*used memory*" represents the amount of storage allocated for nonzero matrix elements.

For the simulation results presented in Table 2.5, the following parameters are chosen respectively for the three different scenarios: (a) $C = 6$, $m = 5$, $\lambda_h = 0.0067$, (b) $C = 30$, $m = 25$, $\lambda_h = 0.0334$, (c) $C = 60$, $m = 50$, $\lambda_h = 0.1667$, while $\mu_n = 1/600$, $\mu_h = 1 /300$, and λ_n varies from 1/600 to 1/50, 1/120 to 1/10 and 1/2 to 1/24, respectively. The new and handoff call arrival rates are chosen to obtain similar values of call blocking/dropping probabilities with the previously computed ones in each scenario in order to make the comparisons appropriate.

As seen in Table 2.5, as the number of channels increases, the values of *CPU time* for the numerical solution methods (both direct and iterative) become significantly greater than the corresponding value for the proposed method. The same observation can also be made for the *used memory*. This should not be surprising since our proposed approximation method has much smaller number of states in its model and a closed form formulation.

2.5 Summary

In this chapter we have examined various guard channel based call admission control schemes in wireless mobile networks to evaluate their performance analytically by using a one dimensional Markov chain model. When the average channel occupancy times for new and handoff calls are significantly different, we have shown that the *traditional* and the *normalized* approaches may not be appropriate to use due to their discrepancies in comparison with the exact results. Even though using a two dimensional Markov chain model could solve this problem and yield exact results, it gives rise to another problem known as the curse of dimensionality since the dimension of the state space in such a model can increase very quickly. With the objective of providing a practical and closed form solution to this problem, we have proposed a new method called *effective holding time*, which gives more accurate results when compared to the existing approximation approaches while keeping the computational cost low.

We have evaluated the accuracy of the proposed method by comparing it with the exact results obtained from the *direct* method based on a multidimensional Markov chain model. When compared with the existing *traditional* and *normalized* approximate methods, it is observed that the proposed *effective holding time* method outperforms the others especially with its high percentage gain in accuracy when computing the handoff call dropping probability. To demonstrate that the proposed *effective holding time* method has very low runtime computational cost, we have presented results showing that the *CPU time* and the amount of *memory* used by the proposed method are very low compared to the direct and iterative numerical methods that can be employed to obtain exact results.

As computational cost plays an important role in real time applications, we believe a better approximation method with low complexity for evaluating the performance of CAC schemes will motivate the practical implementation of these schemes when providing dynamic call admission control in the future. However when a similar performance evaluation problem is addressed with the proposed method in a multi-cell network, not only call arrival rates and call duration times but also changing cell capacity should be considered due to its dependence on the location of users in neighboring cells. We are extending this work to include analytical performance evaluation of multi-service models with multiple channel requests, considering that the level of relative prioritization provided to different service types with different QoS requests is specified by relative blocking/dropping probabilities.

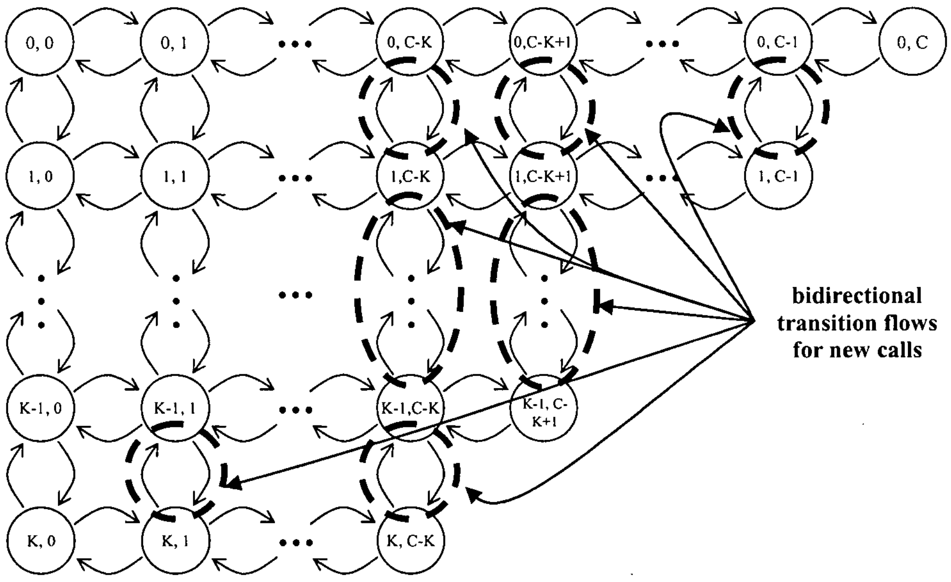

Fig. 2.1 State transition diagram for the new call bounding scheme.

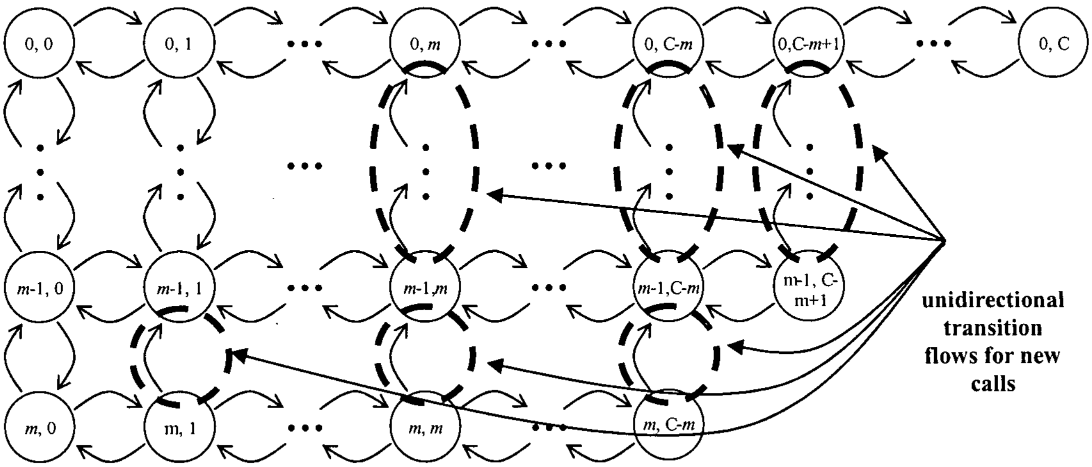

Fig. 2.2. State transition diagram for the cutoff priority scheme.

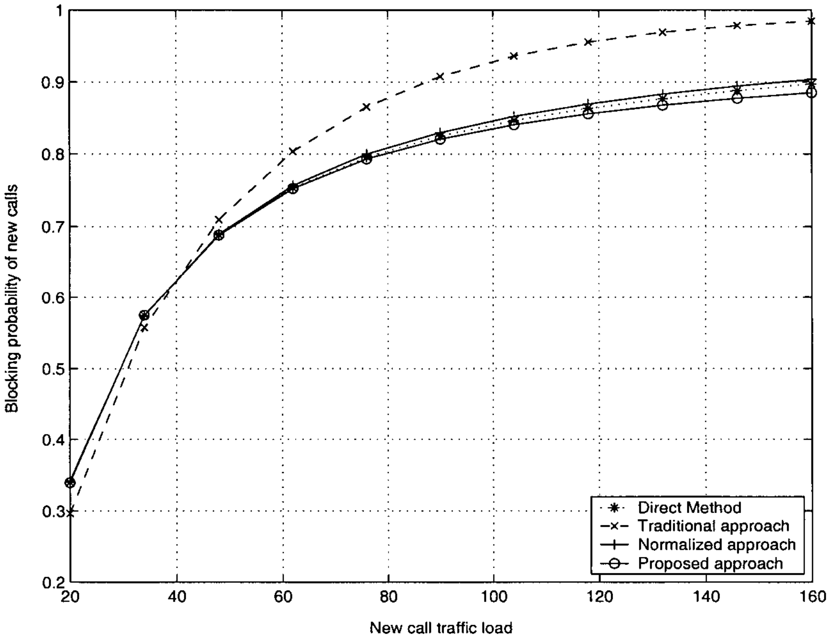

Fig. 2.3 New call blocking probability in the cutoff priority scheme (Case I).

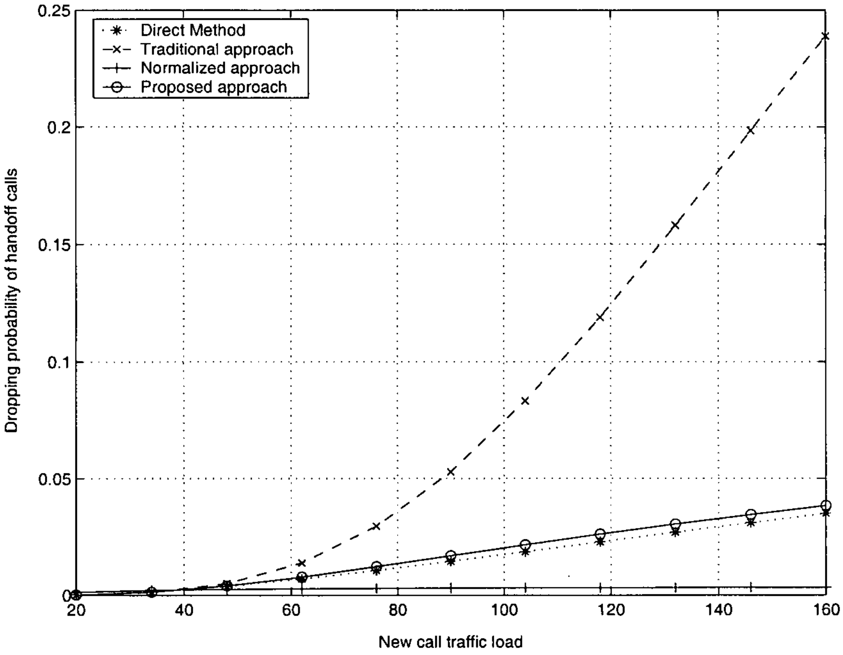

Fig. 2.4 Handoff call dropping probability in the cutoff priority scheme (Case I).

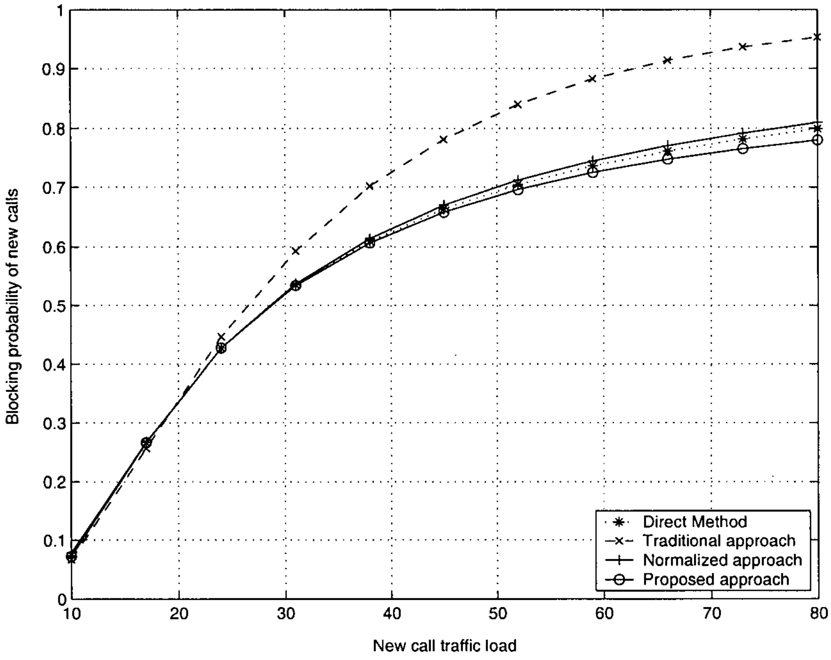

Fig. 2.5 New call blocking probability in the cutoff priority scheme (Case II).

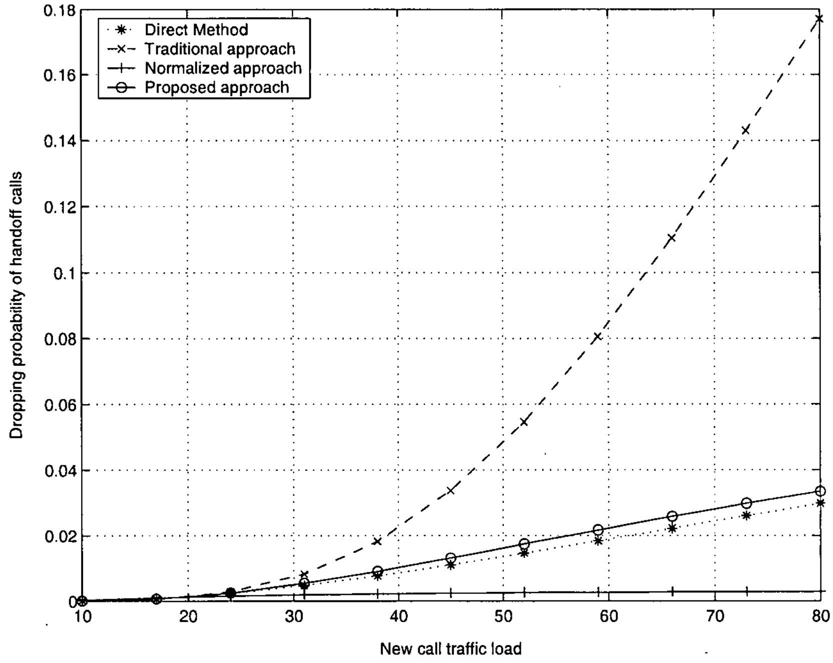

Fig. 2.6 Handoff call dropping probability in the cutoff priority scheme (Case II).

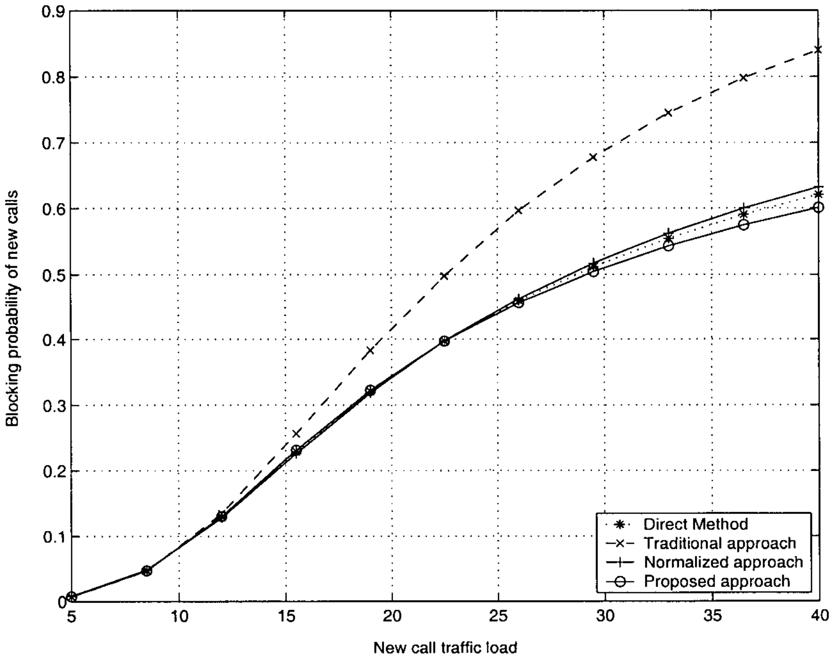

Fig. 2.7 New call blocking probability in the cutoff priority scheme (Case III).

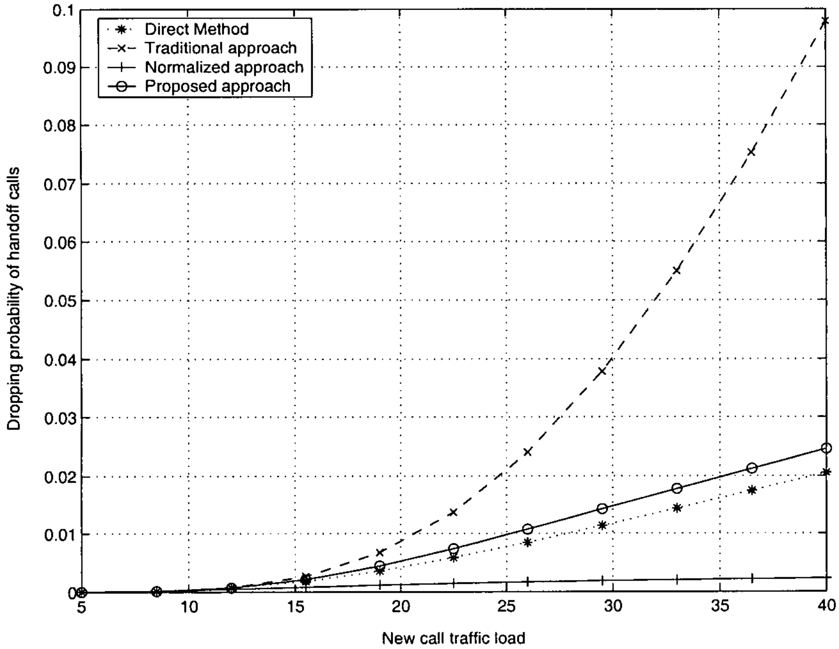

Fig. 2.8 Handoff call dropping probability in the cutoff priority scheme (Case III).

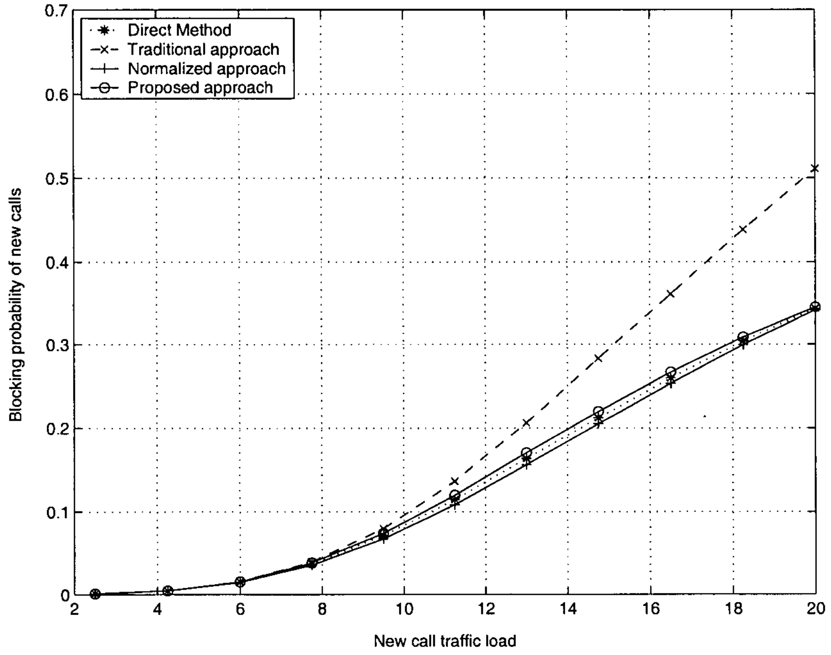

Fig. 2.9 New call blocking probability in the cutoff priority scheme (Case IV).

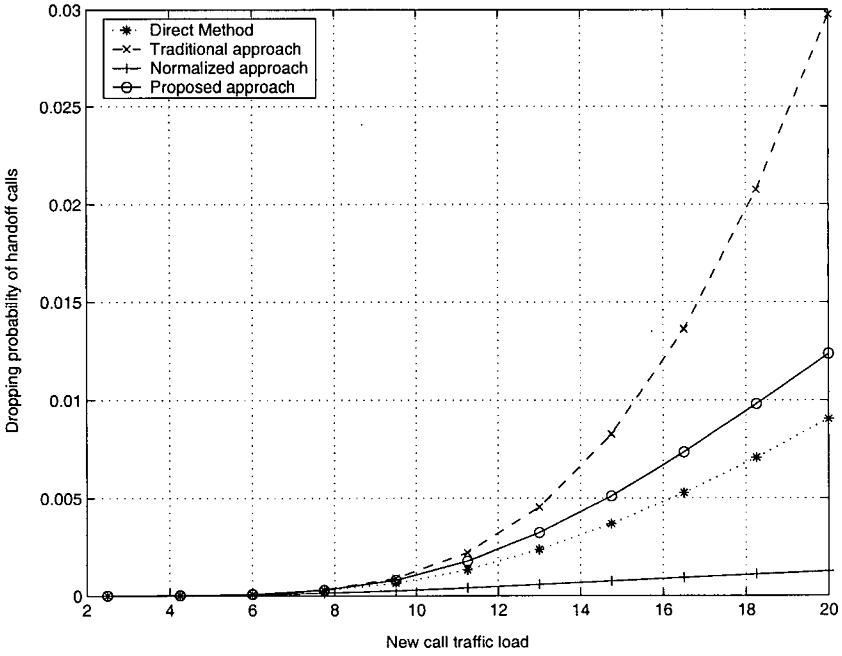

Fig. 2.10 Handoff call dropping probability in the cutoff priority scheme (Case IV).

TABLE 2.1

System parameter values used for each scenario

Cases	λ_n	λ_h	$1/\mu_n$ *(sec)*	$1/\mu_h$ *(sec)*
I	1/40 - 1/5	1/20	800	200
II	1/40 - 1/5	1/20	400	200
III	1/40 - 1/5	1/20	200	200
IV	1/40 - 1/5	1/20	100	200

TABLE 2.2

Errors in handoff call dropping probability approximations relative to direct method

	Traditional approach		*Normalized approach*		*Proposed approach*	
Cases	*MAE*	*RMSE*	*MAE*	*RMSE*	*MAE*	*RMSE*
I	2.76	3.42	1.07	1.53	0.14	0.15
II	2.23	2.81	0.95	1.23	0.16	0.18
III	1.57	2.02	0.79	0.89	0.21	0.22
IV	0.88	1.15	0.66	0.70	0.27	0.30

TABLE 2.3

Errors in new call blocking probability approximations relative to direct method

	Traditional approach		*Normalized approach*		*Proposed approach*	
Cases	*MAE*	*RMSE*	*MAE*	*RMSE*	*MAE*	*RMSE*
I	0.09	0.09	0.005	0.006	0.006	0.007
II	0.15	0.16	0.013	0.02	0.012	0.015
III	0.21	0.25	0.018	0.028	0.016	0.02
IV	0.21	0.27	0.035	0.039	0.026	0.03

TABLE 2.4

Percentage gains in accuracy of handoff call dropping probabilities obtained by the proposed approximation method relative to those obtained by the *normalized* and the *traditional* methods

	% Gain over traditional approach		*% Gain over normalized approach*	
Cases	*MAE*	*RMSE*	*MAE*	*RMSE*
I	94.95	95.55	86.98	90.07
II	92.73	93.75	82.97	85.75
III	86.70	89.08	73.62	75.23
IV	68.78	74.00	58.44	57.20

TABLE 2.5
Comparison of runtime computational costs between the proposed method and the *direct* and *iterative* numerical methods

	Total number of channels = C, total number of shared channels = m								
	$C = 6$, $m = 5$			$C = 30$, $m = 25$			$C = 60$, $m = 50$		
Numerical Methods	CPU time	number of iterations	used memory	CPU time	number of iterations	used memory	CPU time	number of iterations	used memory
LU decomposition	0.156s	-	4157	3m14s	-	1164845	5h35m36s	-	16884530
Method of Jacobi	2.172s	1000*	2728	4m8s	1000*	702606	1h25m50s	1000*	10144576
Method of Gauss-Seidel	0.406s	24	2728	3m49s	59	702606	2h41m17s	172	10144576
Proposed method	0.109s	-	170	0.266s	-	698	1.844s	-	1358

* indicates that the number of iterations for the method of Jacobi is limited to 1000 in each scenario due to divergence to the desired tolerance.

2.6 Bibliography

[1] D. E. Everitt, "Traffic engineering of the radio interface for cellular mobile networks," *Proc. IEEE*, vol. 82, no. 9, pp.1371-1382, 1994.

[2] H. Chen, L. Huang, S. Kumar, and C. C. J. Kuo, *Radio Resource Management for Multimedia QoS Support in Wireless Networks*. Boston: Kluwer Academic Publishers, 2004, chapter 2.

[3] D. Hong and S. S. Rappaport, "Traffic model and performance analysis for cellular mobile radiotelephone systems with prioritized and non-prioritized handoff procedures," *IEEE Transactions on Vehicular Technology*, vol. 35, pp. 77-92, Aug. 1986.

[4] B. Li, C. Lin, and S. T. Chanson, "Analysis of a hybrid cutoff priority scheme for multiple classes of traffic in multimedia wireless networks," *Wireless Networks*, vol. 4, no. 4, pp. 279-290, July 1998.

[5] Y.B. Lin, S. Mohan, and A. Noerpel, "Queuing priority channel assignment strategies for handoff and initial access for a PCS network," *IEEE Transactions on Vehicular Technology*, vol. 43, no. 3, pp. 704-712, Aug. 1994.

[6] R. Ramjee, R. Nagarajan, and D. Towsley, "On optimal call admission control in cellular networks," *Wireless Networks*, vol. 3, no. 1, pp. 29-41, March 1997.

[7] Y. Fang, and Y. Zhang, "Call admission control schemes and performance analysis in wireless mobile networks," *IEEE Transactions on Vehicular Technology*, vol. 51, no. 2, pp. 371-382, March 2002.

[8] M. D. Kulavaratharasah, and A. H. Aghvami, "Teletraffic performance evaluation of microcell personal communication networks (PCNs) with prioritized handoff procedures," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 1, pp. 137-152, Jan. 1999.

[9] R. A. Guerin, "Queuing-blocking system with two arrival streams and guard channels," *IEEE Transactions on Communications*, vol. 36, no. 2, pp. 153-163, Feb. 1988.

[10] E. D. Re, R. Fantacci, and G. Giambene, "Handover queuing strategies with dynamic and fixed channel allocation techniques in low earth orbit mobile satellite systems," *IEEE Transactions on Communications*, vol. 47, no. 1, pp. 89-102, Jan. 1999.

[11] C. H. Yoon, and C. K. Un, "Performance of personal portable radio telephone systems with or without guard channels," *IEEE Journal on Selected Areas in Communications*, vol. 11, no. 6, pp. 911-917, Aug. 1993.

[12] C. Chang, C. J. Chang, and K. R. Lo, "Analysis of a hierarchical cellular system with reneging and dropping for waiting new calls and handoff calls," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 4, pp. 1080-1091, July 1999.

[13] V. K. N. Lau, and S. V. Maric, "Mobility of queued call requests of a new call queuing technique for cellular systems," *IEEE Transactions on Vehicular Technology*, vol. 47, no. 2, pp. 480-488, May 1998.

[14] A. S. Acampora and M. Naghshineh, "Control and quality of service provisioning in high-speed micro-cellular networks," *IEEE Personal Communications*, vol. 1, no. 2, pp. 36-43, 1996.

[15] M. Naghshineh and S. Schwartz, "Distributed call admission control in mobile/wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 14, no. 4, pp. 711-717, May 1996.

[16] D. Levine, I. Akyildiz, and M. Naghshineh, "A resource estimation and call admission algorithm for wireless multimedia networks using the shadow cluster concept," *IEEE/ACM Transactions on Networking*, vol. 5, no. 1, pp. 1-12, Feb. 1997.

[17] C. Oliveira, J. B. Kim, and T. Suda, "An adaptive bandwidth reservation scheme for high-speed multimedia wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 858-874, Aug. 1998.

[18] P. Ramanathan, K. M. Sivalingam, P. Agrawal, and S. Kishore, "Dynamic resource allocation schemes during handoff for mobile multimedia wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 7, pp. 1270-1283, July 1999.

[19] R. Guerin, "Equivalent capacity and its application to bandwidth allocation in high-speed networks," *IEEE Journal on Selected Areas in Communications*, vol. 9, no. 7, pp. 968-981, Sept. 1991.

[20] J. S. Evans and D. Everitt, "Effective bandwidth-based admission control for multi-service CDMA cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 1, Jan. 1999.

[21] Q. Ren and G. Ramamurthy, "A real-time dynamic connection admission controller based on traffic modeling, measurement, and fuzzy logic control," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 2, Feb. 2000.

[22] Y. Fang and I. Chlamtac, "Teletraffic analysis and mobility modeling for PCS networks," *IEEE Transactions on Communications*, vol. 47, pp. 1062-1072, July 1999.

[23] Y. Fang, I. Chlamtac, and Y. B. Lin, "Channel occupancy times and handoff rate for mobile computing and PCS networks," *IEEE Transactions on Computers*, vol. 47, pp. 679-692, June 1998.

[24] W. Li and X. Chao, "Modeling and performance evaluation of a cellular mobile network," *IEEE/ACM Transactions on Networking*, vol. 12, no. 1, Feb. 2004.

[25] S. S. Rappaport, "The multiple call handoff problem in personal communications networks," *IEEE 40th Vehicular Technology Conference*, pp. 287–294, May 1990.

[26] S. S. Rappaport and G. Monte, "Blocking, hand-off and traffic performance for cellular communication systems with mixed platforms," *IEEE 42nd Vehicular Technology Conference*, vol. 2, pp. 1018–1021, May 1992.

[27] K. W. Ross, Multi-service Loss Models for Broadband Telecommunication Networks. London: Springer-Verlag, 1995, chapter 3.

[28] A. Gersht and K. J. Lee, "A bandwidth management strategy in ATM networks," Technical report, GTE Laboratories, 1990.

CHAPTER 3 COMPUTATIONALLY EFFICIENT PERFORMANCE EVALUATION METHODS FOR CALL ADMISSION CONTROL SCHEMES IN MULTI-SERVICE CELLULAR NETWORKS[2]

3.1 Introduction

The emerging global standard for next generation wireless networks has promised to provide not only conventional voice services but also the efficiency and flexibility of multiplexing a wide variety of traffic from data to multimedia applications due to increasing demand coming from users that these services shall also be available on the move. However satisfying the diverse QoS requirements of these services over cellular networks has become even more challenging due to reduced cell size and hence increased user mobility. Call admission control (CAC) schemes are deployed to selectively limit the number of admitted calls from each QoS class to maximize the network utilization while satisfying the QoS constraints [1]. Call admission control for wired and wireless networks has been intensively studied in the past and many priority based call admission control schemes have been proposed [2]-[14]. Calls with more stringent QoS requirements are given higher priorities by having exclusive access to a number of reserved channels. Reducing blocking probabilities of calls with higher priorities increases the probability of blocking for calls with relatively lower priorities resulting in a tradeoff between QoS classes. The goal is to sustain a balance between QoS classes while satisfying the respective QoS requirements. Handoff calls are given priority over new ones to minimize call dropping probabilities since dropping an ongoing call is generally more objectionable to a mobile user than blocking a new call request.

A set of guard channels are reserved for prioritized calls in *Guard Channel (GC)* schemes such as *cutoff priority* [2]-[5], *fractional guard channel* [6], *new call bounding* [7]

[2] **A version of this chapter has been accepted for publication. E. A. Yavuz and V. C. M. Leung, "Efficient Approximations for Call Admission Control Performance Evaluations in Multi-Service Networks,"** ***IEEE GLOBECOM 06*****, San Francisco, CA, November 2006.**

and *rigid division based* [8] schemes. Many dynamic GC schemes have also been proposed to maximize the network utilization by reserving network resources for relatively higher priority calls adaptively so that more relatively lower priority calls can be admitted [9]-[14]. Efficient adaptive reservation depends on reliable and up-to-date system status feedback; however exact analyses of these schemes using multidimensional Markov chain models are intractable in real time due to the need to solve large sets of flow equations. Hence, performance metrics such as call blocking probabilities are generally evaluated using one dimensional Markov chain models based on circuit-switched network architecture under the simplifying assumptions that call arrivals are Poisson, channel occupancy times are exponentially distributed with equal mean values and traffic classes have same capacity requirements. Due to the popularity of Internet and multimedia applications, increasingly traffic carried over wireless networks are packet-switched and statistically multiplexed over shared channels to improve network utilization, which makes performance evaluation of call admission control schemes harder due to the dynamic nature of the traffic. This difficulty can be overcome by using the effective bandwidth [15]-[17] to represent the traffic demand of a packet-switched traffic stream so that application of the above schemes to a packet-switched network can still be evaluated using Markov chain models. This approach has been successfully applied to cellular network by Evans and Everitt in [16].

However, the simplifying assumptions mentioned above may not be appropriate in many situations since calls with different priorities may have different average channel occupancy times if not different distributions [18] [19]. Existing performance evaluation methods based on one dimensional Markov chain approximations, such as *traditional* and *normalized*, lead to significant discrepancies when average channel occupancy times for distinct QoS classes are different [20]. Thus exact analysis methods based on multidimensional Markov chain models appeared to be the only means to obtain accurate solutions for evaluating call admission control schemes. In [21], Rappaport obtained call blocking probabilities for calls of various priorities in a cellular network by using a multidimensional model, whereas with Monte, the authors developed an analytical model for traffic performance analysis using a multidimensional birth death process to take into consideration the effects of various platform types distinguished by different mobility characteristics on performance [22]. These methods suffer from the curse of dimensionality, which results in very high computational cost for large systems, despite providing the exact solutions.

Approximation methods for performance evaluations that have a high accuracy and low computational cost are needed if dynamic call admission control schemes are to be implemented in real time systems that adapt to dynamic changes in traffic statistics. Li and Chao [23] obtained a product form solution by modeling a multicell wireless network as a network of queues employing a hybrid *GC/QP* scheme with transfer of unsuccessful requests to neighboring cells; however their solution is restrictive to the protocol considered and therefore may not be appropriate to be used for the performance evaluation of multi-service models. In [24], Gersht and Lee proposed an iterative algorithm by modifying the approximation suggested by Roberts [25] to improve its accuracy when the service rates differ. However we showed in [20] that starting with an inappropriate initial value leads to significant discrepancies and thus proposed an easy to implement closed form approximation method based on one dimensional Markov chain modeling. We assumed that all classes have same capacity requirements and independent and exponentially distributed channel occupancy times without the necessity of having the same average values. Yet this method applies only when the call traffic is homogeneous. In the absence of a product form solution when capacity requirements of various classes differ, call traffic is heterogeneous, calculating the channel occupancy distribution involves solving the balance equations numerically, which is demanding for all but smallest channel capacities. In [26], Borst and Mitra developed computational algorithms for the multi-service case by coupling the computation of joint channel occupancy probabilities with that of used capacity assuming that channels are occupied independently. The authors solved the balance equations through numerical iteration but the results can only be comparative when the number of existing call arrival types are high due to authors' channel occupancy independence assumption. In this chapter, we classify call admission control schemes into two novel categories based on the nature of connecting links in a scheme's transition diagram; *symmetric* and *asymmetric*. We present performance evaluation methods with high computational efficiency for each category under the simplifying assumptions that call arrival and channel occupancy times for all QoS classes are exponentially distributed, but with different average values in general.

This chapter is organized as follows. In the next section we obtain the product form exact solution formula to evaluate *symmetric* call admission control schemes in multi-service networks. In Section 3.3, we propose a novel performance evaluation approximation method, which we call *state space decomposition*, to evaluate *asymmetric* call admission control schemes in multi-service networks. Section 3.4 presents the numerical results to compare the approximation method proposed in Section 3.3 with the exact analysis and previously proposed approximation methods. We show that the runtime computational cost of the proposed approximation method is significantly lower than that of the exact analysis. Section 3.5 concludes the chapter.

3.2 Performance Evaluation of Symmetric Call Admission Control Schemes

We define a call admission control scheme *symmetric* if each pair of nodes are connected by two unidirectional links in opposite directions in state transition diagram of the scheme's Markov chain model. The widely known *complete sharing (CS)*, *complete partitioning (CP)* and *new call bounding* schemes can then be regarded as symmetric. In this section we obtain a product form exact solution formula to evaluate symmetric call admission control schemes in multi-service networks where all QoS classes have distinct capacity requirements. We consider a cellular system employing a symmetric scheme, new call bounding in this case, and two classes of calls for the benefit of simplicity although any number of classes could be possible; non-prioritized and prioritized, where the latter enjoy a high service priority than the former: Let λ_{np} and λ_p denote the arrival rates, and $1/\mu_{np}$ and $1/\mu_p$ denote the average channel occupancy times for non-prioritized and prioritized calls respectively. Let C denote the total number of channels in a cell and b_{np} and b_p denote the required bandwidth for non-prioritized and prioritized calls, respectively. We assume that the arrival processes for both types of calls are Poisson, and their channel occupancy times are exponentially distributed. Then, let K be the threshold between 0 and C for the new call bounding scheme and a non-prioritized call is admitted only when there are less than K channels occupied by non-prioritized calls in the network.

The *traditional* method, which uses a one dimensional Markov chain model with a fixed average channel holding time for total cell traffic, leads to inaccurate results when average channel occupancy times for all types of calls differ. A product form solution is presented in [7] to obtain the blocking probabilities of non-prioritized, B_{np} and prioritized calls, B_p accurately by exploiting the symmetric nature of the scheme assuming that all QoS classes have same capacity requirements. The authors [7] normalized the average channel occupancy time for both types of calls to allow the arriving traffic for each type of call to be scaled appropriately. In this section we obtain a product form solution where all QoS classes have distinct capacity requirements.

Let $\rho_{np} = \lambda_{np} / \mu_{np}$ and $\rho_p = \lambda_p / \mu_p$, then the prioritized and non-prioritized Poisson call arrival stream is Poisson with arrival rates ρ_p and ρ_{np}, respectively and service rates (corresponding channel occupancy times) that are equal to 1, are equivalent to the original streams with respect to the Markov chain model since only traffic loads are required to obtain the stationary distributions. Yet, the arrival rates in the original system are different. Let $q(n_{np},n_p)$ denote the steady state probability that there are n_{np} non-prioritized calls and n_p prioritized calls in the system. Then we obtain the following stationary distribution:

$$0 \le n_{np}b_{np} \le K$$

$$q(n_{np},n_p) = \frac{\rho_{np}^{\,n_{np}}}{n_{np}!} \cdot \frac{\rho_p^{\,n_p}}{n_p!} \cdot q(0,0), \qquad (n_{np}b_{np} + n_pb_p) \le C$$

$$n_p,\ n_{np} \ge 0$$

where

$$q(0,0) = \left[\sum_{(n_{np},n_p)\in S} \frac{\rho_{np}^{\,n_{np}}}{n_{np}!} \cdot \frac{\rho_p^{\,n_p}}{n_p!} \right]^{-1}$$

$$= \left[\sum_{n_{np}=0}^{\lfloor K/b_{np} \rfloor} \frac{\rho_{np}^{\,n_{np}}}{n_{np}!} \cdot \sum_{n_p=0}^{\lfloor C-((n_{np}\cdot b_{np})/b_p) \rfloor} \frac{\rho_p^{\,n_p}}{n_p!} \right]^{-1}$$

and $\lfloor \ldots \rfloor$ represents the "floor" function which rounds its input to the nearest integer less than or equal to the value of the input itself. Thus, the formulas for non-prioritized and prioritized call blocking probability are as follows:

$$B_{np} = \sum_{n_{np}=0}^{\lfloor K/b_{np} \rfloor} \sum_{n_p=0}^{\lfloor (C-(n_{np} \cdot b_{np}))/b_p \rfloor} q\ (n_{np}, n_p) \tag{1}$$

$$\{(n_p b_p + n_{np} b_{np}) \geq C\}\ U\ \{n_{np} b_{np} \geq K\}$$

$$= \sum_{n_{np}=0}^{\lfloor K/b_{np} \rfloor} \sum_{n_p=0}^{\lfloor (C-(n_{np} \cdot b_{np}))/b_p \rfloor} \frac{\dfrac{\rho_{np}{}^{n_{np}}}{n_{np}!} \cdot \dfrac{\rho_p{}^{n_p}}{n_p!}}{\left[\displaystyle\sum_{n_{np}=0}^{\lfloor K/b_{np} \rfloor} \frac{\rho_{np}^{n_{np}}}{n_{np}!} \cdot \sum_{n_p=0}^{\lfloor C-((n_{np} \cdot b_{np})/b_p) \rfloor} \frac{\rho_p^{n_p}}{n_p!} \right]}$$

$$B_p = \sum_{n_p=0}^{\lfloor K/b_{np} \rfloor} q\ \left(n_p, \lfloor (C-(n_p \cdot b_{np}))/b_p \rfloor\right) \tag{2}$$

$$= \sum_{n_p=0}^{\lfloor K/b_{np} \rfloor} \frac{\dfrac{\rho_{np}{}^{n_p}}{n_p!} \cdot \dfrac{\rho_p{}^{(\lfloor (C-(n_p \cdot b_{np}))/b_p \rfloor)}}{(\lfloor (C-(n_p \cdot b_{np}))/b_p \rfloor)!}}{\left[\displaystyle\sum_{n_{np}=0}^{\lfloor K/b_{np} \rfloor} \frac{\rho_{np}^{n_{np}}}{n_{np}!} \cdot \sum_{n_p=0}^{\lfloor C-((n_{np} \cdot b_{np})/b_p) \rfloor} \frac{\rho_p^{n_p}}{n_p!} \right]}$$

When $K = C$, the *new call bounding* scheme becomes the non-prioritized scheme, however non-prioritized, B_{np}, and prioritized, B_p, call blocking probabilities will not be the same until $b_{np} = b_p$. Apparently when $b_{np} = b_p = 1$, both probabilities become:

$$B_{np} = B_p = \frac{\dfrac{(\rho_{np} + \rho_p)^C}{C!}}{\displaystyle\sum_{j=0}^{C} \frac{(\rho_{np} + \rho_p)^j}{j!}}$$

3.3 Performance Evaluation of Asymmetric Call Admission Control Schemes

We define a call admission control scheme *asymmetric* when some pairs of nodes in the state transition diagram of scheme's Markov chain model have unidirectional links only in one direction. The widely known *cutoff priority* and *fractional guard channel* schemes, which make decisions on whether an arriving non-prioritized call is going to be accepted or not based on the number of total occupied channels in the system, can both then be regarded as asymmetric schemes. We consider a cellular system with two classes of calls where prioritized calls enjoy a higher service priority than the non-prioritized ones. Let λ_p, λ_{np}, μ_p, μ_{np}, b_p, b_{np}, β_i, m and C be defined as before and $q_p(j)$ and $q_{np}(r)$ denote the estimated equilibrium channel occupancy probabilities when exactly j prioritized calls and r non-prioritized calls, respectively, exist in the system. Let β_i $(i = 0, 1,\ldots, C - 1)$ denote the admission probability of an arriving non-prioritized call when the number of busy channels is i and k_j $(j = 0,1,\ldots (\lfloor C/b_p \rfloor - 1))$ denote the admission probability of an arriving prioritized call when j prioritized calls exist in the cell regardless of the number of existing non-prioritized calls. Thus k_j is similar to β_i, however β_i is a predefined user controlled parameter that indicates whether an arriving non-prioritized call will be admitted or not based on the number of occupied channels in the system as opposed to k_j which is extracted from the multidimensional model of the system.

We present the following novel performance evaluation approximation method referred as *state space decomposition*. Instead of evaluating the system using a one dimensional Markov chain model by grouping the nodes with the same total number of occupied channels regardless of the types of calls, we group the nodes with the same number of calls of a certain type to obtain "*supernodes*" to compose a one dimensional Markov chain model for each type of call. By grouping nodes with the same number of prioritized calls such as (0,0), $(b_{np},0),\ldots,(b_{np}(\lfloor m/b_{np} \rfloor - 1),0)$, $(b_{np}(\lfloor m/b_{np} \rfloor),0)$ or $(0, b_p)$, $(b_{np},b_p),\ldots$, $(b_{np}(\lfloor m/b_{np} \rfloor - 1),b_p)$, $(b_{np}(\lfloor m/b_{np} \rfloor),b_p)$ together to obtain supernodes as shown in Fig. 3.1, we can frame a one dimensional Markov chain model that we can solve to obtain the steady state probabilities of each of these supernodes. The same approach can be utilized to group nodes that have the same number of non-prioritized calls such as (0,0), $(0, b_p),\ldots,(0, b_p\cdot(\lfloor C/b_p \rfloor))$ or

(b_{np},0), (b_{np}, b_p),…, (b_{np}, $b_p\cdot(\lfloor (C-b_{np})/b_p \rfloor)$) together. Both one dimensional Markov chain models obtained above are given in Figs. 3.1 and 3.2 for prioritized and non-prioritized calls, respectively. In Fig. 3.1, we observe that for all supernodes except the ones that have at least one member node that represents a system state in which the total number of occupied channels is equal to the total number of channels in the system, C, there exist (m+1) pairs of transitional flows between their member nodes and the corresponding member nodes that belong to their neighboring supernodes. Conversely, for the rest of the supernodes there exist some member nodes that do not have transitional flows in between any of the corresponding nodes that belong to their neighboring supernodes. Same can also be observed for the supernodes shown in Fig. 3.2; however in addition to those mentioned above there exist some other member nodes with unidirectional transition flows.

In Fig. 3.3, we show the one dimensional Markov chain model for prioritized calls where each node represents a supernode composed of a set of nodes shown in Fig. 3.1. We determine the values of the admission probabilities for prioritized calls, k_j, by obtaining the ratio of the sum of occupancy probabilities of the feasible member nodes of a supernode, for which the system admits an arriving prioritized call, to the sum of occupancy probabilities of all feasible member nodes of that particular supernode. Thus, when $j = 0, 1,\ldots (\lfloor (C-(\lfloor m/b_{np}\rfloor \cdot b_{np})/b_p \rfloor -1)$, the admission probabilities for prioritized calls, k_j, are equal to 1. The equilibrium channel occupancy probability when exactly j prioritized calls exist, where $q_p(j), j = 0, 1,\ldots \lfloor C/b_p \rfloor$, can be obtained from the following recursive equation.

$$(\rho_p \cdot k_{j-1}) \cdot q_p(j-1) = j \cdot q_p(j),\ j = 1,\ldots, \lfloor C/b_p \rfloor \qquad (3)$$

Solving for $q_p(0)$ in the equation $\sum_{j=0}^{\lfloor C/b_p \rfloor} q_p(j) = 1$, we obtain

$$q_p(j) = \frac{\prod_{z=0}^{j-1} (\rho_p \cdot k_z)}{j!} \cdot q_p(0), 1 \le j \le \lfloor C/b_p \rfloor \qquad (4)$$

where

$$q_p(0) = \left[1 + \sum_{j=1}^{\lfloor C/b_p \rfloor} \frac{\prod_{n=0}^{j-1} (\rho_p \cdot k_n)}{j!} \right]^{-1} \tag{5}$$

Let h_r, where (r = 0, 1,... $(\lfloor m/b_{np} \rfloor - 1)$), denote the admission probability of an arriving non-prioritized call when r non-prioritized calls exist, regardless of the number of existing prioritized calls. However, h_r shall not be confused with β_i since the latter is a predefined user controlled parameter that indicates whether an arriving non-prioritized call will be admitted or not depending on the number of occupied channels. Similar to, yet slightly different than k_j, we determine the values of h_r by obtaining the ratio of the sum of occupancy probabilities of the feasible member nodes of a supernode, for which an arriving non-prioritized call is admitted, multiplied with β_i to the sum of occupancy probabilities of all the feasible member nodes of that particular supernode.

In Fig. 3.4, we show the one dimensional Markov chain model for non-prioritized calls where each node represents a supernode composed of a set of nodes shown in Fig. 3.2. The equilibrium channel occupancy probabilities, $q_{np}(r)$, could be obtained similarly to prioritized calls if unidirectional transition flows, shown in Fig. 3.2, did not exist. However their existence needs to be taken into account by adjusting μ_{np} affiliated with each supernode appropriately. Therefore we initiate $\mu'_{np}(r)$ to replace μ_{np} affiliated with each supernode in the model given in Fig. 3.4 and determine its value by dividing the number of transition flows departing from the associated supernode with the number of pairs of bidirectional transition flows in between the same particular supernodes.

$$\mu'_{np}(r) = \frac{\lfloor (C-r)/b_p \rfloor + 1}{\lfloor (m-(r-1))/b_p \rfloor} \cdot \mu_{np}(r), r = 1,..., \lfloor m/b_p \rfloor \tag{6}$$

Then we can obtain the occupancy probabilities $q_{np}(r)$, r = 0,1,... $(\lfloor m/b_{np} \rfloor - 1)$, which satisfy the following recursive equation.

$$(\lambda_{np} \cdot h_{r-1}) \cdot q_{np}(r-1) = r \cdot \mu'_{np}(r) \cdot q_{np}(r), r = 1,..., \lfloor m/b_{np} \rfloor \tag{7}$$

Solving for $q_{np}(0)$ in the equation $\sum_{r=0}^{\lfloor m/b_{np} \rfloor} q_{np}(r) = 1$, we obtain

$$q_{np}(r) = \frac{\prod_{z=0}^{r-1}\left(\frac{(\lambda_{np} \cdot h_z)}{\mu'_{np}(r)}\right)}{r!} \cdot q_{np}(0), 1 \le r \le \lfloor m/b_{np} \rfloor \tag{8}$$

where

$$q_{np}(0) = \left[1 + \sum_{r=1}^{\lfloor m/b_{np} \rfloor} \frac{\prod_{n=0}^{r-1}\left(\frac{\lambda_{np} \cdot h_n}{\mu'_{np}(r)}\right)}{r!} \right]^{-1} \tag{9}$$

The admission probabilities for prioritized, k_j, and non-prioritized calls, h_r, cannot be obtained without computing the occupancy probability of each feasible node. Even if the occupancy probabilities of the supernodes for prioritized and non-prioritized calls can be obtained using this method, we still need to compute the occupancy probabilities of certain feasible nodes since joint occupancy probabilities of these supernodes cannot be used due to their dependencies.

To overcome these difficulties, we suggest the following iterative approach:

1. Initialize the value of estimated equilibrium occupancy probabilities ($\hat{q}(n_{np}, n_p)$ for $n_{np} = 0,1 \ldots \lfloor m/b_{np} \rfloor$ and $n_p = 0,1 \ldots \lfloor C/b_p \rfloor$) by setting them equal to 1 / (total number of feasible nodes).

2. Calculate $\mu'_{np}(r)$ for $r = 1, \ldots \lfloor m/b_{np} \rfloor$ using (6).

3. Iterate with the following steps until the changes in the updated values of k_j and h_r are not less than a chosen resolution.

3.1. Calculate and update k_j for $j = 0,1, \ldots (\lfloor C/b_p \rfloor - 1)$ and $q_p(j)$ for $j = 0,1, \ldots \lfloor C/b_p \rfloor$ using (4) and (5).

3.2 Update the values of estimated occupancy probabilities, $\hat{q}(n_{np}, n_p)$, by apportioning the value of the last updated occupancy probability, $q_p(j)$, of the corresponding supernode for prioritized calls amongst its nodes with respect to the value of last updated occupancy probability, $q_{np}(r)$, of the corresponding supernode for non-prioritized calls.

3.3. Calculate and update h_r for $r = 0,1, \ldots (\lfloor m/b_{np} \rfloor - 1)$ and $q_{np}(r)$ for $r = 0,1, \ldots \lfloor m/b_{np} \rfloor$ using (8) and (9).

3.4. Update the values of estimated occupancy probabilities, $\hat{q}(n_{np}, n_p)$, by apportioning the value of the last updated occupancy probability, $q_{np}(r)$, of the corresponding supernode for non-prioritized calls amongst its nodes with respect to the value of last updated occupancy probability, $q_p(j)$, of the corresponding supernode for prioritized calls.

4. Obtain call blocking probabilities for prioritized and non-prioritized calls using $\hat{q}(n_{np}, n_p)$.

The call blocking probabilities for both types of calls are calculated as follows when the final estimated values of equilibrium occupancy probabilities, $\hat{q}(n_{np}, n_p)$, are obtained.

$$B_p = \sum_{n=0}^{\lfloor m/b_{np} \rfloor} \hat{q}\left(n, \lfloor (C - (n \cdot b_{np}))/b_p \rfloor\right) \tag{10}$$

$$B_{np} = \sum_{a=0}^{\lfloor m/b_{np} \rfloor} \sum_{n=0}^{\lfloor (C-a.b_{np}/b_p \rfloor} \hat{q}(a,n) \tag{11}$$

$$(a \cdot b_{np} + n \cdot b_p) \geq \lfloor m/b_{np} \rfloor$$

Despite its iterative nature, we expect the *state space decomposition* method to have low computational cost since decomposing the whole state space into subspaces and forming supernodes to apply one dimensional Markov chain modeling utilize the closed form formulas obtained from one dimensional Markov chain models and make the proposed method easy to implement for real time applications.

3.4 Numerical Results

In this section we compare the performance of the proposed method, *state space decomposition*, with Borst and Mitra's approximation [26] and the *direct* (also known as *LU decomposition*) numerical method. We show that the runtime computational cost of the proposed approximation method is negligible compared to the existing numerical methods' (i.e., direct, method of Jacobi, method of Gauss-Seidel) with respect to *CPU time* and *memory* needed to obtain the results. We investigate the cutoff priority scheme, which is a special case of fractional guard channel scheme, using the following set of parameters: $C = 32$, $m = 24$ ($\beta_i = 1$ for $i = 0, \ldots 23$, 0 otherwise), $\lambda_{np} = 0.1$, $1/\mu_{np} = 200$, $1/\mu_p = 50$ and λ_p is varied from 1 to 0.05. However any fractional guard channel scheme can be chosen since other choices of β_i's would give similar results. We set the capacity requirement for non-prioritized calls, b_{np}, to 1 and vary the capacity requirement for prioritized calls, b_p, by setting its value to 1, 2 and 4. Figs. 3.5 to 3.10 depict the prioritized and non-prioritized call blocking probabilities obtained using the *direct*, Borst and Mitra's and the proposed methods under varying prioritized call traffic load, respectively. We observe for all values of b_p that when prioritized call traffic load is higher than non-prioritized call traffic load (i.e., $\rho_p > \rho_{np}$), both call blocking probabilities approximated by the proposed method match the exact results obtained by the *direct* numerical method very well. However Borst and Mitra's approximation method overestimates the prioritized call blocking probability generously while it underestimates the non-prioritized call blocking probability extensively when $\rho_p > \rho_{np}$. When the prioritized call traffic load is lower than the non-prioritized call traffic load (i.e., $\rho_p < \rho_{np}$), the *state space decomposition* method slightly overestimates the prioritized call blocking probability while it slightly underestimates the non-prioritized call blocking probability with the discrepancy increasing as both traffic loads are decreasing. Yet, Borst and Mitra's method gives a better approximation only when both traffic loads are very low due to its assumption on independent channel occupancy.

The discrepancy observed when non-prioritized call traffic load takes over prioritized call traffic load (i.e., $\rho_p < \rho_{np}$) is due to an assumption that we made in the iterative solution described above, i.e., the steady state probabilities of all nodes that are members of the same particular supernode for prioritized calls are proportional to each other with the same ratio that exists between the steady state probabilities of the corresponding supernodes for non-prioritized calls, and vice versa. Therefore we expect proposed method to approximate steady state probabilities of nodes that are members of supernodes which have relatively less number of member nodes better with respect to the others that are members of supernodes which have relatively more number of member nodes. However this is not a significant problem unless $\rho_{np} >> \rho_p$, since call blockings mostly occur at nodes that are members of supernodes which have relatively less number of member nodes and thus closer to the edges of transition diagrams. When non-prioritized call traffic load takes over the prioritized call traffic load, the steady state probabilities of the nodes that have relatively more number of non-prioritized calls dominates the others needed to compute a particular call blocking probability and thus lead to discrepancy. On the other hand, Borst and Mitra's method approximates better only when both call traffic loads are very low due to the channel occupancy independence assumption that the authors made [26]. When each of the individual classes accounts for a substantial portion of the total amount of capacity in use, it leads to mutual dependence as the traffic load increases. Both approximation methods perform relatively better when the capacity requirement for prioritized calls, b_p, increases.

When we increase the number of shared channels, m, to 28 and keep all the parameters given above same, less number of non-prioritized calls are blocked. We observe that call blocking probabilities for both types of calls are estimated more accurately since scheme's transition diagram has more supernodes for non-prioritized calls that have relatively less number of member nodes. We make another comparison in Figs. 3.11 to 3.16 by using the following set of parameters: C = 32, m = 28 (β_i = 1 for i = 0, ... 27, 0 otherwise), λ_{np} = 0.1, $1/\mu_{np}$ = 200, $1/\mu_p$ = 50 and λ_p is varied from 1 to 0.05. We set the capacity requirement for prioritized calls, b_p, to 1 and vary the capacity requirement for non-prioritized calls, b_{np}, by setting its value to 1, 2 and 4. The results are similar to the first case, however when the capacity requirement for non-prioritized calls, b_{np}, increases, the discrepancy observed in non-prioritized call blocking probabilities obtained from the proposed method increases while it decreases for the ones obtained from Borst and Mitra's method. Yet, Borst and Mitra's method still approximates better than the proposed method only when both traffic loads are very low.

In [20] we proposed a closed form approximation method, *effective duration time*, to evaluate call blocking performance in cellular networks under homogeneous traffic. The method provides accurate results, but the results are sensitive to the average values of channel occupancy times. We use the following set of parameters to compare the performance of this method to the proposed *state space decomposition* method's to observe if it is more sensitive than the previously proposed method under homogeneous traffic: C = 32, m = 28, λ_{np} = 0.1, $1/\mu_{np}$ = 200, $1/\mu_p$ = 50, b_p = b_{np} = 1 and λ_p is varied from 1 to 0.05. Figs. 3.17 and 3.18 depict the prioritized and non-prioritized call blocking probabilities, respectively, under different prioritized call traffic loads. When call traffic load is varied by changing the value of call arrival rates we observe that both methods slightly overestimate the prioritized call blocking probability when $\rho_p < \rho_{np}$ whereas the results obtained from both approximation methods match the results obtained from the *direct* numerical method very well when $\rho_p > \rho_{np}$. The *state space decomposition* method underestimates the non-prioritized call blocking probability while the *effective duration time* method provides results that match well with the exact solutions.

On the other hand, when we keep the set of parameters given above the same but set λ_p to 0.5 and vary the call traffic load by changing the value of average channel occupancy times, $1/\mu_p$, from 5 to 100, we observe that the results obtained from both approximation methods matched the ones obtained by the *direct* numerical method very well when $\rho_p > \rho_{np}$ whereas the effective duration time method degenerates slightly compared to the previous results given above when $\rho_p < \rho_{np}$. Figs. 3.19 and 3.20 show the call blocking probabilities for prioritized and non-prioritized calls, respectively. We observe that the *state space decomposition* method is indifferent to changes in average channel occupancy times as opposed to the *effective duration time* method.

Computationally efficient approximation methods for evaluating call admission control schemes are studied to replace methods such as the *direct* which provides exact solutions by solving large sets of flow equations. Product form solutions are preferable due to their computational efficiency; however it is very difficult to find one to evaluate asymmetric call admission control schemes. Considering that *state space decomposition* method is iterative, we need to compare it with the *direct* and other widely used iterative methods such as the method of *Jacobi*, method of *Gauss-Seidel* and method of *Borst-Mitra* with respect to runtime computational cost to emphasize its benefits. We choose the parameters "*CPU time*" and "*used memory*" to compare the computational efficiencies of the numerical and the approximation methods. We define "*CPU time*" as the total processing time (in seconds) a CPU spends from the time that the computation is started for each method and the "*used memory*" as the amount of storage allocated for nonzero matrix elements. We use the following set of parameters to obtain the numerical results presented in Table 3.1 by evaluating an asymmetric CAC scheme, cutoff priority, for three different scenarios: $C = 6$, $m = 5$, $\lambda_{np} = 0.02$, $C = 32$, $m = 28$, $\lambda_{np} = 0.1$, $C = 64$, $m = 56$, $\lambda_{np} = 0.5$, while $1/\mu_{np} = 200$, $1/\mu_p = 50$, $b_p = b_{np} = 1$ and λ_p varies from 0.2 to 0.01, 1 to 0.05 and 5 to 0.25, respectively. The parameters are chosen to obtain similar values of call blocking probabilities for both classes of calls in all three scenarios to make the comparisons appropriate.

The simulation results given in Table 3.1 show that the *CPU time*s and the *used memory* obtained from our approximation method is almost negligible when compared with the ones obtained from the *direct* numerical solution, method of *Jacobi* and method of *Gauss-Seidel* especially when the number of channels in the system increases. This is not surprising even though the proposed approximation method is iterative, since one dimensional Markov chain models with closed form formula solutions for each call type are utilized to estimate the steady state probabilities. Nevertheless the values of the admission probabilities for the prioritized, k_j, and the non-prioritized, h_r, calls converge fast. Both the method of *state space decomposition* and *Borst-Mitra* perform similar to a closed form formula solution with respect to *CPU time*s and *used memory* given in Table 3.1. The latter computes the solution faster using less memory compared to the former, however the proposed method approximates more accurately compared to the latter despite the insignificant rise in computational cost.

3.5 Summary

In this chapter we have classified call admission control schemes into two categories; *symmetric* and *asymmetric*. We obtained a product form exact solution formula to evaluate symmetric call admission control schemes in multi-service networks in section 3.2 and proposed a novel computationally efficient approximation method that uses an iterative approach to evaluate the call blocking performance of asymmetric call admission control schemes in multi-service networks in section 3.3.

We compared the numerical results obtained from the proposed approximation method, *state space decomposition*, with the ones obtained from a previously proposed approximation method by Borst and Mitra and the numerical method, *direct*, which provides the exact solution. We showed that the total processing time a CPU spends to compute the solution and the amount of storage allocated during this computation are almost negligible when the proposed approximation method is compared with the existing numerical methods' such as *direct*, method of *Jacobi* and method of *Gauss-Seidel*. Numerical results showed that the proposed method provides results that match better with the exact solutions compared to the ones provided by the method of *Borst-Mitra* while keeping the computational cost low. Decomposing the whole state space into subspaces and forming supernodes to apply one dimensional Markov chain modeling to use its closed form formulas iteratively make our proposed approximation method have comparatively low CPU times and memory usage with respect to those obtained from single closed form formula solutions.

Many dynamic call admission control schemes have been proposed to maximize the network utilization. These dynamic schemes help networks to accept more calls with low priorities by adaptively reserving the amount of resources needed for calls with high priorities. However the accuracy of adaptive reservation depends on the quality and up-to-dateness of the feedback received during real time applications. Thus the challenge is to provide this feedback to the call admission control mechanism in real time. Considering the high computational cost of the existing numerical solution methods, finding performance evaluation approximation methods with low computational cost is inevitable if these dynamic call admission control schemes are going to be implemented in cellular networks. We believe that providing easy to implement performance evaluation approximation methods with low computational cost will help motivate the practical implementation of dynamic call admission control schemes.

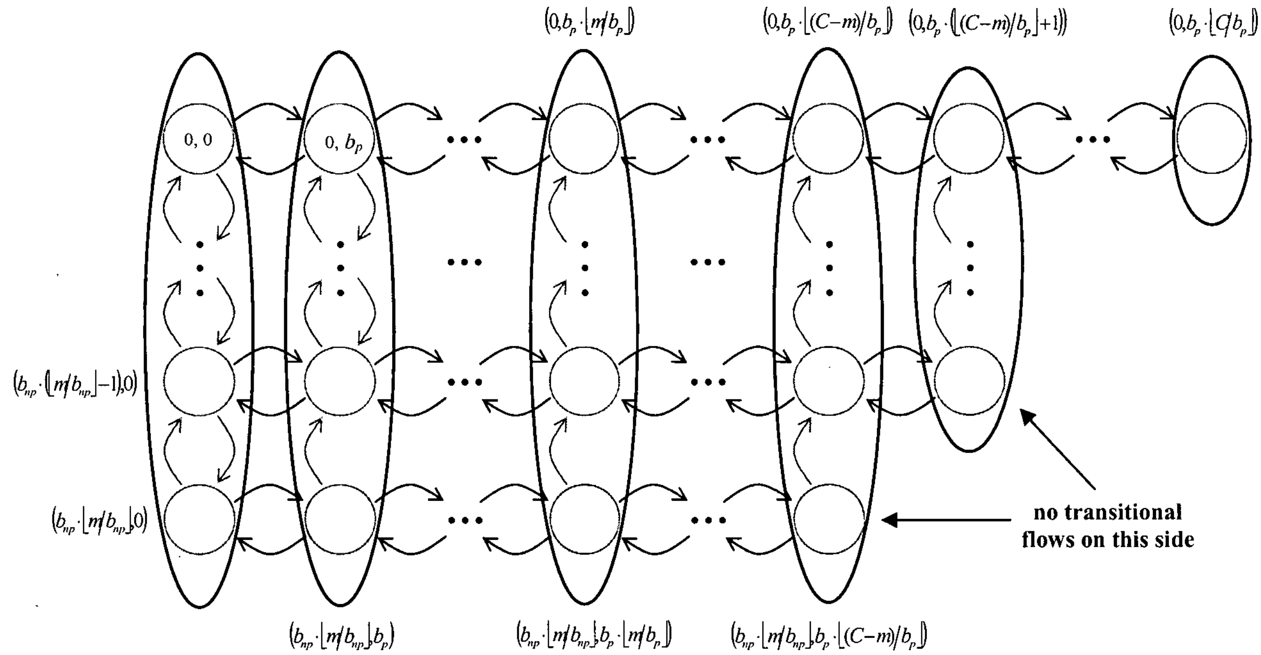
$(0,b_p\cdot\lfloor m/b_p\rfloor)$
$(0,b_p\cdot\lfloor (C-m)/b_p\rfloor)$
$(0,b_p\cdot(\lfloor (C-m)/b_p\rfloor+1))$
$(0,b_p\cdot\lfloor C/b_p\rfloor)$
0, 0
0, b_p
$(b_{np}\cdot(\lfloor m/b_{np}\rfloor-1),0)$
$(b_{np}\cdot\lfloor m/b_{np}\rfloor,0)$
no transitional flows on this side
$(b_{np}\cdot\lfloor m/b_{np}\rfloor,b_p)$
$(b_{np}\cdot\lfloor m/b_{np}\rfloor,b_p\cdot\lfloor m/b_p\rfloor)$
$(b_{np}\cdot\lfloor m/b_{np}\rfloor,b_p\cdot\lfloor (C-m)/b_p\rfloor)$

Fig. 3.1 Transition diagram for asymmetric call admission control schemes with supernodes for prioritized calls.

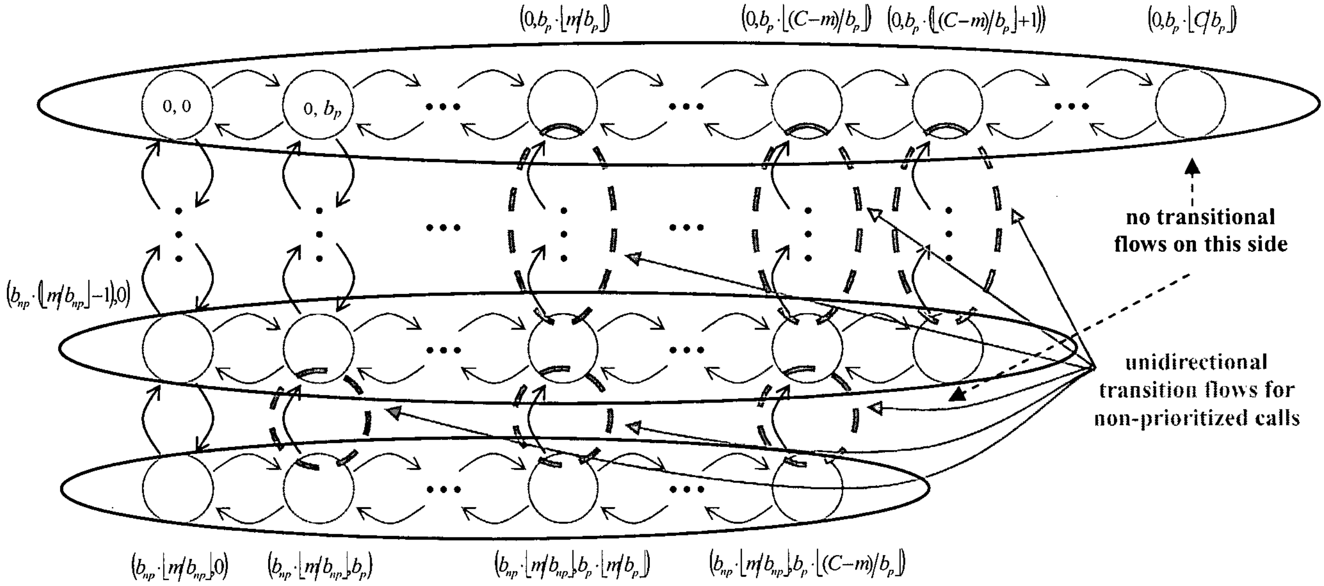
$(0,b_p\cdot\lfloor m/b_p\rfloor)$
$(0,b_p\cdot\lfloor (C-m)/b_p\rfloor)$
$(0,b_p\cdot(\lfloor (C-m)/b_p\rfloor+1))$
$(0,b_p\cdot\lfloor C/b_p\rfloor)$
0, 0
0, b_p
no transitional flows on this side
$(b_{np}\cdot(\lfloor m/b_{np}\rfloor-1),0)$
unidirectional transition flows for non-prioritized calls
$(b_{np}\cdot\lfloor m/b_{np}\rfloor,0)$
$(b_{np}\cdot\lfloor m/b_{np}\rfloor,b_p)$
$(b_{np}\cdot\lfloor m/b_{np}\rfloor,b_p\cdot\lfloor m/b_p\rfloor)$
$(b_{np}\cdot\lfloor m/b_{np}\rfloor,b_p\cdot\lfloor (C-m)/b_p\rfloor)$

Fig. 3.2 Transition diagram for asymmetric call admission control schemes with supernodes for non-prioritized calls.

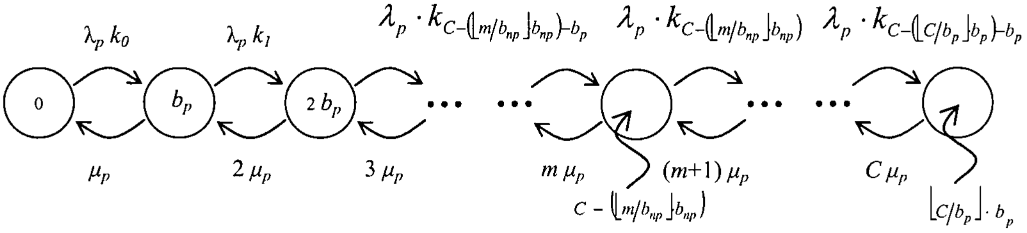
$\lambda_p k_0$
$\lambda_p k_1$
$\lambda_p \cdot k_{C-(\lfloor m/b_{np} \rfloor b_{np})-b_p}$
$\lambda_p \cdot k_{C-(\lfloor m/b_{np} \rfloor b_{np})}$
$\lambda_p \cdot k_{C-(\lfloor C/b_p \rfloor b_p)-b_p}$
0
b_p
$2\,b_p$
μ_p
$2\,\mu_p$
$3\,\mu_p$
$m\,\mu_p$
$(m+1)\,\mu_p$
$C\,\mu_p$
$C-(\lfloor m/b_{np} \rfloor b_{np})$
$\lfloor C/b_p \rfloor \cdot b_p$

Fig. 3.3 One dimensional Markov chain model obtained for the prioritized calls' supernodes.

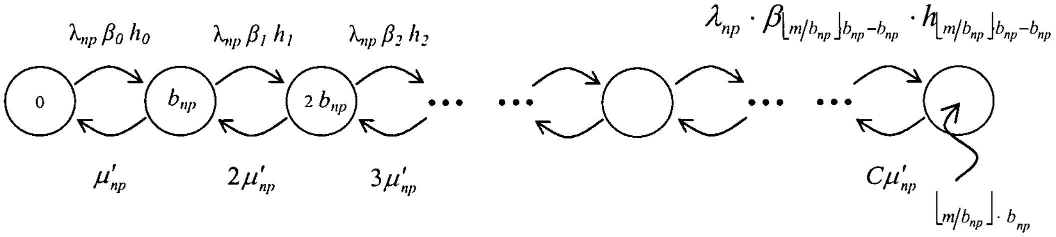
$\lambda_{np}\,\beta_0\,h_0$
$\lambda_{np}\,\beta_1\,h_1$
$\lambda_{np}\,\beta_2\,h_2$
$\lambda_{np} \cdot \beta_{\lfloor m/b_{np} \rfloor b_{np}-b_{np}} \cdot h_{\lfloor m/b_{np} \rfloor b_{np}-b_{np}}$
0
b_{np}
$2\,b_{np}$
μ'_{np}
$2\mu'_{np}$
$3\mu'_{np}$
$C\mu'_{np}$
$\lfloor m/b_{np} \rfloor \cdot b_{np}$

Fig. 3.4 One dimensional Markov chain model obtained for the non-prioritized calls' supernodes.

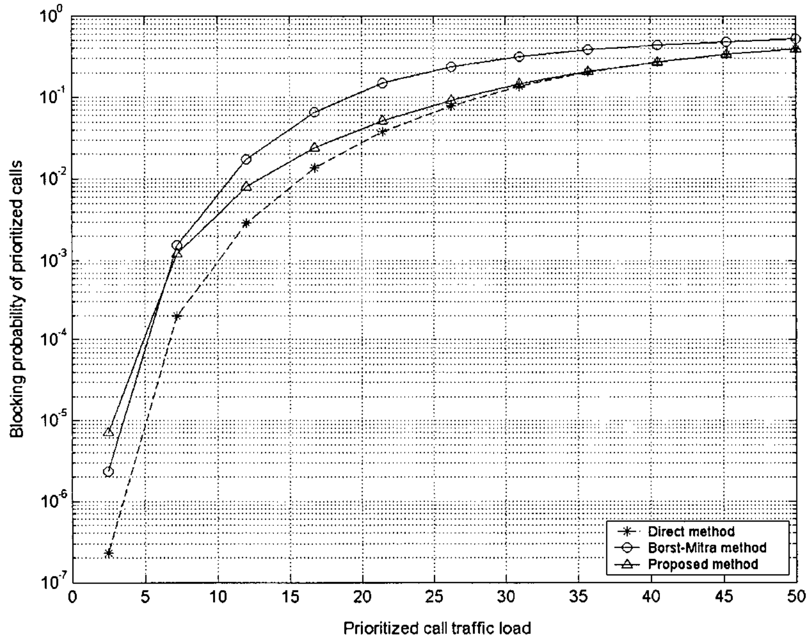

Fig. 3.5 Prioritized call blocking probability for the *cutoff priority* scheme $b_p = 1$, $b_{np} = 1$, and $m = 24$.

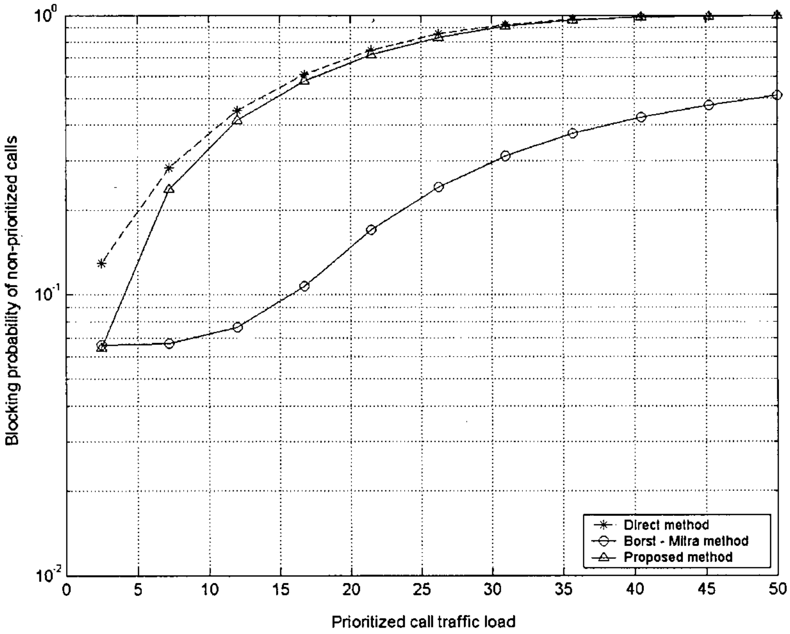

Fig. 3.6 Non-prioritized call blocking probability for the *cutoff priority* scheme $b_p = 1$, $b_{np} = 1$, and $m = 24$.

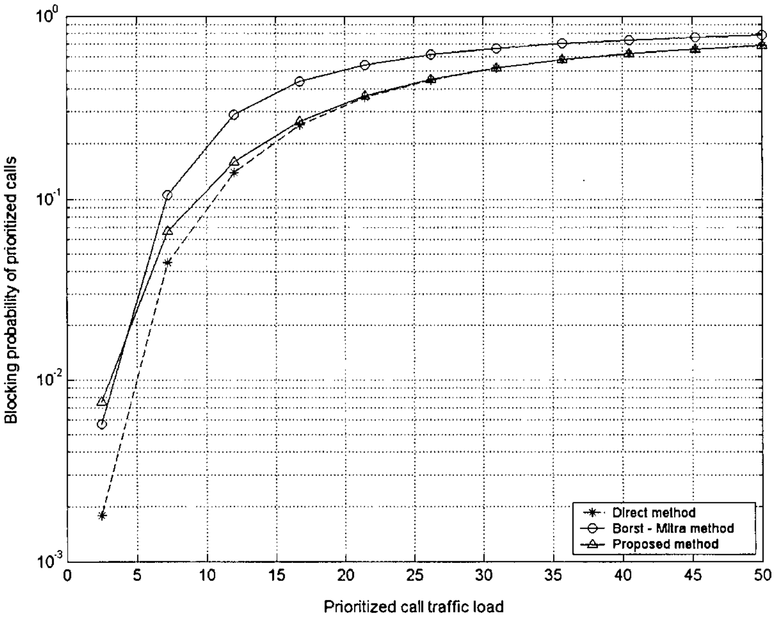

Fig. 3.7 Prioritized call blocking probability for the *cutoff priority* scheme $b_p = 2$, $b_{np} = 1$, and $m = 24$.

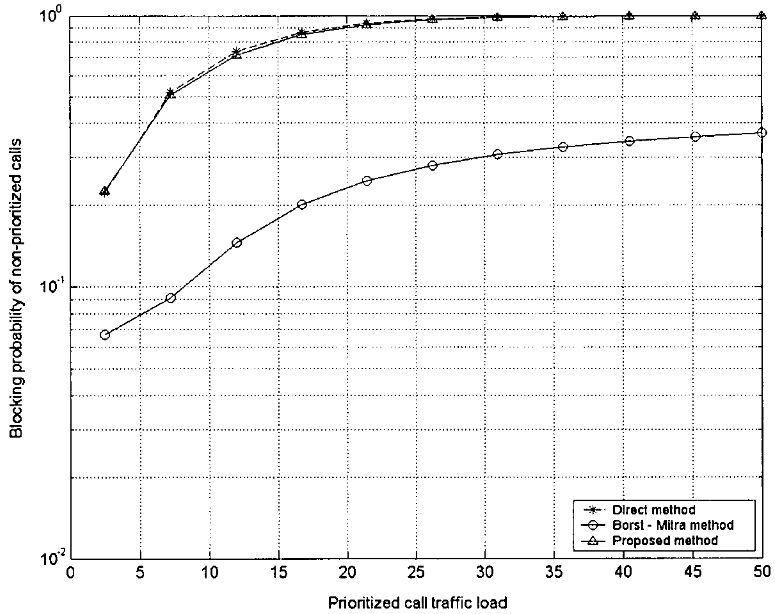

Fig. 3.8 Non-prioritized call blocking probability for the *cutoff priority* scheme $b_p = 2$, $b_{np} = 1$, and $m = 24$.

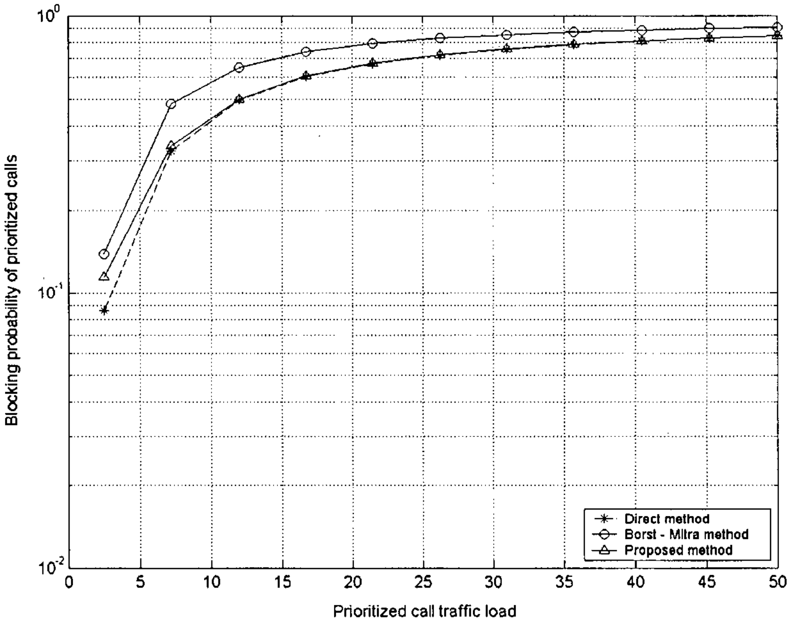

Fig. 3.9 Prioritized call blocking probability for the *cutoff priority* scheme $b_p = 4$, $b_{np} = 1$, and $m = 24$.

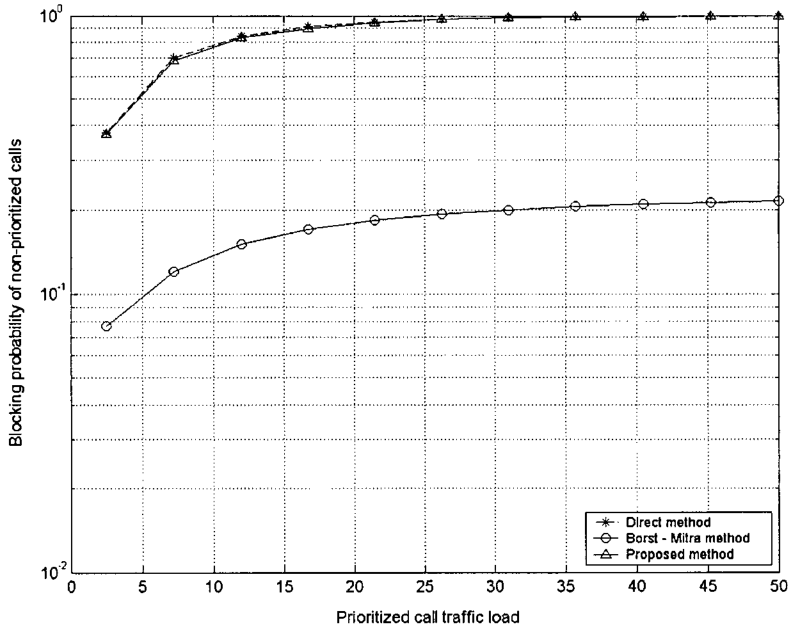

Fig. 3.10 Non-prioritized call blocking probability for the *cutoff priority* scheme $b_p = 4$, $b_{np} = 1$, and $m = 24$.

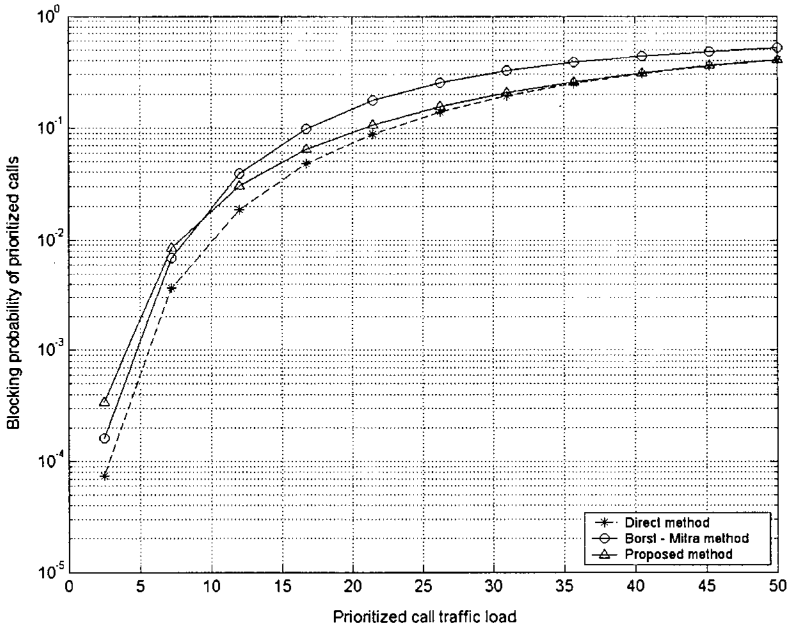

Fig. 3.11 Prioritized call blocking probability for the *cutoff priority* scheme $b_p = 1$, $b_{np} = 1$, and $m = 28$.

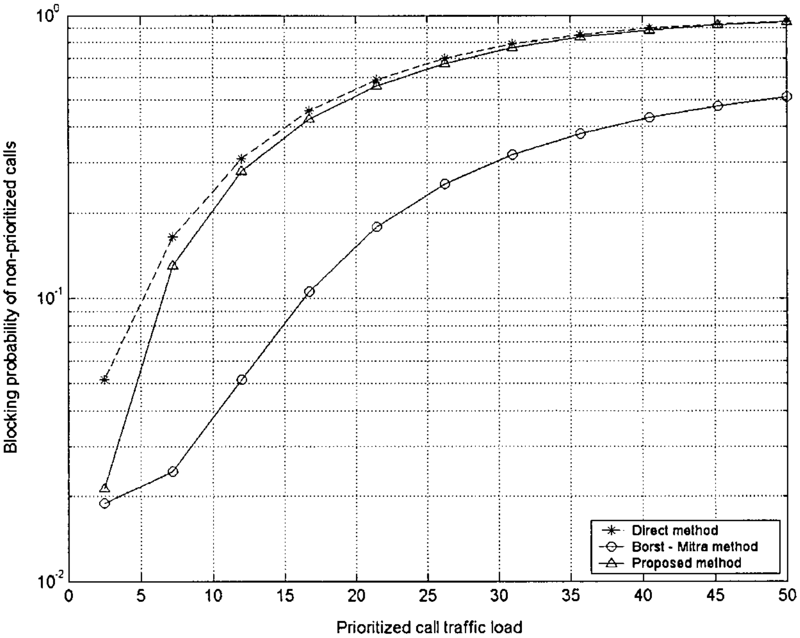

Fig. 3.12 Non-prioritized call blocking probability for the *cutoff priority* scheme $b_p = 1$, $b_{np} = 1$, and $m = 28$.

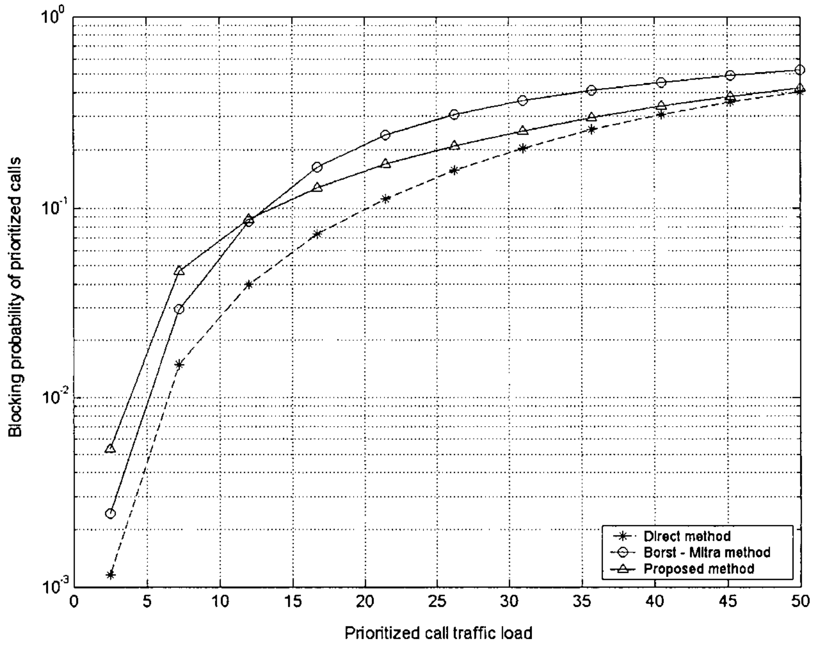

Fig. 3.13 Prioritized call blocking probability for the *cutoff priority* scheme $b_p = 1$, $b_{np} = 2$, and $m = 28$.

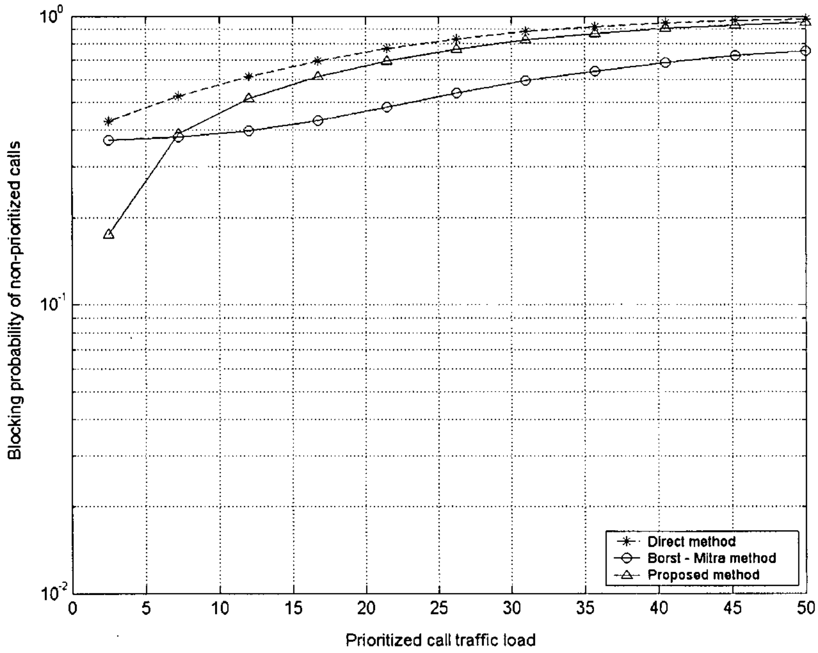

Fig. 3.14 Non-prioritized call blocking probability for the *cutoff priority* scheme $b_p = 1$, $b_{np} = 2$, and $m = 28$.

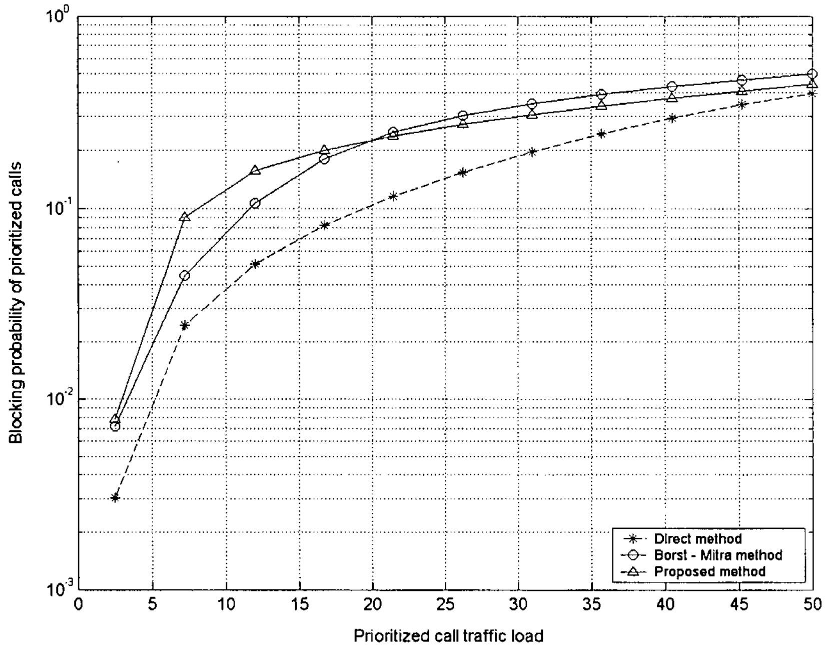

Fig. 3.15 Prioritized call blocking probability for the *cutoff priority* scheme $b_p = 1$, $b_{np} = 4$, and $m = 28$.

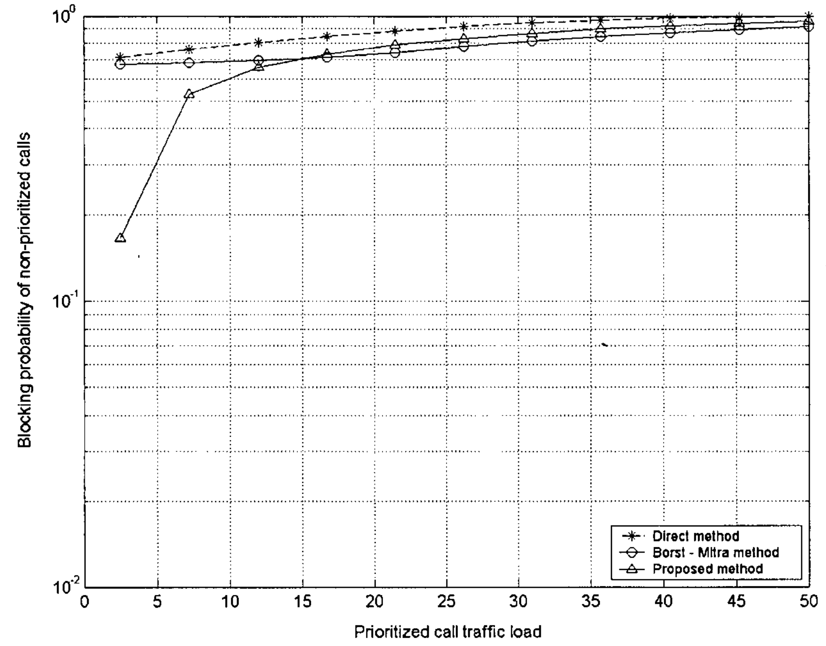

Fig. 3.16 Non-prioritized call blocking probability for the *cutoff priority* scheme $b_p = 1$, $b_{np} = 4$, and $m = 28$.

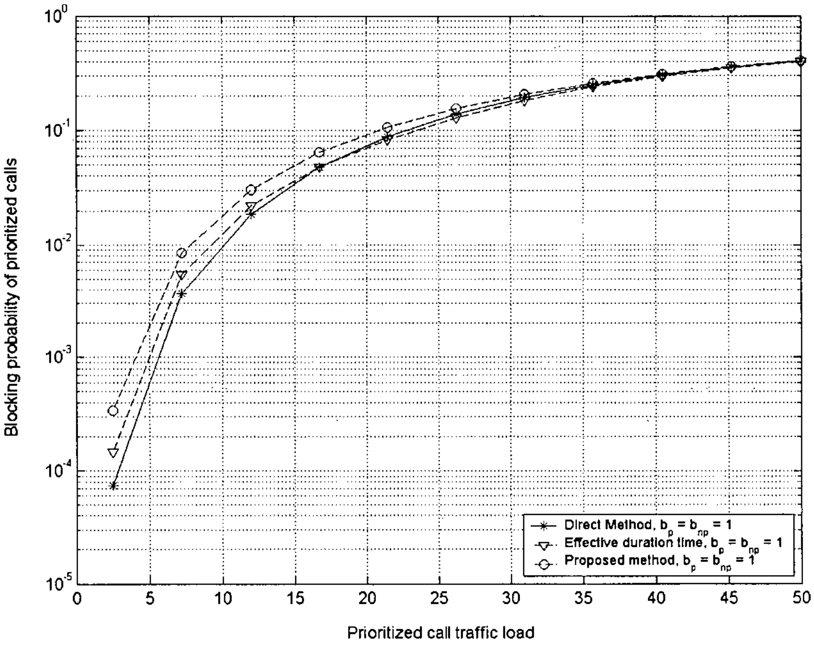

Fig. 3.17 Prioritized call blocking probability for the *cutoff priority* scheme $b_p = b_{np} = 1, \rho_{np} = 20, \rho_p = 2.5$ to $50, \lambda_p = 0.05$ to 1.

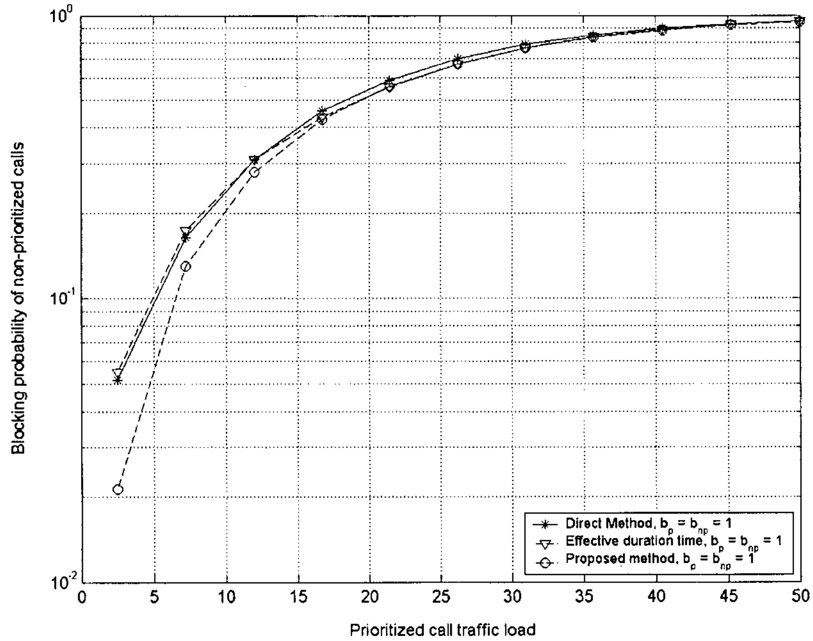

Fig. 3.18 Non-prioritized call blocking probability for the *cutoff priority* scheme $b_p = b_{np} = 1, \rho_{np} = 20, \rho_p = 2.5$ to $50, \lambda_p = 0.05$ to 1.

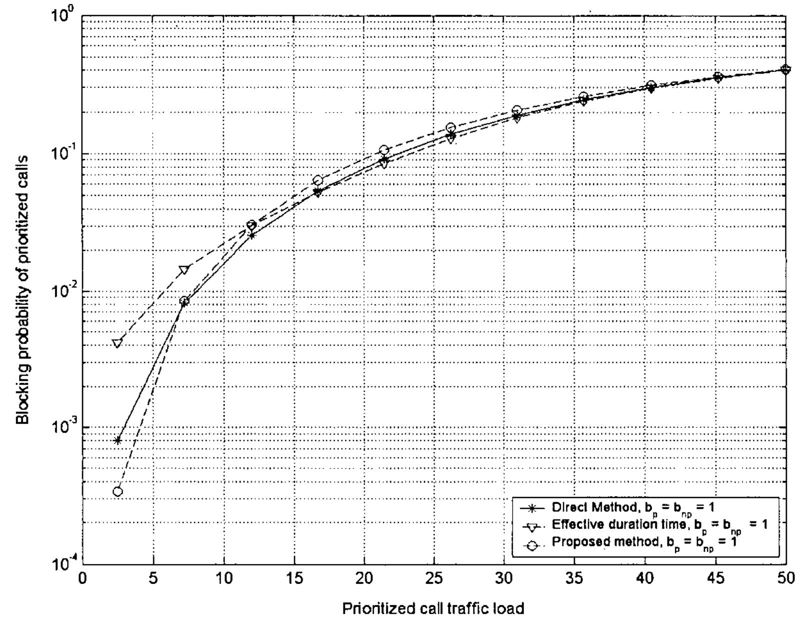

Fig. 3.19 Prioritized call blocking probability for the *cutoff priority* scheme $b_p = b_{np} = 1$, $\rho_{np} = 20$, $\rho_p = 2.5$ to 50, $1/\mu_p = 5$ to 100.

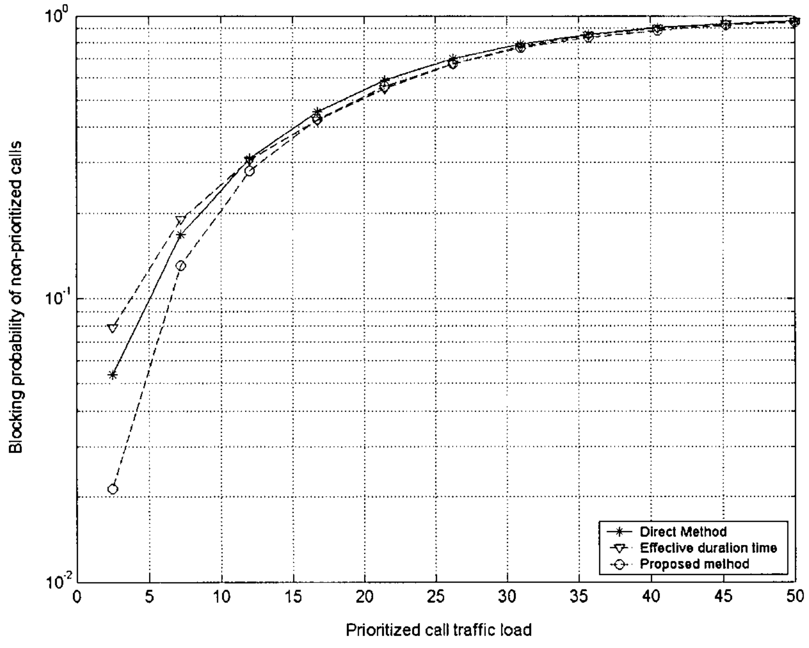

Fig. 3.20 Non-prioritized call blocking probability for the *cutoff priority* scheme $b_p = b_{np} = 1$, $\rho_{np} = 20$, $\rho_p = 2.5$ to 50, $1/\mu_p = 5$ to 100.

TABLE 3.1

Comparison of runtime computational costs between the proposed approximation method and the *Borst-Mitra*, *direct* and *iterative* methods

	Total number of channels = C, total number of shared channels = m					
	$C = 6$, $m = 5$		$C = 32$, $m = 28$		$C = 64$, $m = 56$	
Numerical Methods	CPU time	used memory	CPU time	used memory	CPU time	used memory
Direct Method	0.07s	3.82e+03	30.74s	152e+04	49m8s	222.52e+05
Method of Jacobi	0.25s	2.46e+03	29.96s	91.6e+04	11m24s	133.64e+05
Method of Gauss-Seidel	0.18s	2.46e+03	1m46s	91.6e+04	1h33m24s	133.64e+05
Borst – Mitra Method	0.003s	0.12e+03	0.02s	0.05e+04	0.033s	0.011e+05
Proposed method	0.006s	0.13e+03	0.04s	0.14e+04	0.45s	0.046e+05

3.6 Bibliography

[1] L. Huang, S. Kumar, and C. C. J. Kuo, "Adaptive resource allocation for multimedia services in wireless communication networks," *21^{st} International Conference on Distributed Computing Systems Workshop (ICDCSW '01),* pp. 307-312, 2001.

[2] D. Hong and S. S. Rappaport, "Traffic model and performance analysis for cellular mobile radiotelephone systems with prioritized and non-prioritized handoff procedures," *IEEE Transactions on Vehicular Technology*, vol. 35, pp. 77-92, Aug. 1986.

[3] B. Li, C. Lin, and S. T. Chanson, "Analysis of a hybrid cutoff priority scheme for multiple classes of traffic in multimedia wireless networks," *Wireless Networks*, vol. 4, no. 4, pp. 279-290, July 1998.

[4] Y.B. Lin, S. Mohan, and A. Noerpel, "Queuing priority channel assignment strategies for handoff and initial access for a PCS network," *IEEE Transactions on Vehicular Technology*, vol. 43, no. 3, pp. 704-712, Aug. 1994.

[5] J. Y. Lee, and S. Bahk, "Simple admission control schemes supporting QoS in wireless multimedia networks," *IEE Electronics Letters*, vol. 37, no. 11, pp. 712-713, May 2001.

[6] R. Ramjee, R. Nagarajan, and D. Towsley, "On optimal call admission control in cellular networks," *Wireless Networks*, vol. 3, no. 1, pp. 29-41, March 1997.

[7] Y. Fang, and Y. Zhang, "Call admission control schemes and performance analysis in wireless mobile networks," *IEEE Transactions on Vehicular Technology*, vol. 51, no.2, pp. 371-382, March 2002.

[8] M. D. Kulavaratharasah, and A. H. Aghvami, "Teletraffic performance evaluation of microcell personal communication networks (PCNs) with prioritized handoff procedures," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 1, pp. 137-152, Jan. 1999.

[9] A. S. Acampora and M. Naghshineh, "Control and quality of service provisioning in high-speed micro-cellular networks," *IEEE Personal Communications*, vol. 1, no. 2, pp. 36-43, 1996.

[10] M. Naghshineh and S. Schwartz, "Distributed call admission control in mobile/wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 14, no. 4, pp. 711-717, May 1996.

[11] D. Levine, I. Akyildiz, and M. Naghshineh, "A resource estimation and call admission algorithm for wireless multimedia networks using the shadow cluster concept," *IEEE/ACM Transactions on Networking*, vol. 5, no. 1, pp. 1-12, Feb. 1997.

[12] A. Sutivong and J. M. Peha, "Novel heuristics for call admission control in cellular systems," *1997 IEEE 6th International Conference on Universal Personal Communications Record*, vol. 1, pp. 129-133, Oct. 1997.

[13] C. Oliveira, J. B. Kim, and T. Suda, "An adaptive bandwidth reservation scheme for high-speed multimedia wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 858-874, Aug. 1998.

[14] P. Ramanathan, K. M. Sivalingam, P. Agrawal, and S. Kishore, "Dynamic resource allocation schemes during handoff for mobile multimedia wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 7, pp. 1270-1283, July 1999.

[15] R. Guerin, "Equivalent capacity and its application to bandwidth allocation in high-speed networks," *IEEE Journal on Selected Areas in Communications*, vol. 9, no. 7, pp. 968-981, Sept. 1991.

[16] J. S. Evans and D. Everitt, "Effective bandwidth-based admission control for multi-service CDMA cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 1, Jan. 1999.

[17] Q. Ren and G. Ramamurthy, "A real-time dynamic connection admission controller based on traffic modeling, measurement, and fuzzy logic control," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 2, Feb. 2000.

[18] Y. Fang and I. Chlamtac, "Teletraffic analysis and mobility modeling for PCS networks," *IEEE Transactions on Communications*, vol. 47, pp. 1062-1072, July 1999.

[19] Y. Fang, I. Chlamtac, and Y. B. Lin, "Channel occupancy times and handoff rate for mobile computing and PCS networks," *IEEE Transactions on Computers*, vol. 47, pp. 679-692, June 1998.

[20] E. A. Yavuz and V. C. M. Leung, "Computationally efficient method to evaluate the performance of guard-channel-based call admission control in cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 4, pp. 1412-1424, July 2006.

[21] S. S. Rappaport, "The multiple call handoff problem in personal communications networks," *IEEE 40th Vehicular Technology Conference*, pp. 287–294, May 1990.

[22] S. S. Rappaport and G. Monte, "Blocking, hand-off and traffic performance for cellular communication systems with mixed platforms," *IEEE 42nd Vehicular Technology Conference*, vol. 2, pp. 1018–1021, May 1992.

[23] W. Li and X. Chao, "Modeling and performance evaluation of a cellular mobile network," *IEEE/ACM Transactions on Networking*, vol. 12, no. 1, Feb. 2004.

[24] A. Gersht and K. J. Lee, "A bandwidth management strategy in ATM networks," Technical report, GTE Laboratories, 1990.

[25] J. W. Roberts, "Teletraffic models for the Telecom 1 integrated services network," *Proceedings of the 10th International Teletraffic Conference*, Montreal, 1983.

[26] S. C. Borst and D. Mitra, "Virtual partitioning for robust resource sharing: computational techniques for heterogeneous traffic," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 5, pp. 668-678, June 1998.

[27] K. W. Ross, Multi-service Loss Models for Broadband Telecommunication Networks. London: Springer-Verlag, 1995, chapter 3.

[28] W. J. Stewart, Introduction to the Numerical Solution of Markov Chains. Princeton University Press, Princeton, New Jersey, 1994.

CHAPTER 4 PROBABILITY DISTRIBUTION OF CHANNEL OCCUPANCY TIMES AND NUMBER OF USER HANDOFF IN CELLULAR NETWORKS[3]

4.1 Introduction

The evolution of cellular networks has been accompanied not only by voice but also by the development, growth, and use of a wide variety of network applications. Packet-switched services are gradually integrated with the conventional circuit-switched ones such as voice to accommodate network applications that range from text-based utilities such as SMS messaging, file transfer, remote login and electronic mail to MMS messaging, videoconferencing, multimedia streaming, web surfing and electronic commerce. The advantage of having packet-switched data services overlaid on circuit-switched technology over the same air interface is the utilization of excess network capacity available in each cell. Statistical multiplexing is used to transmit data packets over radio interface to provide a QoS level comparable to that of circuit-switched services. Statistical multiplexing gain arises from the talk spurt to silence ratio found in speech which makes it possible to multiplex more than one service on to the same radio channel. However accurate voice traffic statistics are needed to understand the length and frequency distributions of idle periods of cellular channels assigned to voice in order to exploit the statistical multiplexing gain in cellular networks.

Traffic statistics are important not only for statistical multiplexing and thus network management but also for performance evaluation in communication networks, billing, optimization and allocation of safety buffers. These statistics simulate network traffic when evaluating networks using mathematical models. Call holding and channel occupancy times are two of these traffic statistics that are needed to compute performance metrics such as call blocking/dropping probabilities. *Call holding time* is defined as the duration of requested call connection which corresponds to holding time for a wired phone call or session time in a computer system, whereas *channel occupancy time* is defined as the time that a mobile

[3] **A version of this chapter has been accepted for publication. E. A. Yavuz and V. C. M. Leung, "Modeling Channel Occupancy Times for Voice Traffic in Cellular Networks,"** ***IEEE ICC 2007*****, Glasgow, Scotland, June 2007.**

occupies channel(s) within a cell during its residence in the cell. In the literature, wireless network traffic has been approximated based on wireline traffic statistics where call holding times are generally considered to be exponentially distributed. Hong and Rappaport [1] proposed a traffic model for cellular mobile radio telephone systems and showed that channel occupancy time distribution can be approximated by exponential distribution only when call holding times are exponentially distributed. In the literature cellular networks have been mostly evaluated under this assumption due to its tractability. Ramjee *et al.* [2], Fang and Zhang [3], Naghshineh and Schwartz [4], Gersht and Lee [5], Borst and Mitra [6] and Yavuz and Leung [7] studied performance of various call admission control schemes using one dimensional Markov chain models assuming that channel occupancy times are exponentially distributed. In [8], Rappaport developed multidimensional models under the same assumption and with Monte, the author obtained call blocking probabilities using this model [9]. However simulation studies and field data have shown that these assumptions are not perpetually valid. Guerin [10] used a simulation model to show that channel occupancy time distribution displays a rather poor agreement with the exponential fitting for mobile users with low change rate of movement direction. Jedrzycki and Leung showed in [11] that exponential distribution assumption for channel occupancy times is not correct and a lognormal model approximation fits much better using real cellular data. In [12] and [13], Fang *et al.* demonstrated that channel occupancy times in a cellular network depend not only on call holding times but also on users' mobility which can be characterized by cell residence time distribution. The authors showed in [14] that channel occupancy time is exponentially distributed only if cell residence time is exponentially distributed. However it is also observed in the same study that channel occupancy time distribution have a good approximation by exponential distribution in general when the mobility is low. In [15], Barcelo and Jordan analyzed a cellular network based on a fully empirical approach and observed that channel occupancy is less spread out than if exponential distribution was assumed.

In this chapter, we present an analysis of real traffic data obtained from a number of cell sites in Bell Mobility Canada's cellular network. Similar to the empirical study presented in [11], we obtain the probability distributions for channel occupancy times however we classify them according to their occupancy types to perform goodness-of-fit tests for each type of channel occupancy times and users. In order to facilitate this classification, we briefly review the life cycle of a typical cellular call first. In a cellular network, the service area is covered by base stations whose radio coverage defines the corresponding cell. Each base station is assigned a set of mobile users. When a new call is originated by a mobile user in a cell, one of the channels assigned to the base station is used for communication between that mobile user and the base station if a channel is available. If a channel can be assigned to a call, it will be kept until the call is completed or the user moves out of the corresponding cell. When the user moves into a new cell while having an active call, a new channel needs to be acquired in the new cell using a "handoff procedure." We name the amount of time that a call occupies any channel during its holding time *total* regardless of being started in a cell and completed in the same cell or handoff to another. We classify channel occupancy times based on the following occupancy type characteristics:

- channels occupied by calls that are started and completed in the same cell (*new2same*).
- channels occupied by calls that are started in a cell but handed off to a neighboring cell (*new2ho*).
- channels occupied by calls that are started in a cell and either completed in the same cell or handed off to a neighboring cell (*new2sameorho*). This type of channel occupancy is very important for Markov chain modeling of cellular networks and classified as "new calls" in the models.
- channels occupied by calls that are handed off to a cell and completed in that cell (*ho2same*).
- channels occupied by calls that are handed off to a cell but handed off once more before completed (*ho2ho*).
- channels occupied by calls that are handed off to a cell and either completed in that cell or handed off once more before completed (*ho2sameorho*). This type of channel occupancy is very important for Markov chain modeling of cellular networks and classified as "handoff calls" in the models.

The benefits of good knowledge about various types of channel occupancy times are basically twofold.

1. Analytical models are developed using Markov chain modeling to evaluate performance in cellular networks. When a call admission control scheme is modeled for each cell in a cellular network, arriving calls to a particular cell are grouped into QoS classes or call types, such as new and handoff, based on their first appearance in the corresponding cell. Channel occupancy time distribution for each QoS class or call type includes respective channel occupancy times counted only until the corresponding calls discard the occupied channels in the cell due to call termination or handoff. Call holding time distribution, on the other hand, includes the amount of times that the channels are occupied by a call until it terminates either in its originating cell or another. The results presented in this paper are useful for providing sufficiently representative channel occupancy time statistics to develop analytical models since call holding time statistics are not sufficient alone.
2. Simulation provides a second approach towards network performance evaluation by building a mathematical model of the network to analyze its behavior as time progresses. The results presented in this chapter are very useful for feeding simulations with realistic traffic statistics to obtain network performance metrics.

This chapter is organized as follows. In Section 4.2, we explain the data acquisition method that we use to obtain statistics for various types of channel occupancy times and discuss system related anomalies that we observe. Section 4.3 consists of the candidate distributions that we propose to represent the empirical data set, the statistical tools, parameter estimation techniques and goodness-of-fit tests utilized in the study. In Section 4.4, we present the statistical results obtained from the goodness-of-fit tests performed for each aforementioned type of channel occupancy times along with the observed data histograms and fitted distributions. We examine how modeling call holding times with the best fitting candidate distribution would impact performance metrics such as call blocking probabilities instead of modeling with the traditionally accepted exponential distribution. We provide the statistical results obtained from the goodness-of-fit tests performed for channel occupancy and call holding time distributions for stationary and mobile users. Finally we present the statistical results obtained from the goodness-of-fit test performed to fit the distribution of the number of handoffs committed by a user to a candidate distribution. We will conclude the chapter in Section 4.5.

4.2 Data Acquisition and System-Related Anomalies

In this chapter, we analyze cellular call data obtained from the CDMA system deployed by Bell Mobility in Ontario, Canada. In Bell's CDMA system a call can be in up to 6 way soft/softer handoff at anytime. However call handoffs have been modeled differently in the literature in traditional mathematical and simulation network models developed for evaluating performance in cellular networks: a call is traditionally assumed to communicate via one primary sector at any given time unless it is in handoff. In the empirical data set, we obtain the values of E_c/I_0, ratio of the pilot signal energy to the total power in the channel, for each active call by observing the corresponding "Neighbor List Tuning Data Array" messages to determine a call's primary sector. An example of a "Neighbor List Tuning Data Array" message is given in Table 4.1. For a call at any given time, we take the sector which the call has the highest E_c/I_0 value in its active set as the primary sector of that call for that particular time. We compute E_c/I_0 value of a call in a sector by dividing the value of call's "pilot strength" given for that sector in the "Neighbor List Tuning Data Array" message by -2. For example for the sample message given in Table 4.1, E_c/I_0 value of the active call is equal to -12 for its primary sector where the corresponding "pilot strength" is equal to 24.

We assume that handoffs are technology independent which happen at the equal power boundaries. We use the "Neighbor List Tuning Data Array" messages to detect the committed handoffs by observing a call's primary sector being replaced by other sectors that have the highest E_c/I_0 value in the corresponding call's active set at the time of observation. We take an entirely empirical approach in this work based on true data collected from actual working systems. However unlike analytical and simulation approaches, the empirical approach depends on the environment and therefore may contain system related anomalies. The results presented in this paper might have been different if taken in a different place, time etc. We believe that all approaches are complementary and have more or less advantages or disadvantages depending on the specific application. Figs. 4.1 to 4.7 depict the distribution of channel occupancy times classified as *new2same*, *new2ho*, *new2sameorho*, *ho2same*, *ho2ho*, *ho2sameorho* and *total* respectively. Upon close inspection of the distributions given in Figs. 4.1 to 4.7, we observe two sorts of anomalies: unusually high number of short channel occupancy times and spikes. The unexpected behavior of channel occupancy times classified as *ho2ho* (see Fig. 4.5) and *ho2sameorho* (see Fig. 4.6) are due to the "pilot pollution" in the data set since several pilot signals are observed to be close to the "add-drop" thresholds and thus come in and get out of their active sets frequently (no dominant primary sector). The unexpected behavior of the channel occupancy times classified as *new2same* (see Fig. 4.1), *new2sameorho* (see Fig. 4.3) and *total* (see Fig. 4.7) is similar to what is observed by Bolotin in [16]. The author categorized the observed short channel occupancy times into four different classes in order of increasing average times [16]:

- various abandonment before the connection over the network is established.
- outgoing calls that encounter busy condition and therefore being abandoned by the caller.
- outgoing calls that encounter no answer condition and therefore being abandoned by the caller.
- outgoing calls that encounter a busy or no answer condition and therefore followed by a voice mail message left by the caller.

Second type of anomaly is the spikes that we observe within channel occupancy time samples classified as *new2same* around 16th second and *ho2ho* around 4th second. The former can be explained due to calls that encounter no answer and therefore go to voice mail while the latter can be explained due to "immediate handoff candidatecy". The spikes observed around 16th second within channel occupancy time samples classified as *new2sameorho* and *total* and the spikes observed around 4th second within channel occupancy time samples classified as *ho2sameorho* are due to same samples that created the spikes within channel occupancy time samples classified as *new2same* and *ho2ho*, respectively.

4.3 Statistical Methods

4.3.1 Candidate Probability Distributions

We choose the following theoretical distribution models that can reasonably represent the observed empirical distributions before carrying out the fitting estimations. We propose a list of candidate theoretical continuous distributions and describe their properties along with examples of processes for which they can serve as models. The list of the proposed candidate distributions is given below along with the main statistics provided in the Appendix.

1. *Exponential Distribution*: This distribution is often used to model the time between events that happen at a constant average rate. It is the only continuous *memoryless* probability distribution and thus highly appreciated for achieving analytical results in Markov processes, making it easier to analytically solve systems involving queues. The exponential distribution plays a strong role in the theory of congestion systems and modeling processes such as duration of traditional phone calls, the time between failures of certain types of electronic devices, and the time between interrupts received by a CPU in a computer system.
2. *Lognormal Distribution*: The lognormal distribution is the probability distribution of any random variable whose logarithm is normally distributed. A variable might be modeled as lognormal if it can be thought of the product of many small independent positive variables. A typical example is modeling the time until a system fails or the time to perform manual tasks such as assembly, inspection or repair.
3. *Gamma Distribution*: The gamma distribution is a continuous distribution that is often used to model the average lifetime or the sum of lifetimes of various items that have exponential lifetimes. It describes the time until n consecutive rare random events occur in a process with no memory. It is a popular candidate for modeling processes such as the time to perform a manual task and the CPU time a job requires due to its ability to assume many shapes. The exponential distribution and the Erlang distribution, which can be expressed as the sum of independent

exponential distributions, are the two important special cases of the gamma distribution.

4. *Weibull Distribution*: This distribution is often used in reliability analysis due to its versatility to model the distribution of time until failure. In particular every exponential distribution is also a Weibull distribution since a Weibull distribution is defined by modifying the constant event occurrence rate of an exponential distribution to make it time dependent.

Other distributions such as Beta, Poisson or Pareto are not proposed as candidate distributions in this study due to lack of theoretical and empirical criteria to support them as candidates.

4.3.2 Parameter Estimation

Once the candidate distributions are proposed, we need the parameters for each distribution estimated from the empirical data set. There are many methods to estimate a particular parameter of a given distribution such as probability plotting, method of moments and maximum likelihood estimation. A *probability plot* is a graphical comparison of an empirical distribution with that of a candidate distribution. The *method of moments* consists of equating the first few moments obtained from an empirical data set with the corresponding moments of a candidate distribution to solve the number of equations that match with the number of unknown parameters to obtain the required estimates. This method usually yields fairly simple and consistent estimators, however the given estimators can be biased and inefficient. The *maximum likelihood estimation* (MLE) method consists of obtaining parameters that maximize the probability of obtaining the empirical data in the whole sample set [17]. The principle of this method is to select a value as an estimate for which the observed sample is most likely to occur.

In this chapter, we use the maximum likelihood estimation method to obtain the required parameters since it usually produces consistent estimators. It is also shown to be the most efficient method under certain regularity conditions when the sample size approaches infinity [18]. The advantage of using MLE over other methods is that it gives better fit results than method of moments when the goodness of fit test is applied [17]. MLE captures the shape of the empirical distribution much better than the method of moments since it is not tied to the first moments of the sample. Even though the method usually works quite well, it has some drawbacks: the estimators may be biased for small sample sizes; the method has a higher complexity of parameter computation; and the moments of the empirical and candidate distributions may not agree. The difference between the moments of the candidate and empirical distribution is slight for distributions with good fit; however the system of equations that needs to be solved to compute the required parameters may not always be a closed form solution or a unique solution may not even exists.

4.3.3 Goodness-of-fit Tests

Statistical tests that determine whether a given theoretical probability distribution is appropriate to characterize an observed sample data set are called goodness-of-fit tests [19]. These tests are statistical hypothesis tests that are used to assess formally whether the observed data are independent samples from a particular distribution. However failure to reject the null hypothesis that claims the observed data samples to be IID random variables with a particular distribution function, should not be interpreted as accepting the null hypothesis as being true. Law and Kelton noted in [17] that these tests are often not very powerful for small to moderate sample sizes and thus should be regarded as a systematic approach for detecting fairly gross differences instead. Yet if the sample size is very large, the authors observed that these tests almost always reject the null hypothesis. Even an instant departure from the hypothesized distribution is detected for a large sample size, since the null hypothesis is virtually never exactly true. Therefore it is more important to find the theoretical distribution that fits the empirical data best even if it may be rejected with a "relatively small" margin by the respective goodness of fit test. In this chapter we are not only looking for an accepted theoretical distribution, but also an order in which the candidate distributions fit the empirical data set since it is usually sufficient to have a distribution that is "nearly" correct due to its benefits such as [15];

- building appropriate queuing models with the corresponding general distributions to simulate systems since highly accurate performance metrics can also be obtained using approximations.
- exploiting idle times in the traffic channels of the mobile systems for the insertion of data on a non-preemptive basis.
- scheduling the channel to be interrupted when an emergency call arrives at a blocked system (to interrupt the channel with the lowest expected remaining occupancy time or to predict the first available channel for handoff).

The *chi-square* test is the oldest goodness-of-fit hypothesis test that can be thought of as a more formal comparison of a histogram with the proposed candidate probability distribution. To compute the chi-square test statistic in either continuous or discrete case, the entire range of candidate distribution must be divided into k adjacent intervals $[t_0, t_1)$, $[t_1, t_2)$, … , $[t_{k-1}, t_k)$, where t_0 and t_k can either or both be -∞ and +∞, respectively. Then we check

$$N_j = \text{number of observations in the } j\text{th interval } [t_{k-1}, t_k)$$

for $j = 1, 2, \ldots, k$ and compute the expected proportion p_j of the observations that would fall in the jth interval if we were sampling from the candidate probability distribution. In the continuous case,

$$p_j = \int_{t_{j-1}}^{t_j} f_c(x) \cdot dx \qquad (1)$$

where $f_c(x)$ is the probability function of the candidate continuous distribution. For discrete data,

$$p_j = \sum_{t_{j-1} \le x_i < t_j} f_d(x_i) \tag{2}$$

where $f_d(x_i)$ is the probability function of the candidate discrete distribution. Finally, the test statistic is

$$\chi^2 = \sum_{j=1}^{k} \frac{(N_j - np_j)^2}{np_j} \tag{3}$$

The test statistic is expected to be small if the fit were good since np_j will then be the expected number of observations that would fall in the *j*th interval. The most troublesome aspect of carrying a chi-square goodness of fit test is choosing the number and size of the bins. This is a difficult problem and no definitive prescription can be given that is guaranteed to produce good results in terms of validity (actual level of the test close to the desired level α) and high power for all hypothesized distributions and all sample sizes. Therefore the major drawback of the chi-square test is the lack of clear prescription for interval selection. However Law and Kelton suggested a few guidelines in [17]. The authors proposed the *equiprobable approach* which chooses the bin intervals so that the expected proportion of the empirical data set that fall in each interval will be equal to each other. Even though this might be inconvenient to apply to some continuous distributions since the distribution function of the candidate distribution must be inverted, it will be possible to make the values of the expected proportion of the empirical data set approximately equal for discrete distributions. It is also stated in [17] that the chi-square test will be approximately valid if the number of bins is greater than 3 and the minimum number of expected number of observations in a bin is 5 for equiprobable intervals.

Kolmogorov-Smirnov (K-S) tests, on the other hand, compare an empirical data set with a candidate distribution without grouping the data. Thus, no information is lost when applying this test which eliminates the troublesome problem of interval specification. It is valid for any sample size however the major drawback of a K-S test is its range of applicability which is more limited than that for chi-square tests [17]. These tests are valid only if all the parameters of the hypothesized distribution are known and the distribution is continuous, which means that the parameters cannot be estimated from the empirical data set and even if it will; this in fact will produce a conservative test. The K-S test statistic is the largest (vertical) distance between the empirical and the candidate distribution function, however giving the same weight to this distance for all observed values is another drawback of these tests since many distributions of interest differ primarily in their tails. The *Anderson-Darling* (A-D) test is a modification of the K-S test that is designed to detect these discrepancies in the tails. It has higher power than the K-S test against many alternative distributions yet it is only available for a few specific distributions [17].

In this study, we use the chi-square test which remains in wide use since it can be applied to any hypothesized distribution with parameters estimated from the observed data. We perform the following steps:

1. Divide the data into groups with respect to occupancy types and sort it by channel occupancy times. This will let us visually see the data distribution for each group with a different occupancy type as a histogram. The normalized histograms for all channel occupancy types are shown in Figs. 4.1 to 4.7.
2. Choose a list of candidate distributions to which each group of data will be fitted and calculate the parameters for each distribution using the corresponding maximum likelihood estimation equations.
3. Choose the number of bins into which each group of data will be divided, calculate the expected number of observations in each bin and confirm that it exceeds the *minimum number of expected number of observations* given in [17]. Each bin should be created to assure that its expected number of observations is equal to that of others created for the same candidate distribution being tested. Therefore the bins have to be recalculated every time a different set of distribution parameters are used or a new distribution is tested. For each group of data, the bin boundaries should satisfy

$$\lfloor np_j = n \cdot \left(pr\{X < t_j\} - pr\{X < t_{(j-1)}\} \right) \rfloor \geq 5 \qquad (4)$$

where $j = 1, 2, \ldots k$

p_j is the probability of a data item from the distribution that falls into bin j,

k is the number of created bins,

$t_0, t_1, \ldots t_k$ are the upper bin boundaries,

n is the total number of observations in the data set and

np_j is the expected number of observations in the jth bin.

4. Divide the observed data in each group into its corresponding created bins whose upper boundaries are given in the previous step. However each data group should be made continuous in advance by spreading it evenly within 0.5 seconds of their discrete values since all candidate distributions that we propose are continuous due to their nature.
5. Calculate the test statistics for each data group using (3).
6. The null hypothesis is rejected in each case if the value of the test statistics calculated by the previous step is greater than the value of the chi-square statistics with $k - 1 - z$ degrees of freedom, where z is the number of parameters estimated. We set the significance level of the performed tests equal to the traditional value of 0.95 ($\alpha = 0.05$).

4.4 Numerical Results

4.4.1 Statistical Results for Channel Occupancy Times

In this section, we present the statistical results obtained from the goodness-of-fit tests performed for each aforementioned type of channel occupancy times. The results show that all types of channel occupancy times resemble either lognormal or weibull distributions. However only for two of these, *new2ho* and *ho2new*, the candidate probability distributions were able to pass the chi-square goodness-of-fit test with lognormal being a better fit than weibull. Upon close inspection of the rest of data histograms given in Figs. 4.1, 4.3, 4.5, 4.6 and 4.7, we observed two sorts of anomalies which we addressed in section 4.2: unusually high number of short channel occupancy times and spikes. None of these distributions fits statistically to a proposed candidate distribution due to the observed anomalies. Thus, filtering these empirical data sets is inevitable since no candidate distribution can otherwise be fitted to the data.

All short channel occupancy times less than 3 seconds are discarded from the respective data sets of channel occupancy time distributions classified as *new2same*, *new2sameorho* and *total* since most of them are calls terminated abnormally due to reasons given in section 4.2. The excess channel occupancy times observed at the 16th second in the same data sets given above and at the 4th second in the data sets of channel occupancy time distributions classified as *ho2ho* and *ho2sameorho* are stripped from the rest using simple means. The test and the chi-square statistics obtained from the revised goodness-of-fit tests for each type of channel occupancy are presented in Tables 4.2 to 4.8. The null hypothesis is rejected in each case if the value of the test statistics is greater than the value of the chi-square statistics. All significant levels are 0.95, the candidate probability distributions which pass the chi-square goodness of fit test are marked with a star (*) and the test statistics for the closest fitted candidate distributions given in the tables are **bolded**. Figs. 4.1 to 4.7 show the closest fitted candidate distribution and the fitted exponential distribution along with the normalized histograms for all types of channel occupancy times, respectively.

Figs. 4.1, 4.3 and 4.7 show the traditionally accepted fitted exponential distribution underestimating the short channel occupancy times while overestimating the rest except around 16th second, where unusually high number of channel occupancy times is observed. However the closest fitted candidate distribution, lognormal, slightly overestimates the short channel occupancy times except at 16th second while it matches well with the rest. For channel occupancy types of *new2ho* and *ho2same*, it is shown in Figs. 4.2 and 4.4 that the fitted exponential distribution underestimates the short channel occupancy times while it overestimates the rest. The closest fitted candidate distribution, lognormal, matches the observed distribution very well except for very short channel occupancy times which it slightly overestimates. Figs. 4.5 and 4.6 show the channel occupancy time distributions for types of *ho2ho* and *ho2sameorho*. The traditionally accepted fitted exponential distribution underestimates the short channel occupancy times while it overestimates the rest. Yet, the closest fitted candidate distribution, lognormal, slightly overestimates the short channel occupancy times except around 4th second while it matches well with the rest. We observe in general that lognormal distribution is by far the closest fitted distribution for each channel occupancy type when compared with other candidate distributions.

4.4.2 The Effects of Traffic Remodeling on Performance Evaluation

In this section, we demonstrate how modeling call holding times with a lognormal distribution would impact performance metrics such as call blocking probabilities instead of modeling with a traditionally accepted exponential distribution. Let us assume that we are modeling a cellular network in a single cell with a fixed amount of bandwidth capacity, *C*. We assume that a Poisson process describes the arrival of each call that requires a single channel and when all channels are occupied, all arriving calls are assumed to be rejected and thus lost. We can compare this model to an *M/M/C* system with no queues available. First, we simulate our model using exponentially distributed call holding times with various mean values ($1/\lambda$) within the range of 10 sec/call to 120 sec/call that are set equal to the mean values of all types of channel occupancy times obtained from the empirical data set using the maximum likelihood estimation. We assume that the bandwidth capacity of the cell, *C*, is equal to 20 and we choose a constant average call arrival rate of 1 call/sec to obtain call blocking probabilities that overlay [0, 1] interval exclusively. Then we simulate the model using lognormally distributed call holding times with the corresponding μ's and σ's obtained from the same empirical data set to have equivalent mean values with the exponentially distributed call holding times. The results are given in Fig. 4.8. We observe that call blocking probabilities obtained when call holding times are exponentially distributed match well with the call blocking probabilities obtained when call holding times are lognormally distributed provided that both distributions have means that are alike. However, we shall note that it is more challenging to compute the parameters (μ and σ) of a lognormal distribution using the expected value obtained from an empirical data set when compared to computing the mean value for an exponential distribution.

4.4.3 Channel Occupancy and Call Holding Times of Stationary and Mobile Users

In this section, we present the statistical results obtained from the goodness-of-fit tests performed to fit a proposed candidate distribution to observed channel occupancy and call holding times when these distributions are grouped with respect to user mobility: stationary and mobile. In [12] and [13], the authors demonstrated theoretically that traffic characteristics such as channel occupancy times depend not only on call holding times but also on users' mobility which can be characterized by cell residence times. However, it is difficult to obtain cell residence times from the data sets since the observed data are collected from network nodes which track idle mobile users by exchanging messages very infrequently. Thus, we classify users in our data set with respect to their mobility characteristics based on the number of handoffs that they commit during a call. We identify each user with zero number of handoffs *stationary* or *low mobility* and the rest (with number of handoffs more than zero) *mobile.* Note that some stationary users may in fact be physically mobile within a cell yet we still consider them stationary with respect to their cell residency. We consider users *mobile* if only the number of handoffs that they commit is more than a certain threshold (set to 3 handoffs in this study) since "pilot pollution" may cause a stationary call to commit handoff once in a while.

Users that are affiliated with *new2same* type of channel occupancy times can be considered stationary whereas users affiliated with *new2ho*, *new2sameorho*, *ho2same*, *ho2ho* and *ho2sameorho* types of channel occupancy times can be considered mobile. Thus, call holding time distribution for stationary users is equivalent to channel occupancy time distribution for *new2same.* However we have to obtain call holding time distribution for mobile users (*total_mobility*) separately from the data set since the call holding time distribution obtained previously includes times affiliated with both stationary and mobile users. The test results and chi-square statistics for all types of channel occupancy times are previously given. We apply the chi-square goodness of fit test only to the *total_mobility* data set after the observed anomalies were stripped using simple means. The results are presented in Table 4.9 where significant level is 0.95, the candidate probability distribution which passes the chi-square goodness of fit test is marked with a star (*) and the test statistics for the closest fitted candidate distribution given in the table is **bolded**. Figure 4.9 shows the closest fitted candidate distribution, lognormal, along with the normalized histogram for mobile users' call holding times.

In [14], Fang, Chlamtac and Lin showed analytically that when cell residence times are exponentially distributed, channel occupancy time distribution for "new calls" can be approximated by the fitted exponential distribution for stationary users (or when mobility is low) and yet for mobile users there is a significant mismatch between channel occupancy time distribution for "handoff calls" and the fitted exponential distribution. In this study, we observe that not only channel occupancy and call holding time distribution for mobile users fit lognormal distribution very strongly compared to other proposed candidate distributions but also both distributions for stationary users fit lognormal distribution while radically differing from the fitted exponential distribution. Hence we expect cell residence times also not to be exponentially distributed.

4.4.4 Statistical Results for Number of Handoffs Committed by Users

In this section, we present the statistical results obtained from the goodness-of-fit test performed to fit a proposed candidate distribution to the distribution of number of handoffs committed by mobile users in a cellular network. The distribution of the number of handoffs becomes significant when feeding network simulations. The test and chi-square statistics are presented in Table 4.10 where significant level is 0.95, the candidate probability distribution which passes the chi-square goodness of fit test is marked with a star (*) and the test statistics for the closest fitted candidate distribution given in the table is **bolded**. Figure 4.10 shows the closest fitted candidate distribution, lognormal, along with the normalized histogram for users' number of committed handoffs.

4.5 Summary

In this chapter, we have presented an empirical approach to determine the probability distribution functions that fit various types of channel occupancy times for voice service in cellular networks. The results are environment dependent however we have made no assumptions that can influence the results as opposed to previous analytical and simulation studies where the obtained results can be highly dependent on the assumptions made by the authors. We have explained the data acquisition method that we used to obtain the statistics for various types of channel occupancy times and discussed the system related anomalies that we have observed. The statistical results obtained from the goodness-of-fit tests have been presented along with the candidate probability distributions that may fit the empirical data. We have shown that not only call holding times but also various types of channel occupancy times can be approximated by lognormal distribution. We have examined how modeling call holding times with a lognormal distribution would impact the value of performance metrics such as call blocking probabilities instead of modeling with an equivalent traditionally accepted exponential distribution. We have observed that call blocking probabilities obtained when call holding times are exponentially distributed match well with the call blocking probabilities obtained when call holding times are lognormally distributed provided that both distributions have equal means.

We have discovered that not only channel occupancy and call holding time distributions for mobile users fit the lognormal distribution very strongly compared to other proposed candidate distributions but also both distributions for stationary users fit the lognormal distribution very well while it radically differs from the exponential distribution. We have presented the statistical results obtained from the goodness-of-fit test performed to fit a proposed candidate distribution to the number of handoffs committed by mobile users in a cellular network. We have shown that the closest fitted candidate distribution to approximate the distribution of number of handoffs committed by users in a cellular network is also a lognormal distribution.

The results are expected to be useful in traffic and network modeling, performance evaluation, billing, network management and optimization in cellular networks.

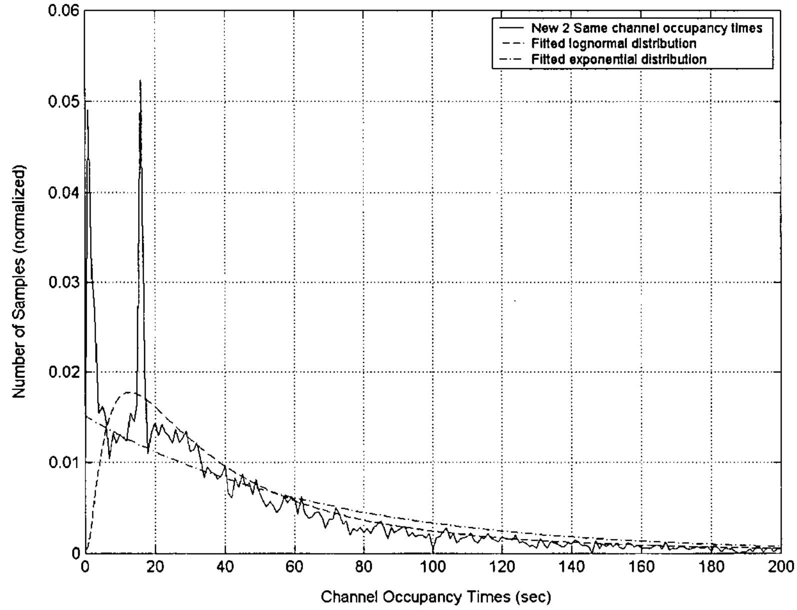

Fig. 4.1 Distribution of channel occupancy times (*new2same*) and the fitted *lognormal* and *exponential* distributions.

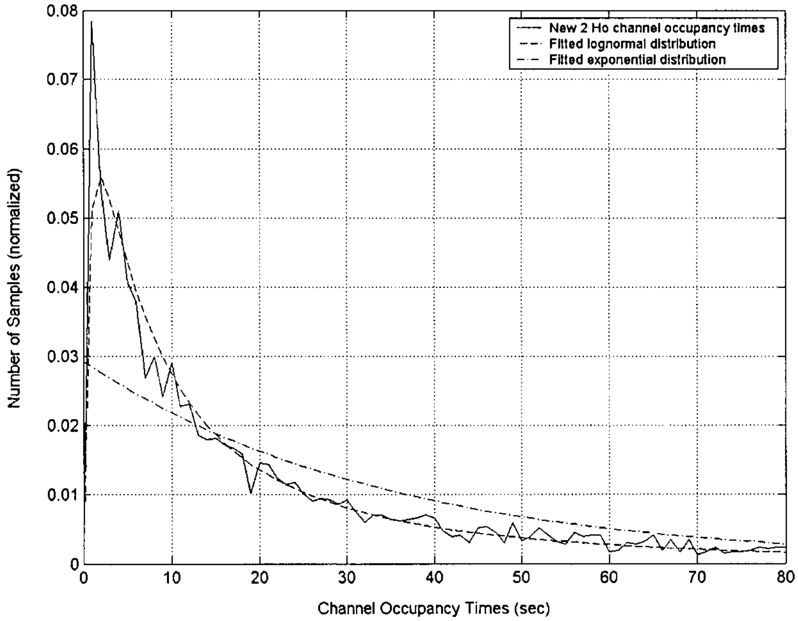

Fig. 4.2 Distribution of channel occupancy times (*new2ho*) and the fitted *lognormal* and *exponential* distributions.

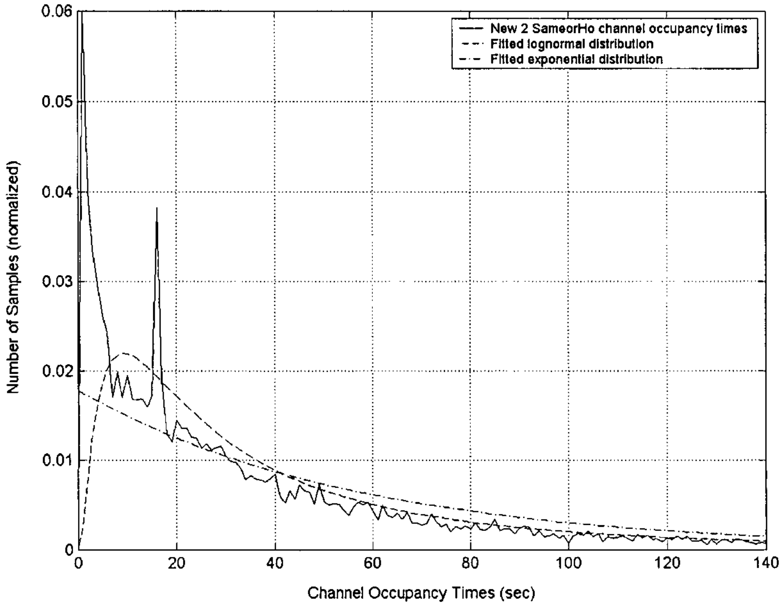

Fig. 4.3 Distribution of channel occupancy times (*new2sameorho*) and the fitted *lognormal* and *exponential* distributions.

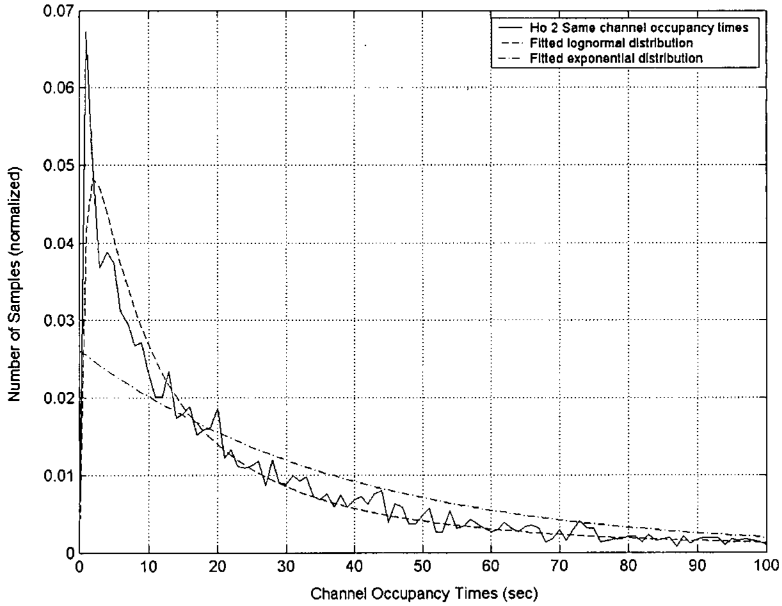

Fig. 4.4 Distribution of channel occupancy times (*ho2same*) and the fitted *lognormal* and *exponential* distributions.

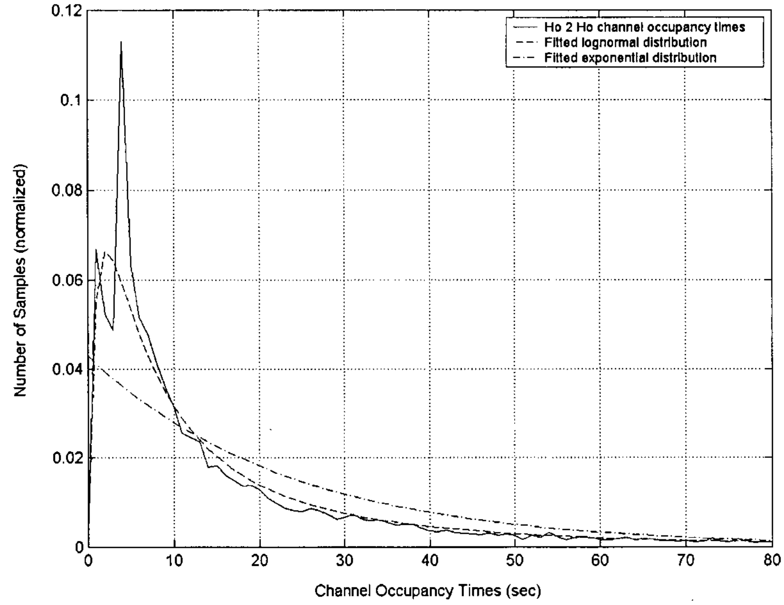

Fig. 4.5 Distribution of channel occupancy times (*ho2ho*) and the fitted *lognormal* and *exponential* distributions.

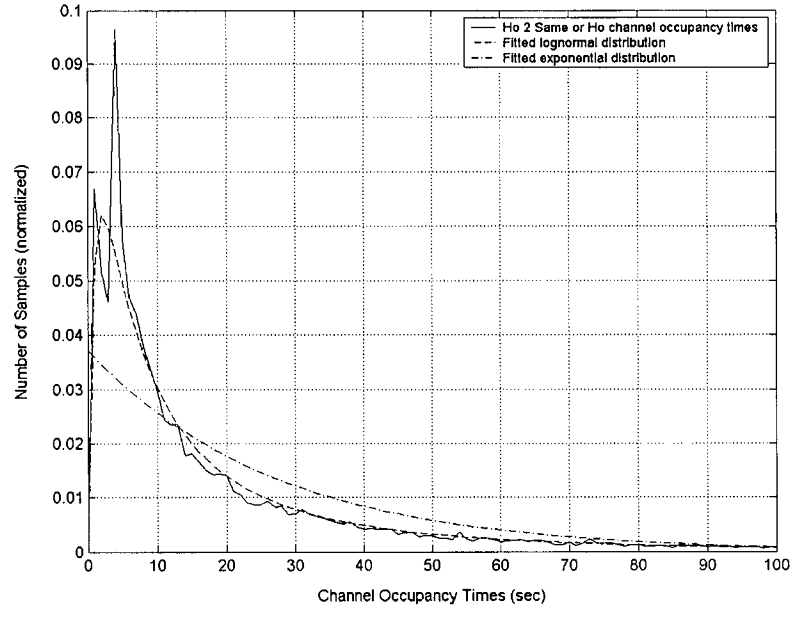

Fig. 4.6 Distribution of channel occupancy times (*ho2sameorho*) and the fitted *lognormal* and *exponential* distributions.

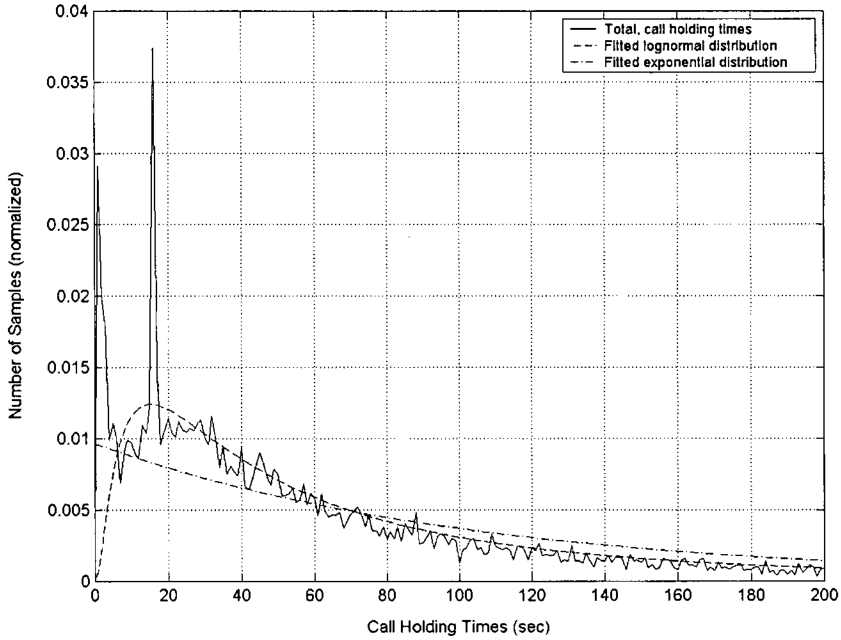

Fig. 4.7 Distribution of call holding times (*total*) and the fitted *lognormal* and *exponential* distributions.

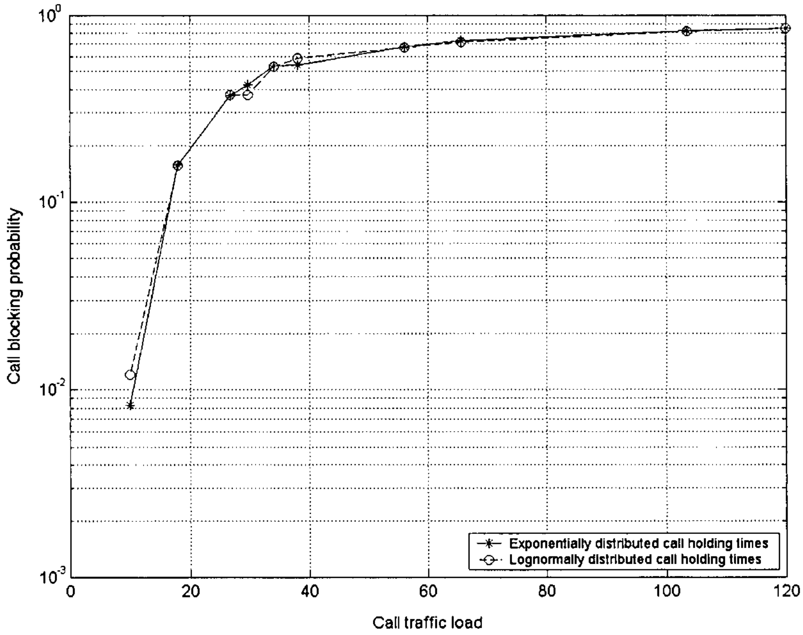

Fig. 4.8 Call blocking probabilities for *exponentially* and *lognormally* distributed call holding times.

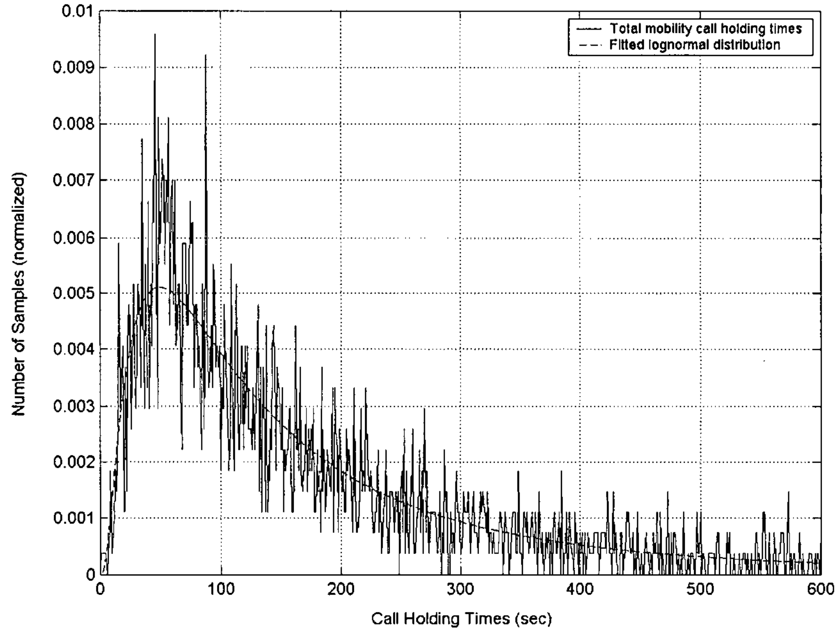

Fig. 4.9 Distribution of call holding times (*total_mobility*) and the fitted *lognormal* distribution.

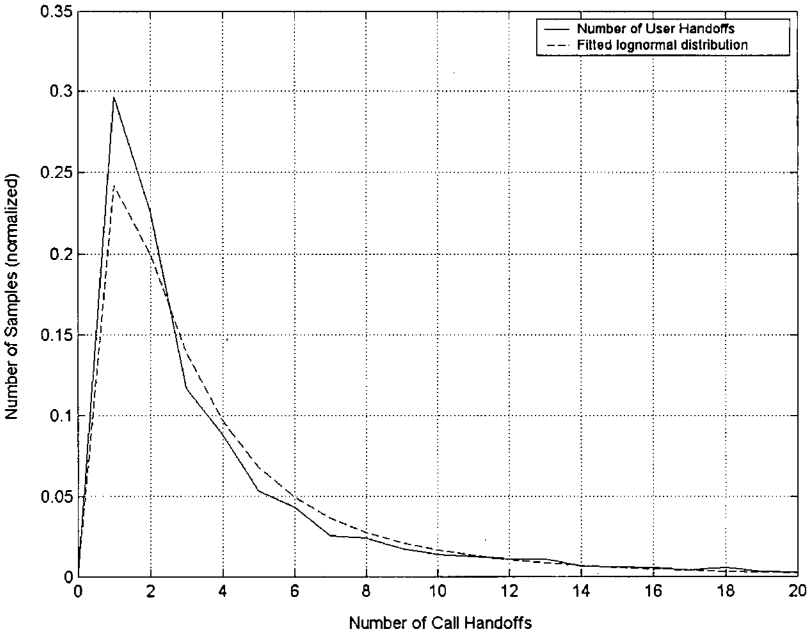

Fig. 4.10 Distribution of number of user handoffs (*user_handoffs*) and the fitted *lognormal* distribution.

TABLE 4.1

An example of a "*Neighbor List Tuning Data Array*" message taken from Bell Mobility cellular call data set.

Attribute Name:	Neighbor List Tuning Data Array
TimeStamp:	2003/09/19 – 11:46:28.240
Source Node Id:	0xd6ad45
Call Id:	0x00000520cb286664
Resource Info:	Frame 4 Shelf 1 Slot 15 DSP 7

BaseId	**KeepBit**	**PilotStrength**	**PNOffset**	**PNPhase**
0x09771101	1	32	36	0
0x09770e21	1	31	222	14225
0x09771372	1	26	300	19209
0x09770e72	1	28	336	21507
0x09771073	1	34	396	25357
0x09771081	1	27	114	7304
0x09770e71	1	40	330	21124
0x09771382	1	44	372	23828
0x09771071	1	24	384	24589

TABLE 4.2

Goodness of fit test results for Channel Occupancy Time distribution fitting (*new2same*).

Distribution	**Test Statistics**	**Chi-square Statistics**
Exponential	9.8955e+02	5.5102e+02
*Lognormal**	**3.1824e+02**	5.4997e+02
Gamma	1.3068e+03	5.4997e+02
Weibull	1.178e+03	5.4997e+02
number of bins: 500, number of data samples per bin: 11.892		

TABLE 4.3

Goodness of fit test results for Channel Occupancy Time distribution fitting (*new2ho*).

Distribution	Test Statistics	Chi-square Statistics
Exponential	1.2794e+03	5.5102e+02
*Lognormal**	**3.0301e+02**	5.4997e+02
Gamma	6.8892e+03	5.49975e+02
*Weibull**	5.4559e+02	5.4997e+02
number of bins: 500, number of data samples per bin: 9.336		

TABLE 4.4

Goodness of fit test results for Channel Occupancy Time distribution fitting (*new2sameorho*).

Distribution	Test Statistics	Chi-square Statistics
Exponential	1.7965e+03	5.5102e+02
Lognormal	**6.9142e+02**	5.4997e+02
Gamma	1.7968e+03	5.4997e+02
Weibull	1.67174e+03	5.4997e+02
number of bins: 500, number of data samples per bin: 19.554		

TABLE 4.5

Goodness of fit test results for Channel Occupancy Time distribution fitting (*ho2same*).

Distribution	Test Statistics	Chi-square Statistics
Exponential	1.1922e+03	5.5102e+02
*Lognormal**	**3.325e+02**	5.4997e+02
Gamma	6.651e+02	5.4997e+02
*Weibull**	5.1167e+02	5.4997e+02
number of bins: 500, number of data samples per bin: 10.832		

TABLE 4.6

Goodness of fit test results for Channel Occupancy Time distribution fitting (*ho2ho*).

Distribution	**Test Statistics**	**Chi-square Statistics**
Exponential	3.4501e+03	5.5102e+02
Lognormal	**5.5853e+02**	5.4997e+02
Gamma	2.9586e+03	5.4997e+02
Weibull	2.3558e+03	5.4997e+02
number of bins: 500, number of data samples per bin: 35.13		

TABLE 4.7

Goodness of fit test results for Channel Occupancy Time distribution fitting (*ho2sameorho*).

Distribution	**Test Statistics**	**Chi-square Statistics**
Exponential	4.6679e+03	2.8574e+02
Lognormal	**6.4289e+02**	2.8466e+02
Gamma	3.6019e+03	2.8466e+02
Weibull	2.7327e+03	2.8466e+02
number of bins: 250, number of data samples per bin: 91.86		

TABLE 4.8

Goodness of fit test results for Call Holding Time distribution fitting (*total*).

Distribution	**Test Statistics**	**Chi-square Statistics**
Exponential	1.8421e+03	5.5102e+02
*Lognormal**	**4.3353e+02**	5.4997e+02
Gamma	2.8244e+03	5.4997e+02
*Weibull**	1.7486e+03	5.4997e+02
number of bins: 500, number of data samples per bin: 22.426		

TABLE 4.9

Goodness of fit test results for Call Holding Time distribution fitting (*total_mobility*).

Distribution	**Test Statistics**	**Chi-square Statistics**
Exponential	7.0154e+02	5.5102e+02
*Lognormal**	**3.1515e+02**	5.4997e+02
Gamma	6.4451e+02	5.4997e+02
Weibull	6.8223e+02	5.4997e+02
number of bins: 500, number of data samples per bin: 5.398		

TABLE 4.10

Goodness of fit test results for Number of User Handoffs distribution fitting (*user_handoffs*).

Distribution	**Test Statistics**	**Chi-square Statistics**
Exponential	1.0844e+03	2.5884e+02
*Lognormal**	**2.0881e+02**	2.5776e+02
Gamma	1.8877e+03	2.8466e+02
Weibull	9.62e+02	2.5344e+02
number of bins: 250, number of data samples per bin: 22.704		

4.6 Bibliography

[1] D. Hong and S. S. Rappaport, “Traffic model and performance analysis for cellular mobile radiotelephone systems with prioritized and non-prioritized handoff procedures,” *IEEE Transactions on Vehicular Technology*, vol. 35, pp. 77-92, Aug. 1986.

[2] R. Ramjee, R. Nagarajan, and D. Towsley, “On optimal call admission control in cellular networks,” *Wireless Networks*, vol. 3, no. 1, pp. 29-41, March 1997.

[3] Y. Fang, and Y. Zhang, “Call admission control schemes and performance analysis in wireless mobile networks,” *IEEE Transactions on Vehicular Technology*, vol. 51, no.2, pp. 371-382, March 2002.

[4] M. Naghshineh and S. Schwartz, “Distributed call admission control in mobile/wireless networks,” *IEEE Journal on Selected Areas in Communications*, vol. 14, no. 4, pp.711-717, May 1996.

[5] A. Gersht and K. J. Lee, “A bandwidth management strategy in ATM networks,” Technical report, GTE Laboratories, 1990.

[6] S. C. Borst and D. Mitra, “Virtual partitioning for robust resource sharing: computational techniques for heterogeneous traffic,” *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 5, pp. 668-678, June 1998.

[7] E. A. Yavuz and V. C. M. Leung, “Computationally efficient method to evaluate the performance of guard-channel-based call admission control in cellular networks,” *IEEE Transactions on Vehicular Technology*, vol. 55, no. 4, pp. 1412-1424, July 2006.

[8] S. S. Rappaport, “The multiple call handoff problem in personal communications networks,” *IEEE 40th Vehicular Technology Conference*, pp. 287-294, May 1990.

[9] S. S. Rappaport and G. Monte, “Blocking, hand-off and traffic performance for cellular communication systems with mixed platforms,” *IEEE 42nd Vehicular Technology Conference*, vol. 2, pp. 1018-1021, May 1992.

[10] R. A. Guerin, “Channel occupancy time distribution in a cellular radio system,” *IEEE Transactions Vehicular Technology*, vol. 35, no. 3, pp. 89-99, 1987.

[11] C. Jedrzycki, and V. C. M. Leung, “Probability distribution of channel holding time in cellular telephony systems,” *IEEE Vehicular Technology Conference (VTC’96)*, vol. 1, pp. 247-251, Apr. 1996.

[12] Y. Fang, I. Chlamtac, and Y.B. Lin, “Call performance for a PCS network,” *IEEE Journal on Selected Areas in Communications* vol. 15, no. 8, pp. 1568-1581, Oct. 1997.

[13] Y. Fang, I. Chlamtac, and Y.B. Lin, "Modeling PCS networks under general call holding times and cell residence time distributions," *IEEE Transactions on Networking* vol. 5, no. 6, pp. 893-906, Dec. 1997.

[14] Y. Fang, I. Chlamtac and Y.B. Lin, "Channel occupancy times and handoff rate for mobile computing and PCS networks," *IEEE Transactions on Computers* vol. 47, no. 6, pp. 679-692, June 1998.

[15] F. Barcelo, and J. Jordan, "Channel holding time distribution in public telephony systems (PAMR and PCS)," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 5, pp. 1615-1625, Sep. 2000.

[16] V. A. Bolotin, "Modeling call holding time distributions for CCS network design and performance analysis," *IEEE Journal on Selected Areas in Communications*, vol. 12, no.3, pp. 433-438, April 1994.

[17] A. M. Law and W. D. Kelton, *Simulation, Modeling and Analysis*, 3rd edition, New York: McGraw-Hill, 2000.

[18] K. S. Trivedi, *Probability and Statistics with Reliability, Queuing and Computer Science Applications*, New Jersey: Prentice-Hall, 1982.

[19] M. R. Sheldon, Introduction to Probability and Statistics for Engineers and Scientists, 3rd edition, USA: Elsevier Academic Press, 2004.

CHAPTER 5 CONCLUSION AND RECOMMENDATIONS FOR FURTHER WORK

5.1 Conclusion

We conclude this dissertation with a summary of our contributions and directions for future work. One dimensional Markov chain models are commonly used to evaluate call admission control schemes in cellular networks assuming that call requests that originate from different types of users are independently Poisson distributed, channel occupancy times for each call are exponentially distributed with equal mean values and each call requires an equal channel capacity. These assumptions may not be appropriate since calls with different priorities, such as new and handoff, may have different average channel occupancy times if not different distributions. When average channel occupancy times for different call types are not equal, existing performance evaluation approximation methods based on one dimensional Markov chain models lead to significant discrepancies. Thus, accurate solutions can only be obtained by exact analysis methods based on multidimensional Markov chain models. However, these methods suffer from the curse of dimensionality, which results in very high computational cost for large systems. In chapter 2, we proposed an easy to implement analytical performance evaluation approximation method, *effective holding time*, with low computational cost for several widely known call admission control schemes under more general assumptions. The proposed approximation method provides a highly accurate closed form solution and thus has low computational cost. We compared the accuracy of the proposed method with the existing approximation methods' with respect to exact results obtained from the direct method based on multidimensional Markov chain modeling. The results showed that the proposed method outperforms the existing approximation methods in accuracy while keeping the computational cost low. An accurate performance evaluation approximation method with low computational cost will motivate the practical implementation of dynamic call admission control schemes.

Conventional circuit-switched services such as voice are gradually being replaced by packet-switched data and multimedia applications due to increasing demand coming from users that these services shall also be available on the move. Conventional call admission control schemes, on the other hand, will continue to be useful when applied with suitable scheduling techniques to guarantee QoS at the packet level since most data and multimedia applications are inherently connection oriented and packet-switched connections can be provisioned to their effective bandwidths. Performance of call admission control schemes for multi-service cellular networks, which provide packet-switched services, can be evaluated by multidimensional Markov chain modeling since they have a similar form to circuit-switched networks. However calculating channel occupancy distribution of a multi-service cellular network involves numerically solving the balance equations when a multidimensional Markov chain model is used. In the absence of a product form solution this is demanding for all but smallest channel capacities, yet a computationally efficient one dimensional Markov chain model can only be used when all circuit and packet switched services have equal capacity requirements. Existing performance evaluation approximation methods for multi-service cellular networks are only accurate when call traffic loads are very low due to the assumption that channels are occupied independently. In chapter 3, we classified call admission control schemes into two categories called *symmetric* and *asymmetric*. We presented a product form solution formula to evaluate symmetric call admission control schemes and proposed a novel computationally efficient performance evaluation approximation method, *state space decomposition*, to evaluate asymmetric call admission control schemes when all services have distinct capacity requirements. We compared the numerical results obtained from the proposed method with the ones obtained from previously proposed approximation methods and the numerical exact method based on multidimensional Markov chain modeling. The results showed that proposed method provides more accurate solutions while keeping the computational cost low.

Packet-switched services are overlaid on circuit-switched technology over the same air interface to utilize cellular network's access capacity. More than one service can be multiplexed statistically on to the same radio channel since it is unlikely that all services transmit at their peak rates at the same time. The network can allocate each user less resource than the corresponding requested peak capacity to meet its statistical performance requirements. Thus, data packets can be transmitted efficiently to provide packet-switched services a QoS level comparable to that of circuit-switched services. However accurate voice traffic statistics are needed to understand the length and frequency distributions of idle periods of cellular channels assigned to voice in order to exploit the statistical multiplexing gain in cellular networks.

Traffic statistics are also used to feed network simulations with realistic traffic data or develop mathematical models to evaluate network performance analytically. Two of these statistics, call holding and channel occupancy times, are key elements to compute performance metrics such as call blocking and dropping probabilities. Channel occupancy times are measured from call starting time till the occupied channel in the respective cell is discarded due to call termination or handoff while call holding times are measured from call starting time till call termination regardless of occupying a channel in the same cell or not. In classical voice traffic modeling call holding times are approximated by exponential distribution, yet it has been shown that a lognormal distribution approximation fits much closer. However performance evaluation models for call admission control require the distribution of channel occupancy times rather than distribution of call holding times. In chapter 4, we presented an empirical approach to determine the probability distribution functions that fit channel occupancy times classified according to their occupancy types to provide sufficiently representative statistics. The results are environment dependent but no assumptions that can be influential have been made as opposed to previous analytical and simulation studies where the obtained results are highly dependent on the assumptions made by the authors. We showed that all types of channel occupancy times can be approximated by lognormal distribution. For stationary users channel occupancy times are approximated by exponential distribution due to its tractability assuming that cell residence times are also exponentially distributed. Yet we observed that lognormal distribution fits much better to channel occupancy times when users are stationary. We examined the impact of modeling call holding times with lognormal distribution on performance metrics instead of modeling with the traditionally accepted exponential distribution. When averages are same both distributions provide very close results in a single service system with one type of call. We showed that lognormal distribution is the closest fitted candidate to approximate the number of handoffs committed by a user. We expect the results to play an important role in traffic and network modeling, performance evaluation, billing planning, network management and optimization in cellular networks.

5.2 Future Work

While this thesis provides achievements for evaluating performance of call admission control mechanisms in single and multi service cellular networks by proposing computationally efficient approximation methods and modeling channel occupancy times, there are emerging issues deserving further investigation.

To facilitate the practical implementation of proposed approximation methods, we need to investigate techniques that can provide good estimations of call arrival rates and average channel occupancy times of all types of calls in real-time.

In [1], we showed that the proposed *state space decomposition* method outperforms the Borst-Mitra approach when evaluating the performance of call admission control schemes under high traffic load when the number of existing call arrival types are low. However the two-class model with the heterogeneous traffic may be the most challenging one to Borst and Mitra's approach due to their assumption on independent channel occupancy [2]. Since each of the individual classes accounts for a substantial portion of the total amount of capacity in use, it is worth investigating to what extent the performance of authors' approach can improve under high traffic load with respect to the proposed approach when more than two-class models are considered.

We showed in [3] that channel occupancy times can be approximated by lognormal distribution along with call holding times. However results obtained from a simulated cellular network revealed that performance metrics such as call blocking probabilities are not affected significantly when channel occupancy times are modeled with traditionally accepted exponential distribution providing that average values for both distributions are same. The exponential distribution underestimates the distribution of channel occupancy times for its short values while it overestimates for long ones. When the distribution of channel occupancy times is obtained using empirical data collected during certain times of a day such as the lunch time or rush hours or when the cellular operator offers discounts, this misestimation may extend. Considering that the results in [3] are environment dependent, we need to seek further investigations for system response using empirical data collected at various times.

Bandwidth scarceness is another important problem in cellular networks. Making radio cells smaller is one solution, however as cell size is reduced, more users will probably require handoffs. On the other hand, more handoffs can be observed in environments such as highway stretches where users move very fast. These are typical situations where high mobility of users may affect cell residence times and thus the characteristics of channel occupancy times [4]-[6]. It may be necessary to analyze empirical data collected from base stations serving cell sites with high mobility profile to determine its effects on distributions of the classified channel occupancy times given in [3].

Fang, Chlamtac and Lin showed analytically in [7] that channel occupancy time distributions for stationary users can be approximated by exponential distribution when cell residence times are exponentially distributed. Yet, it is difficult to obtain cell residence time data since network nodes which track idle mobile users exchange messages very infrequently. Thus, we classified users with respect to their mobility characteristics and observed that channel occupancy and call holding time distributions for stationary users fit lognormal distribution better rather than exponential distribution as opposed to results obtained in [7]. It may be interesting to investigate which distribution fits best to cell residence times and how cell residence times are related to channel occupancy times using corresponding empirical data.

5.3 Bibliography

[1] E. A. Yavuz and V. C. M. Leung, "Efficient approximations for call admission control performance evaluations in multi-service networks," presented at IEEE GLOBECOM 06, San Francisco, LA, November 2006.

[2] S. C. Borst and D. Mitra, "Virtual partitioning for robust resource sharing: computational techniques for heterogeneous traffic," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 5, pp. 668-678, June 1998.

[3] E. A. Yavuz and V. C. M. Leung, "Modeling channel occupancy times for voice traffic in cellular networks," to be presented at IEEE ICC 2007, Glasgow, Scotland, June 2007.

[4] H. Zeng, and I. E. Chlamtac, "Handoff traffic distribution in cellular networks," IEEE WCNC Wireless Communications and Networking Conference, vol. 1, pp. 413-417, Sept 1999.

[5] F. Khan and D. Zeghlache, "Effect of cell residence time distribution on the performance of cellular mobile networks," IEEE 47th Vehicular Technology Conference, vol. 2, pp. 949-953, May 1997.

[6] R. Bolla and M. Repetto, "A new model for network traffic forecast based on user's mobility in cellular networks with highway stretches," *International Journal of Communication Systems*, vol. 17, no. 10, pp. 911-934, Dec 2004.

[7] Y. Fang, I. Chlamtac, and Y. B. Lin, "Channel occupancy times and handoff rate for mobile computing and PCS networks," *IEEE Transactions on Computers*, vol. 47, pp. 679-692, June 1998.

APPENDICES

Appendix A

PROBABILITY DENSITY FUNCTIONS

1. Exponential distribution

$$f(x) = \lambda \cdot e^{-\lambda x}, x \geq 0$$

Maximum Likelihood Estimators: $\hat{\lambda} = \frac{1}{\bar{x}}$, where $\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$

2. Lognormal distribution

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \cdot e^{\left(-\left[\frac{\ln(x)-\mu}{\sigma}\right]^2 / 2\right)}$$

Maximum Likelihood Estimators: $\hat{\mu} = \frac{\sum_{i=1}^{n} \ln x_i}{n}$, and $\hat{\sigma} = \frac{\sum_{i=1}^{n} (\ln x_i - \hat{\mu})^2}{n}$

3. Gamma distribution

$$f(x) = x^{k-1} \cdot \frac{e^{-(x/\theta)}}{\theta^k \cdot \Gamma(k)}, x > 0$$

where $k > 0$ is the shape parameter, $\theta > 0$ is the scale parameter and $\Gamma(k) = \int_0^\infty t^{k-1} \cdot e^{-t} dt$

Maximum Likelihood Estimators:

$$\theta = \frac{1}{kn} \cdot \sum_{i=1}^{n} x_i \text{ and } \ln(k) - \varphi(k) = \ln\left(\frac{1}{n} \cdot \sum_{i=1}^{n} x_i\right) - \frac{1}{n} \cdot \sum_{i=1}^{n} \ln(x_i), \text{ where } \varphi(k)) = \frac{\Gamma'(k)}{\Gamma(k)}$$

There is no close form solution for *k*. The numerical solution can be found using Newton's method.

4. Weibull distribution

$$f(x) = (k/\lambda) \cdot (x/\lambda)^{(k-1)} \cdot e^{-(x/\lambda)^k}, x \geq 0$$

where *k* > *0* is the shape parameter and λ > *0* is the scale parameter of the distribution.

Maximum Likelihood Estimators: $\lambda = \frac{\sum_{i=1}^{n} x_i^k}{n}$ and $\frac{\sum_{i=1}^{n} x_i^k \cdot \ln(x_i)}{\sum_{i=1}^{n} x_i^k} - \frac{1}{k} - \frac{1}{n} \cdot \sum_{i=1}^{n} \ln(x_i) = 0$

There is no close form solution for *k*. The numerical solution can be found using Newton's method.